JMIR Medical Education

Impact Factor (2024): 3.2
Volume 11 (2025) ISSN 2369-3762 Editor-in-Chief: Blake J. Lesselroth, MD, MBI, FACP, FAMIA

Contents

Original Papers

Collaborative Development of Feedback Concept Maps for Virtual Patient–Based Clinical Reasoning Education: Mixed Methods Study (e57331)
Anja Mayer, Inga Hege, Andrzej Kononowicz, Anja Müller, Małgorzata Sudacka. 8

Using Web-Based Continuing Education to Improve New Diagnoses of Alzheimer Disease in Claims Data: Retrospective Case-Control Study (e72000)
Katie Lucero, Thomas Finnegan, Soo Borson. 151

Exploring Health Care Professionals' Perspectives on Education, Awareness, and Preferences for Digital Educational Resources to Support Transgender, Nonbinary, and Intersex Care: Interview Study (e67993)
Sravya Katta, Nadia Davoody. 159

Evaluation of an Interdisciplinary Educational Program to Foster Learning Health Systems: Education Evaluation (e54152)
Sathana Dushyanthen, Nadia Zamri, Wendy Chapman, Daniel Capurro, Kayley Lyons. 182

Recruiting Medical, Dental, and Biomedical Students as First Responders in the Immediate Aftermath of the COVID-19 Pandemic: Prospective Follow-Up Study (e63018)
Nicolas Schnetzler, Victor Taramarcaz, Tara Herren, Eric Golay, Simon Regard, François Mach, Amanta Nasution, Robert Larribau, Melanie Suppan, Eduardo Schiffer, Laurent Suppan. 196

Comparison of Learning Outcomes Among Medical Students in Thailand to Determine the Right Time to Teach Forensic Medicine: Retrospective Study (e57634)
Ubon Chudoung, Wilaipon Saengon, Vichan Peonim, Wisarn Worasuwannarak. 209

Understanding Community Health Care Through Problem-Based Learning With Real-Patient Videos: Single-Arm Pre-Post Mixed Methods Study (e68743)
Kiyoshi Shikino, Kazuyo Yamauchi, Nobuyuki Araki, Ikuo Shimizu, Hajime Kasai, Tomoko Tsukamoto, Hiroshi Tajima, Yu Li, Misaki Onodera, Shoichi Ito. 216

Organizational Leaders' Views on Digital Health Competencies in Medical Education: Qualitative Semistructured Interview Study (e64768)
Humairah Zainal, Xin Xiao Hui, Julian Thumboo, Fong Kok Yong. 228

Performance of Plug-In Augmented ChatGPT and Its Ability to Quantify Uncertainty: Simulation Study on the German Medical Board Examination (e58375)
Julian Madrid, Philipp Diehl, Mischa Selig, Bernd Rolauffs, Felix Hans, Hans-Jörg Busch, Tobias Scheef, Leo Benning. 242

XSL•FO
RenderX

Evolution of Learning Styles in Surgery Comparing Residents and Teachers: Cross-Sectional Study (e64767)
Gabriela Gouvea Silva, Carlos da Silva Costa, Bruno Cardoso Gonçalves, Luiz Vianney Saldanha Cidrão Nunes, Emerson Roberto dos Santos, Natalia Almeida de Arnaldo Rodriguez Castro, Alba de Abreu Lima, Vânia Sabadoto Brienze, Antônio Oliani, Júlio André. 255

Faculty Perceptions on the Roles of Mentoring, Advising, and Coaching in an Anesthesiology Residency Program: Mixed Methods Study (e60255)
Sydney Nykiel-Bailey, Kathryn Burrows, Bianca Szafarowicz, Rachel Moquin. 263

A Brief Web-Based Person-Centered Care Group Training Program for the Management of Generalized Anxiety Disorder: Feasibility Randomized Controlled Trial in Spain (e50060)
Vanesa Ramos-García, Amado Rivero-Santana, Wenceslao Peñate-Castro, Yolanda Álvarez-Pérez, Andrea Duarte-Díaz, Alezandra Torres-Castaño, María Trujillo-Martín, Ana González-González, Pedro Serrano-Aguilar, Lilisbeth Perestelo-Pérez. 274

Exploring Connections Between Mental Health, Burnout, and Academic Factors Among Medical Students at an Iranian University: Cross-Sectional Questionnaire Study (e58008)
Elham Faghihzadeh, Ali Eghtesad, Muhammad Fawad, Xiaolin Xu. 286

Barriers to and Facilitators of Implementing Team-Based Extracorporeal Membrane Oxygenation Simulation Study: Exploratory Analysis (e57424)
Joan Brown, Sophia De-Oliveira, Christopher Mitchell, Rachel Cesar, Li Ding, Melissa Fix, Daniel Stemen, Krisda Yacharn, Se Wong, Anahat Dhillon. 297

Evaluation of the Inverted Classroom Approach in a Case-Study Course on Antithrombotic Drug Use in a PharmD Curriculum: French Monocentric Randomized Study (e67419)
Georges Jourdi, Mayssa Selmi, Pascale Gaussem, Jennifer Truchot, Isabelle Margaill, Virginie Siguret. 311

Preclinical Medical Students' Perspectives and Experiences With Structured Web-Based English for Medical Purposes Courses: Cross-Sectional Study (e65779)
Radhakrishnan Muthukumar, Isaraporn Thepwongsa, Poompong Sripa, Bangonsri Jindawong, Kamonwan Jenwitheesuk, Surapol Virasiri. . .
3 2 4

Factors Influencing Educators' Perspectives on Accepting Extended Reality in Health Care Education: Qualitative Study (e65042)
Zuheir Khlaif, Nisreen Salama, Bilal Hamamra, Allam Mousa. 338

Assessing ChatGPT's Capability as a New Age Standardized Patient: Qualitative Study (e63353)
Joseph Cross, Tarron Kayalackakom, Raymond Robinson, Andrea Vaughans, Roopa Sebastian, Ricardo Hood, Courtney Lewis, Sumanth Devaraju, Prasanna Honnavar, Sheetal Naik, Jillwin Joseph, Nikhilesh Anand, Abdalla Mohammed, Asjah Johnson, Eliran Cohen, Teniola Adeniji, Aisling Nnenna Nnaji, Julia George. 354

Virtual Reality Simulation for Undergraduate Nursing Students for Care of Patients With Infectious Diseases: Mixed Methods Study (e64780)
Wen Chang, Chun-Chih Lin, Julia Crilly, Hui-Ling Lee, Li-Chin Chen, Chin-Yen Han. 367

Mono-Professional Simulation-Based Obstetric Training in a Low-Resource Setting: Stepped-Wedge Cluster Randomized Trial (e54911)
Anne van Tetering, Ella de Vries, Peter Ntuyo, E van den Heuvel, Annemarie Fransen, M van der Hout-van der Jagt, Imelda Namagembe, Josaphat Byamugisha, S Oei. 380

Awareness and Attitude Toward Artificial Intelligence Among Medical Students and Pathology Trainees: Survey Study (e62669)
Anwar Rjoop, Mohammad Al-Qudah, Raja Alkhasawneh, Nesreen Bataineh, Maram Abdaljaleel, Moayad Rjoub, Mustafa Alkhateeb, Mohammad Abdelraheem, Salem Al-Omari, Omar Bani-Mari, Anas Alkabalan, Saoud Altulaih, Iyad Rjoub, Rula Alshimi. 391

XSL•FO
RenderX

Reviewing Mobile Apps for Teaching Human Anatomy: Search and Quality Evaluation Study (e64550)
Guadalupe Rivera García, Miriam Cervantes López, Juan Ramírez Vázquez, Arturo Llanes Castillo, Jaime Cruz Casados. 399

Making Medical Education Courses Visible: Theory-Based Development of a National Database (e62838)
Andi Gashi, Monika Brodmann Maeder, Eva Hennel. 417

Large Language Models in Biochemistry Education: Comparative Evaluation of Performance (e67244)
Olena Bolgova, Inna Shypilova, Volodymyr Mavrych. 432

Medical Students' Acceptance of Tailored e–Mental Health Apps to Foster Their Mental Health: Cross-Sectional Study (e58183)
Catharina Grüneberg, Alexander Bäuerle, Sophia Karunakaran, Dogus Darici, Nora Dörrie, Martin Teufel, Sven Benson, Anita Robitzsch. . . 4 4 2

Integration of an Audiovisual Learning Resource in a Podiatric Medical Infectious Disease Course: Multiple Cohort Pilot Study (e55206)
Garrik Hoyt, Chandra Bakshi, Paramita Basu. 453

Performance of ChatGPT-3.5 and ChatGPT-4 in the Taiwan National Pharmacist Licensing Examination: Comparative Evaluation Study (e56850)
Ying-Mei Wang, Hung-Wei Shen, Tzeng-Ji Chen, Shu-Chiung Chiang, Ting-Guan Lin. 464

Guidelines for Patient-Centered Documentation in the Era of Open Notes: Qualitative Study (e59301)
Anita Vanka, Katherine Johnston, Tom Delbanco, Catherine DesRoches, Annalays Garcia, Liz Salmi, Charlotte Blease. 474

Implementing the H&P 360 in Three Medical Institutions: Usability Study (e66221)
Rupinder Hayer, Joyce Tang, Julia Bisschops, Gregory Schneider, Kate Kirley, Tamkeen Khan, Erin Rieger, Eric Walford, Irsk Anderson, Valerie Press, Brent Williams. 486

Resilience Training Web App for National Health Service Keyworkers: Pilot Usability Study (e51101)
Joanna Burrell, Felicity Baker, Matthew Bennion. 499

Evaluation of a Simulation Program for Providing Telenursing Training to Nursing Students: Cohort Study (e67804)
Ola Ali-Saleh, Layalleh Massalha, Ofra Halperin. 506

Distance Learning During the COVID-19 Lockdown and Self-Assessed Competency Development Among Radiology Residents in China: Cross-Sectional Survey (e54228)
Peicheng Wang, Ziye Wu, Jingfeng Zhang, Yanrong He, Maoqing Jiang, Jianjun Zheng, Zhenchang Wang, Zhenghan Yang, Yanhua Chen, Jiming Zhu. 516

Comparison of an Emergency Medicine Asynchronous Learning Platform Usage Before and During the COVID-19 Pandemic: Retrospective Analysis Study (e58100)
Blake Briggs, Madhuri Mulekar, Hannah Morales, Iltifat Husain. 532

Exploring Social Media Use Among Medical Students Applying for Residency Training: Cross-Sectional Survey Study (e59417)
Simi Jandu, Jennifer Carey. 541

Ethical Use of Social Media and Sharing of Patient Information by Medical Students at a University Hospital in Saudi Arabia: Cross-Sectional Survey (e57812)
Sara Farsi, Alaa Sabbahi, Deyala Sait, Raghad Kabli, Ghaliah Abduljabar. 551

Instagram as a Tool to Improve Human Histology Learning in Medical Education: Descriptive Study (e55861)
Alejandro Escamilla-Sanchez, Juan López-Villodres, Carmen Alba-Tercedor, María Ortega-Jiménez, Francisca Rius-Díaz, Raquel Sanchez-Varo, Diego Bermúdez. 561

XSL•FO
RenderX

Exploring the Role of Immersive Virtual Reality Simulation in Health Professions Education: Thematic Analysis (e62803)
Jordan Talan, Molly Forster, Leian Joseph, Deepak Pradhan. 575

Knowledge Mapping and Global Trends in Simulation in Medical Education: Bibliometric and Visual Analysis (e71844)
Hongjun Ba, Lili Zhang, Xiufang He, Shujuan Li. 587

Effectiveness of an Interactive Web-Based Clinical Practice Monitoring System on Enhancing Motivation in Clinical Learning Among Undergraduate Nursing Students: Longitudinal Quasi-Experimental Study in Tanzania (e45912)
Patricia Herman, Stephen M Kibusi, Walter C Millanzi. 597

Media-Induced and Psychological Factors That Foster Empathy Through Virtual Reality in Nursing Education: 2×2 Between-Subjects Experimental Study (e59083)
Kuo-Ting Huang, Zexin Ma, Lan Yao. 617

Case-Based Virtual Reality Simulation for Severe Pelvic Trauma Clinical Skill Training in Medical Students: Design and Pilot Study (e59850)
Peng Teng, Youran Xu, Kaoliang Qian, Ming Lu, Jun Hu. 628

Extended Reality–Enhanced Mental Health Consultation Training: Quantitative Evaluation Study (e64619)
Katherine Hiley, Zanib Bi-Mohammad, Luke Taylor, Rebecca Burgess-Dawson, Dominic Patterson, Devon Puttick-Whiteman, Christopher Gay, Janette Hiscoe, Chris Munsch, Sally Richardson, Mark Knowles-Lee, Celia Beecham, Neil Ralph, Arunangsu Chatterjee, Ryan Mathew, Faisal Mushtaq. 645

Exploring the Impact of the COVID-19 Pandemic on Learning Experience, Mental Health, Adaptability, and Resilience Among Health Informatics Master's Students: Focus Group Study (e63708)
Nadia Davoody, Natalia Stathakarou, Cara Swain, Stefano Bonacina. 663

Feedback From Dental Students Using Two Alternate Coaching Methods: Qualitative Focus Group Study (e68309)
Lulwah Alreshaid, Rana Alkattan. 686

Refining Established Practices for Research Question Definition to Foster Interdisciplinary Research Skills in a Digital Age: Consensus Study With Nominal Group Technique (e56369)
Jana Sedlakova, Mina Staniki , Felix Gille, Jürgen Bernard, Andrea Horn, Markus Wolf, Christina Haag, Joel Floris, Gabriela Morgenshtern, Gerold Schneider, Aleksandra Zumbrunn Wojczy ska, Corine Mouton Dorey, Dominik Ettlin, Daniel Gero, Thomas Friemel, Ziyuan Lu, Kimon Papadopoulos, Sonja Schläpfer, Ning Wang, Viktor von Wyl. 708

Global Health care Professionals' Perceptions of Large Language Model Use In Practice: Cross-Sectional Survey Study (e58801)
Ecem Ozkan, Aysun Tekin, Mahmut Ozkan, Daniel Cabrera, Alexander Niven, Yue Dong. 723

Novel Blended Learning on Artificial Intelligence for Medical Students: Qualitative Interview Study (e65220)
Zoe Oftring, Kim Deutsch, Daniel Tolks, Florian Jungmann, Sebastian Kuhn. 736

Chatbots' Role in Generating Single Best Answer Questions for Undergraduate Medical Student Assessment: Comparative Analysis (e69521)
Enjy Abouzeid, Rita Wassef, Ayesha Jawwad, Patricia Harris. 752

ChatGPT's Performance on Portuguese Medical Examination Questions: Comparative Analysis of ChatGPT-3.5 Turbo and ChatGPT-4o Mini (e65108)
Filipe Prazeres. 763

XSL•FO
RenderX

Performance of ChatGPT-4 on Taiwanese Traditional Chinese Medicine Licensing Examinations: Cross-Sectional Study (e58897)
Liang-Wei Tseng, Yi-Chin Lu, Liang-Chi Tseng, Yu-Chun Chen, Hsing-Yu Chen. 771

Generative Artificial Intelligence in Medical Education—Policies and Training at US Osteopathic Medical Schools: Descriptive Cross-Sectional Survey (e58766)
Tsunagu Ichikawa, Elizabeth Olsen, Arathi Vinod, Noah Glenn, Karim Hanna, Gregg Lund, Stacey Pierce-Talsma. 788

Assessing Familiarity, Usage Patterns, and Attitudes of Medical Students Toward ChatGPT and Other Chat-Based AI Apps in Medical Education: Cross-Sectional Questionnaire Study (e63065)
Safia Elhassan, Muhammad Sajid, Amina Syed, Sidrah Fathima, Bushra Khan, Hala Tamim. 796

Performance Evaluation and Implications of Large Language Models in Radiology Board Exams: Prospective Comparative Analysis (e64284)
Boxiong Wei. 804

Factors Associated With the Accuracy of Large Language Models in Basic Medical Science Examinations: Cross-Sectional Study (e58898)
Naritsaret Kaewboonlert, Jiraphon Poontananggul, Natthipong Pongsuwan, Gun Bhakdisongkhram. 811

Enhancing Medical Student Engagement Through Cinematic Clinical Narratives: Multimodal Generative AI–Based Mixed Methods Study (e63865)
Tyler Bland. 821

Perceptions and Earliest Experiences of Medical Students and Faculty With ChatGPT in Medical Education: Qualitative Study (e63400)
Noura Abouammoh, Khalid Alhasan, Fadi Aljamaan, Rupesh Raina, Khalid Malki, Ibraheem Altamimi, Ruaim Muaygil, Hayfaa Wahabi, Amr Jamal, Ali Alhaboob, Rasha Assiri, Jaffar Al-Tawfiq, Ayman Al-Eyadhy, Mona Soliman, Mohamad-Hani Temsah. 833

Detecting Artificial Intelligence–Generated Versus Human-Written Medical Student Essays: Semirandomized Controlled Study (e62779)
Berin Doru, Christoph Maier, Johanna Busse, Thomas Lücke, Judith Schönhoff, Elena Enax- Krumova, Steffen Hessler, Maria Berger, Marianne Tokic. 845

AIFM-ed Curriculum Framework for Postgraduate Family Medicine Education on Artificial Intelligence: Mixed Methods Study (e66828)
Raymond Tolentino, Fanny Hersson-Edery, Mark Yaffe, Samira Abbasgholizadeh-Rahimi. 859

Leveraging Datathons to Teach AI in Undergraduate Medical Education: Case Study (e63602)
Michael Yao, Lawrence Huang, Emily Leventhal, Clara Sun, Steve Stephen, Lathan Liou. 874

Effect of Immersive Virtual Reality Teamwork Training on Safety Behaviors During Surgical Cases: Nonrandomized Intervention Versus Controlled Pilot Study (e66186)
Lukasz Mazur, Logan Butler, Cody Mitchell, Shaian Lashani, Shawna Buchanan, Christi Fenison, Karthik Adapa, Xianming Tan, Selina An, Jin Ra. 887

Impact of Clinical Decision Support Systems on Medical Students' Case-Solving Performance: Comparison Study with a Focus Group (e55709)
Marco Montagna, Filippo Chiabrando, Rebecca De Lorenzo, Patrizia Rovere Querini, Medical Students. 898

Reviews

Multidisciplinary Oncology Education Among Postgraduate Trainees: Systematic Review (e63655)
Houman Tahmasebi, Gary Ko, Christine Lam, Idil Bilgen, Zachary Freeman, Rhea Varghese, Emma Reel, Marina Englesakis, Tulin Cil. 22

XSL•FO
RenderX

Bridging Gaps in Telemedicine Education in Romania to Support Future Health Care: Scoping Review (e66458)
Mircea Focsa, Virgil Rotaru, Octavian Andronic, Marius Marginean, Sorin Florescu. 39

Virtual Simulation Tools for Communication Skills Training in Health Care Professionals: Literature Review (e63082)
Manuel Fernández-Alcántara, Silvia Escribano, Rocío Juliá-Sanchis, Ana Castillo-López, Antonio Pérez-Manzano, M Macur, Sedina Kalender-Smajlovi , Sofía García-Sanjuán, María Cabañero-Martínez. 51

AI in the Health Sector: Systematic Review of Key Skills for Future Health Professionals (e58161)
Javier Gazquez-Garcia, Carlos Sánchez-Bocanegra, Jose Sevillano. 69

Motivation Theories and Constructs in Experimental Studies of Online Instruction: Systematic Review and Directed Content Analysis (e64179)
Adam Gavarkovs, Erin Miller, Jaimie Coleman, Tharsiga Gunasegaran, Rashmi Kusurkar, Kulamakan Kulasegaram, Melanie Anderson, Ryan Brydges. 81

Online-Based and Technology-Assisted Psychiatric Education for Trainees: Scoping Review (e64773)
Mohd Mohd Kassim, Sidi Azli Shah, Jane Lim, Tuti Mohd Daud. 95

Viewpoints

Leveraging Generative Artificial Intelligence to Improve Motivation and Retrieval in Higher Education Learners (e59210)
Noahlana Monzon, Franklin Hays. 116

Digital Dentists: A Curriculum for the 21st Century (e54153)
Michelle Mun, Samantha Byrne, Louise Shaw, Kayley Lyons. 679

Tutorials

Designing Personalized Multimodal Mnemonics With AI: A Medical Student's Implementation Tutorial (e67926)
Noor Elabd, Zafirah Rahman, Salma Abu Alinnin, Samiyah Jahan, Luciana Campos, Ovidiu Baltatu. 127

Creation of the ECHO Idaho Podcast: Tutorial and Pilot Assessment (e55313)
Ryan Wiet, Madeline Casanova, Jonathan Moore, Sarah Deming, Russell Baker Jr. 135

Enhancing Access to Neuraxial Ultrasound Phantoms for Medical Education of Pediatric Anesthesia Trainees: Tutorial (e63682)
Leah Webb, Melissa Masaracchia, Kim Strupp. 143

Letters to the Editor

Citation Accuracy Challenges Posed by Large Language Models (e72998)
Manlin Zhang, Tianyu Zhao. 696

XSL•FO
RenderX

Authors' Reply: Citation Accuracy Challenges Posed by Large Language Models (e73698)
Mohamad-Hani Temsah, Ayman Al-Eyadhy, Amr Jamal, Khalid Alhasan, Khalid Malki. 698

Examining Multimodal AI Resources in Medical Education: The Role of Immersion, Motivation, and Fidelity in AI Narrative Learning (e72190)
Chris Jacobs. 700

Enhancing AI-Driven Medical Translations: Considerations for Language Concordance (e70420)
Stephanie Quon, Sarah Zhou. 702

Author's Reply: Examining Multimodal AI Resources in Medical Education: The Role of Immersion, Motivation, and Fidelity in AI Narrative Learning (e72336)
Tyler Bland. 704

Authors' Reply: Enhancing AI-Driven Medical Translations: Considerations for Language Concordance (e71721)
Joyce Teng, Roberto Novoa, Maria Aleshin, Jenna Lester, Kira Seiger, Fiatsogbe Dzuali, Roxana Daneshjou. 706

Commentary

Transforming Medical Education to Make Patient Safety Part of the Genome of a Modern Health Care Worker (e68046)
Peter Lachman, John Fitzsimons. 908

XSL•FO
RenderX

Collaborative Development of Feedback Concept Maps for Virtual Patient–Based Clinical Reasoning Education: Mixed Methods Study

Anja Mayer[1], MPH; Inga Hege[2], MSc, MD; Andrzej A Kononowicz[3], PhD; Anja Müller[1], MD; Małgorzata Sudacka[4], MD

[1]Medical Education Sciences, University of Augsburg, Augsburg, Germany

[2]Institute for Research in Health Science Education, Brandenburg Medical School Theodor Fontane, Neuruppin, Germany

[3]Department of Bioinformatics and Telemedicine, Jagiellonian University Medical College, Kraków, Poland

[4]Department of Medical Education, Center for Innovative Medical Education, Jagiellonian University Medical College, Kraków, Poland

Corresponding Author:
Anja Mayer, MPH
Medical Education Sciences, University of Augsburg, Augsburg, Germany

Abstract

Background: Concept maps are a suitable method for teaching clinical reasoning (CR). For example, in a concept map, findings, tests, differential diagnoses, and treatment options can be documented and connected to each other. When combined with virtual patients, automated feedback can be provided to the students' concept maps. However, as CR is a nonlinear process, feedback concept maps that are created together by several individuals might address this issue and cover perspectives from different health professionals.

Objective: In this study, we aimed to develop a collaborative process for creating feedback concept maps in virtual patient–based CR education.

Methods: Health professionals of different specialties, nationalities, and levels of experience in education individually created concept maps and afterward reached a consensus on them in structured workshops. Then, medical students discussed the health professionals' concept maps in focus groups. We performed a qualitative content analysis of the transcribed audio records and field notes and a descriptive comparison of the produced concept maps.

Results: A total of 14 health professionals participated in 4 workshops, each with 3 - 4 participants. In each workshop, they reached a consensus on 1 concept map, after discussing content and presentation, as well as rationales, and next steps. Overall, the structure of the workshops was well-received. The comparison of the produced concept maps showed that they varied widely in their scope and content. Consensus concept maps tended to contain more nodes and connections than individual ones. A total of 9 medical students participated in 2 focus groups of 4 and 5 participants. Their opinions on the concept maps' features varied widely, balancing between the wish for an in-depth explanation and the flexibility of CR.

Conclusions: Although the number of participating health professionals and students was relatively low, we were able to show that consensus workshops are a constructive method to create feedback concept maps that include different perspectives of health professionals with content that is useful to and accepted by students. Further research is needed to determine which features of feedback concept maps are most likely to improve learner outcomes and how to facilitate their construction in collaborative consensus workshops.

(JMIR Med Educ 2025;11:e57331) doi:10.2196/57331

KEYWORDS

clinical reasoning; consensus building process; concept map; consensus map; virtual patient; international collaboration; health professionals' education; undergraduate; collaborative; development; feedback; content analysis; health professional; medical student; mixed method; Europe; questionnaire; descriptive analysis

Introduction

Background

"Clinical reasoning encompasses health professionals thinking and acting in assessment, diagnostic, and management processes in clinical situations taking into account the patient's specific circumstances and preferences" [1]. It is evident that health professionals in different disciplines (eg, physicians and nurses) differ in their reasoning approaches [2], and there are differences between novices and experts [3]. Even experienced health professionals of the same discipline do not follow the same

XSL•FO
RenderX

diagnostic process, even when they are confronted with the same medical case, and ultimately arrive at the same diagnosis [4,5]. A study by Charlin et al [6] showed that experts' case solutions also varied depending on the situation, for example, whether they were asked to give answers as an examinee or as a panel member.

The variety and nonlinearity of possible clinical reasoning (CR) approaches make CR training and assessment a highly complex matter [4,7]. Therefore, concept maps have been suggested as a useful method for training the CR skills of medical students [4,8], especially in terms of problem representation [9].

Concept mapping is a method used to represent concepts and their relationships in a visual diagram, using explanatory terms to relate concepts to each other [10]. A typical use in health education is to present students with a case scenario and have them create a concept map to represent their thought process as the case unfolds [9,11]. They can record relevant findings, tests, differential diagnoses, and treatment options and connect concepts to each other to visualize their CR process [8]. Torre et al [12] show that concept maps promote the connection between theory and practice and facilitate knowledge integration and critical thinking.

Teachers can ask students to create concept maps in different forms, depending on the purpose, such as freely from scratch or in a preconstructed form [10,11]. Because creating a comprehensive and accurate concept map is time-consuming and students need some time to learn how to do it, Daley and Torre [8] suggest the use of semistructured concept maps.

Concept maps have also been found to be suitable for measuring learning outcomes [13], and various ways of assessing and scoring concept maps, both qualitatively and quantitatively, have been described in the literature [14-17]. A study by Morse and Jutras [18] showed that working with concept maps had an effect on the students' problem-solving performance only when feedback was provided. However, in order to provide students with feedback on their concept map, some form of "expert concept map" is needed to compare students' results with [19], which can then be provided in real time in digital environments. Such "expert concept maps" can be created by a single teacher or by a panel of professionals or experts [19-21]. In their systematic review of different methods for assessing CR skills, Daniel et al [9] concluded that "using written cases, expert consensus is the most prevalent method" used to create concept maps as feedback for students. However, little is known about the process and challenges involved when health professionals are asked to reach a consensus on a concept map for teaching CR.

Recent studies suggest that virtual patients (VPs) are an appropriate method for training CR [22-24], especially for some components of this process, such as collecting data, generating differential diagnoses, or developing a treatment plan [25,26]. VPs are computer-based patient case scenarios that students can interact with [27]. Often, such scenarios are designed so that the cases gradually lead the student to the final diagnosis by providing more and more information over time [28,29]. VPs provide a safe environment, in which mistakes can be made without harming real patients [30]. It has been suggested that combining concept map activities with VPs can reinforce the educational effect of VPs in CR outcomes [31]. The importance of VPs has increased over the years [32], especially since the beginning of the COVID-19 pandemic, when direct patient contact and opportunities for CR training were limited [33].

Objectives

In this study, we aimed to develop a collaborative process for creating feedback concept maps in VP-based CR education. From this, we derive the following research questions: (1) What are the similarities and differences of concept maps for teaching CR that have been created by individual health professionals and groups? (2) What themes emerge when health professionals are asked to jointly create a concept map in a consensus workshop? (3) What are the challenges and benefits of such consensus workshops? (4) What aspects of the consensus concept maps do medical students find helpful in learning CR?

Methods

Study Design

This study followed a convergent mixed methods approach. First, we asked health professionals from different disciplines to individually create concept maps for 2 VPs that would serve as feedback for medical students. We then conducted structured digital workshops for those health professionals in which they reached a consensus on the concept maps. After the workshops were finished, we conducted focus groups with medical students to discuss which aspects of the professionals' concept maps they found helpful for learning CR.

Ethical Considerations

The study was approved by the institutional review board of the Ludwig-Maximilians-University, Munich, Germany (21 - 0941), and adhered to ethical guidelines. Informed consent was obtained from all participants prior to their participation in the study, with assurances of anonymity and confidentiality. Participants were informed of the objectives of the study and how the data collected would be used. In addition, strict measures were taken to protect the privacy and confidentiality of the study data. Students who participated in the focus groups received a US $16 voucher as compensation for their time.

Data Collection

Between November 2021 and January 2022, we sent out emails to invite health professionals of different specialties, nationalities, and levels of experience in medical education to participate in our study. The email included study information and a written informed consent form. After returning the signed written consent by email, participants were asked to create concept maps for 2 VP cases. We used the software platform CASUS (Instruct gGmbH), which is a VP player and authoring environment with integrated concept map functionality [31]. The study participants were told that these concept maps would serve as feedback for medical students. We carefully chose the VPs with regard to their sociodemographic features, key symptoms, and difficulty levels. They were a 19-year-old female student with mononucleosis and a 58-year-old male nurse with hepatitis E. We chose them for providing patients of different

XSL•FO
RenderX

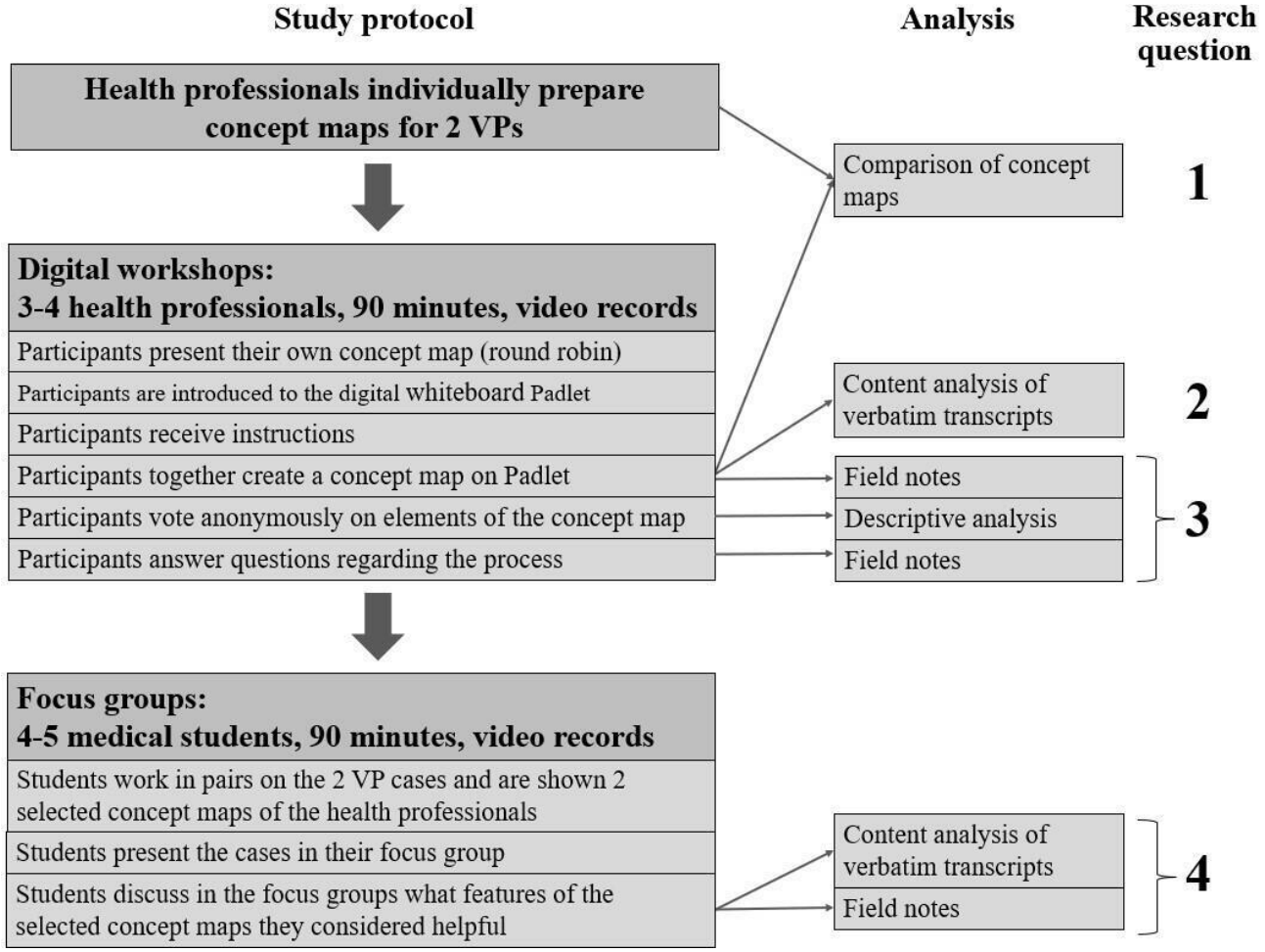
sex, age, and profession. Both were heterosexual and of Caucasian origin. We deliberately chose these VPs because they provided different levels of difficulty for the students but had key symptoms that are common in daily practice and easily recognizable by health professionals of different specialties. The 2 VPs can be found on the CASUS platform and are part of a collection of over 200 freely available VPs in 6 languages [34]. These VPs include a semistructured concept map that students fill out while solving the case, covering 4 categories: findings, differential diagnoses, tests or examinations, and treatments [31].

Participants were also asked to complete a web-based questionnaire that included personal data and level of experience in concept mapping and teaching CR. We used a convenience sampling strategy, inviting partners from 2 recent Erasmus+ projects, iCoViP (International Collection of Virtual Patients) and DID-ACT (Developing, Implementing, and Disseminating an Adaptive Clinical Reasoning Curriculum for Healthcare Students and Educators) [35,36], who were interested in teaching CR with concept maps. As the VPs are available in multiple languages, we valued the international composition of the study group that would reduce local bias in clinical practice. Participants were given 10 days to create the individual concept maps and were reminded of the task 3 days before the workshop. After that, we held structured digital workshops of 90 minutes, where they met in groups of 3‑4 to reach a consensus on a common concept map. The workshops took place on the Zoom platform (Zoom Video Communications) and were video recorded. A Mayer and MS facilitated the workshops, following a predefined structure according to the nominal group technique [37,38] (Figure 1): first, all participants explained their own concept maps to the others in a round robin and described their reasoning. Then, A Mayer and MS introduced them to the digital whiteboard Padlet (Wallwisher Inc), and the participants had the opportunity to try it out. Once they felt comfortable with the tool, A Mayer and MS gave them instructions on how to create a concept map together. They then created a new concept map on Padlet based on their individual concept maps. A Mayer and MS answered participants' questions, kept track of the timeline, and reminded participants of the original assignment if they strayed from the topic. When the concept map was complete, A Mayer and MS provided the opportunity to anonymously rate the concepts and connections with a thumbs up or down mechanism on Padlet. Afterward, they asked the participants about their experience of creating the concept map together. IH and AAK attended the workshops as neutral observers and, together with A Mayer and MS, took field notes, which they all discussed immediately after the workshop. The study was piloted as a face-to-face workshop in October 2021. Afterward, we decided that web-based meetings would be equally feasible and made minor changes to the study protocol, such as adding an anonymous voting round.

After all workshops were completed, IH and MS selected 4 individual and 4 consensus concept maps to be discussed by medical students in focus groups. For this purpose, 9 international medical students were recruited to participate in 90-minute focus groups during a transnational meeting of the iCoViP project. Written informed consent was obtained prior to participation. In the beginning, the students were asked to work in small teams (2‑3 students) and solve 1 of the 2 VP cases together. Afterward, the teams were shown 2 of the selected concept maps from the workshops to compare and decide which one they would prefer to have as feedback for their case and why. Then, 2 teams of students who had worked on different cases were brought together as a focus group. They presented their cases to the others and then started a group discussion, facilitated by A Mayer and MS, about which of the presented concept maps they found most helpful and how different features of the concept maps could improve their CR process.

XSL•FO
RenderX

Figure 1. Flowchart of study protocol and analysis plan with regard to the research questions. VP: virtual patient.



Data Analysis

This study used a convergent mixed methods design. In the quantitative part, a descriptive statistical analysis of the questionnaire, concept maps, and votes was performed using Microsoft Excel. The individual and consensus concept maps were analyzed for scope (number of nodes and connections), agreement (number of "likes" of nodes or connections in the consensus phases of the group concept map authors), and content (number of times a particular concept, eg, "fever," appeared in the individual and consensus concept maps). We extracted information from the concept maps and compared them separately for each of the 2 cases.

The qualitative part of the study involved the thematic analysis of the transcripts and field notes from the workshops and focus groups. It was conducted in several steps. The recordings of the workshops were transcribed verbatim and anonymized. Two authors (A Mayer and A Müller) performed a thematic analysis of the transcripts, following the 6 steps for qualitative content analysis proposed by Kuckartz [39]. Using an inductive approach, they independently created codes for the first 2 workshops and reached a consensus on an initial coding framework. They then coded 1 workshop at a time, applying and refining the coding framework in an iterative process. They used MAXQDA software (version Analytics Pro 2022; VERBI GmbH) for coding and discussed discrepancies until a consensus was reached. A Mayer, A Müller, and IH then grouped similar codes into themes. Throughout the process, AAK and MS reviewed the coding framework and emerging themes and provided feedback; discrepancies were discussed until a consensus was reached.

We analyzed field notes taken during the workshops and participants' responses during the round of questions for challenges and benefits. Student focus group recordings were transcribed verbatim and anonymized. Two authors (A Mayer and A Müller) independently extracted statements from the transcripts about what students found helpful in the selected concept maps, grouped them into themes, and discussed discrepancies until a consensus was reached. Finally, we looked for confirmation or discrepancies of the results obtained from the mixed methods.

Results

Participants

A total of 14 health professionals from 6 European countries participated in our study, of whom 9 were female and 5 were male. On average, participants were 37 (SD 10) years of age and had 10 (SD 9) years of professional experience. Participants worked in different disciplines (Table 1) and had an average of 6 (SD 5) years of experience in health education. Participants differed only slightly in their teaching experience with concept maps or CR.

XSL•FO
RenderX

Table . Characteristics of participating health professionals (N=14).

Characteristics	Values
Age (years), mean (SD)	37 (10)
Sex, n (%)	
Female	9 (64)
Male	5 (36)
Country (place of work), n (%)	
France	1 (7)
Germany	3 (21)
Poland	2 (14)
Portugal	1 (7)
Spain	4 (29)
Sweden	3 (21)
Specialty, n (%)	
Internal medicine	4 (29)
Nursing	2 (14)
Biochemistry	2 (14)
Rheumatology	2 (14)
Family medicine	1 (7)
Neurology	1 (7)
Paramedic	1 (7)
Occupational medicine	1 (7)
Professional experience of physicians (n=9), n (%)	
Resident	6 (67)
Consultant	3 (33)
Working experience (years), mean (SD)	10 (9)
Experience in health teaching (years), mean (SD)	6 (5)
Experience in teaching with concept maps, n (%)	
None	9 (64)
Some	5 (36)
Much	0 (0)
Experience in teaching clinical reasoning, n (%)	
None	5 (36)
Some	9 (64)
Much	0 (0)

Participants created 13 individual concept maps prior to the workshops. We held 4 digital workshops with 3‐4 participants each, resulting in 4 consensus concept maps (2 hepatitis E and 2 mononucleosis). We also conducted 2 focus groups with 4 and 5 medical students, respectively. The students were in their final year of study (sixth year), with an average age of 24 (SD 0.5) years. In total, 8 students were female, and 1 was male. We chose students from Portugal (n=5) and Poland (n=4) because these countries represent educational systems from different parts of Europe.

Research Question 1: Comparison of Individual and Consensus Concept Maps

The individual concept maps varied widely from each other regarding scope and content. We found most similarities in the final diagnoses and treatment options and only a few similarities regarding findings, differential diagnoses, and tests. The same was true when comparing the consensus concept maps.

When we compared the consensus concept maps to the individual versions, we found that they all had a bigger scope than the individual concept maps, as can be seen in Table 2 and

XSL•FO
RenderX

in the examples given in Figure 2 (original images are provided in Multimedia Appendix 1). We also found that most of the nodes from the individual concept maps were present in the consensus versions, and only in a few cases were nodes left out or new nodes added during the workshops. Altogether, the consensus versions showed higher similarities to the underlying individual versions than to each other. All consensus concept maps included connections, while these were missing in 5 of the individual versions.

Table . Comparison of number of elements in consensus and individual concept maps.

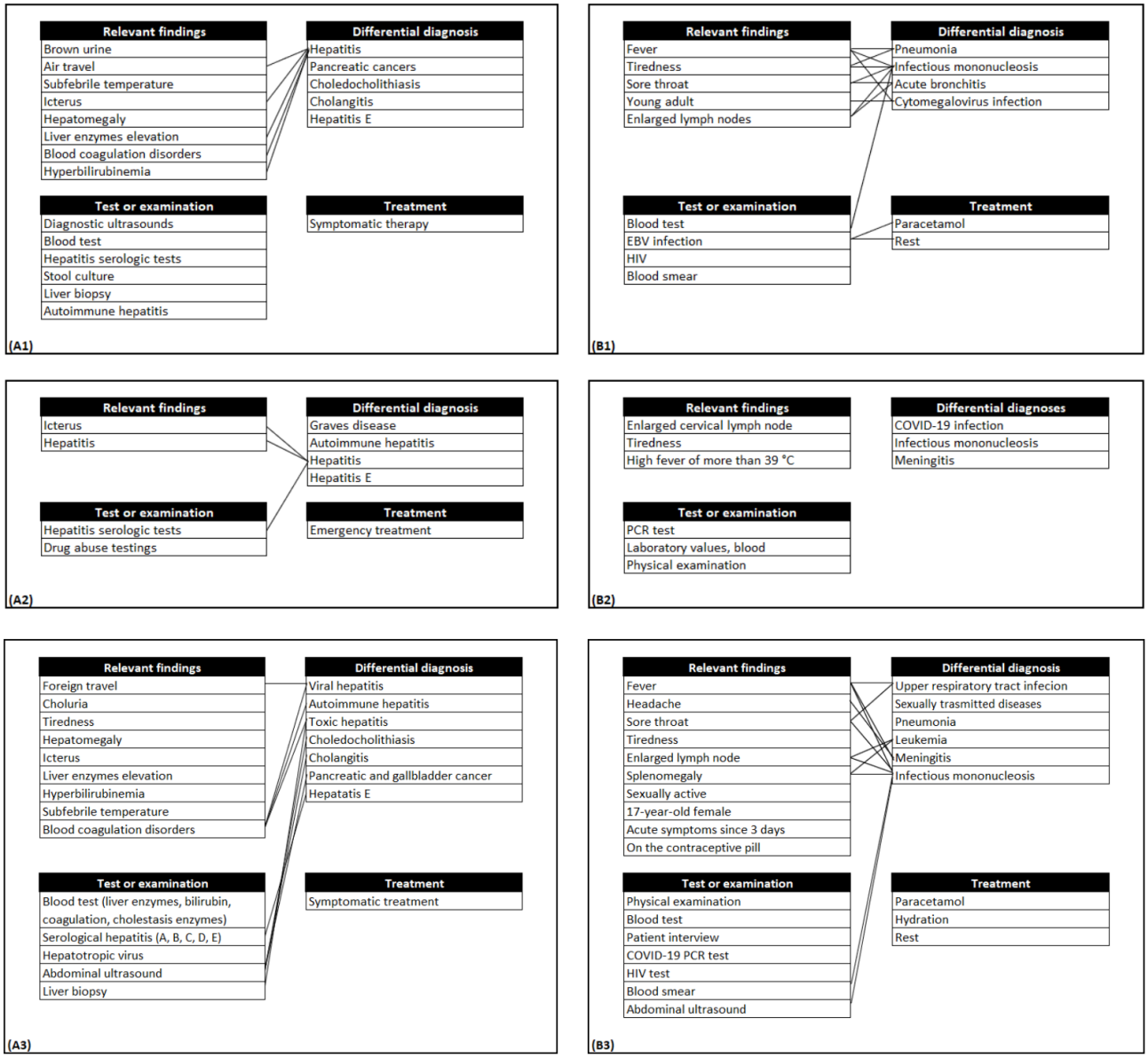
	Hepatitis E				Mononucleosis			
	Workshop 1		Workshop 2		Workshop 3		Workshop 4	
	n (GRP)[a] (%)	Rn (IND)[b]	n (GRP) (%)	Rn (IND)	n (GRP) (%)	Rn (IND)	n (GRP) (%)	Rn (IND)
Total	56 (100)	10-29	33 (100)	11-25	39 (100)	13-22	43 (100)	9-11
Nodes								
Findings	9 (16)	6-8	9 (27)	1-8	10 (26)	3-6	12 (28)	3-4
Examinations or tests	11 (20)	2-8	5 (15)	2-5	7 (18)	2-4	7 (16)	0-3
Differential diagnoses	11 (20)	1-8	7 (21)	3-6	7 (18)	4-7	6 (14)	3-7
Treatments	1 (2)	1-1	1 (3)	1-1	3 (8)	0-2	2 (5)	0-0
Connections	24 (43)	0-13	11 (33)	3-5	12 (31)	0-13	16 (37)	0-0

[a]n (GRP): number of elements in the group consensus concept map.

[b]Rn (IND): range of element number in the individual concept maps.

XSL•FO
RenderX

Figure 2. Examples of individual concept maps and consensus versions. (A1 and A2) Individual concept map (hepatitis E), (A3) consensus concept map (hepatitis E), (B1 and B2) individual concept map (mononucleosis), and (B3) consensus concept map (mononucleosis). EBV: Epstein-Barr virus; PCR: polymerase chain reaction



Research Question 2: Themes Emerging During the Consensus Process

From the qualitative content analysis of the creation phase, we identified 4 themes: the first theme covered the content of the consensus concept maps, that is, participants discussed which findings, examinations or tests, differential diagnoses, treatments, and connections should be included. Related to this, the second theme was the rationales they gave during their discussion, that is, why they thought something should (not) be part of the consensus concept map. The third theme covered the presentation of the consensus concept map, that is, participants discussed how to present the content. In the fourth theme, participants discussed their next steps, that is, how to approach the creation of the consensus concept map. Table 3 shows the 4 themes and associated subthemes, including sample quotes, and their frequency in the verbatim transcripts of the workshops.

XSL•FO
RenderX

Table . Themes and associated subthemes derived from the qualitative content analysis.

Themes and subthemes	Description	Sample quotes	Values, n (%)
Content of the consensus concept maps (n=677)			
Relevant findings	Which findings (not) to include in the concept map	“Sore throat” [Workshop 4]	227 (41)
Differential diagnoses	Which differentials (not) to include in the concept map	“Should we add EBV and CMV as a differential?” [Workshop 1]	200 (30)
Tests or examinations	Which tests or examinations (not) to include in the concept map	“Liver biopsy could be relevant” [Workshop 2]	170 (25)
Treatment	Which treatments (not) to include in the concept map	“Paracetamol I think is perfect in this case” [Workshop 3]	25 (4)
Connections	Which connections (not) to include in the concept map	“Maybe this splenomegaly should be also connected?“ [Workshop 3]	55 (8)
Presentation of the content (n=89)			
Layout	Visual aspects of the concept map or possible actions such as highlighting, crossing out, rearranging, merging, splitting, enlarging, or reducing nodes	“I’m just putting this in a nice order” [Workshop 3], “Can I change the color of the connections?” [Workshop 1]	40 (45)
Categorization	Which heading a node should be assigned to	“Could we add other headings, for example ‘recommendations’?” [Workshop 4]	6 (7)
Phrasing	Use of synonyms or abbreviations	“Is writing ‘STDs’ appropriate or should I use ‘sexually transmitted disease’?” [Workshop 3]	32 (36)
Level of granularity	Detail level of the concepts	“Viral hepatitis, autoimmune hepatitis or [just] hepatitis?” [Workshop 2]	11 (12)
Rationales (n=318)			
Medical relations	Medical relations between the concepts, including the probability of differential diagnoses	“I think about it because it’s a young female on the pill” [Workshop 3], “We think about it because it’s quite common” [Workshop 1]	96 (30)
Relevance	Highlighting the medical urgency or indicating that most participants are of the same opinion	“It’s a potentially dangerous situation for our patient” [Workshop 3], “And also COVID-19, I think we all agree” [Workshop 4]	42 (13)
Individual concept maps	Referring to individual concept maps or clinical reasoning process when creating them	“In the individual mapping, we have PCR-test, someone wrote that” [Workshop 4], “Was this something you came up with now during this process or [when creating your concept map]?” [Workshop 1]	62 (19)
Referring to the case	Referring to the case by quoting or repeating facts	“He’s not saying that he takes any drugs” [Workshop 2], “The case has provided us a biopsy” [Workshop 1]	65 (20)
Professional experience	What participants have experienced in daily practice or what they are accustomed to doing	“This is usually the first serology I order” [Workshop 1]	19 (6)
Common knowledge	General phenomena in society or “universal truths”	“People lie – he might be an alcoholic” [Workshop 2] “One would expect that this nurse is already immunized” [Workshop 1]	5 (2)

XSL•FO
RenderX

Themes and subthemes	Description	Sample quotes	Values, n (%)
Encounter setting	Regional standards or differences between facilities (hospital, general practice, etc)	"How [do] you have it in Spain or Germany?" [Workshop 4], "I was wondering whether I would have done urine analysis in the [general] practice" [Workshop 1]	9 (3)
Hindsight	Assumption that the consensus process might be unconsciously guided by already knowing the final diagnosis	"[We think so because] we already know that it's mononucleosis" [Workshop 3]	6 (2)
Didactical aspects	What could be helpful for the students or is the content appropriate for their level of knowledge, etc	"It could be a good training for students, to think what can cause hepatitis" [Workshop 1], "I think it's too specialistic" [Workshop 3]	8 (3)
Functionality of the VP[a] platform	Features, navigation, or structure of the CASUS platform	"I don't know if this is possible on CASUS" [Workshop 2]	6 (2)
Next steps (n=107)			
Developing a strategy	How to approach the creation of the concept map	"[Let's] do differentials first before adding anything to tests" [Workshop 3]	78 (73)
Referring to facilitators	Referring to instructions given by facilitators or directly asking them for advice	"[It depends on] what is wanted or what is expected" [Workshop 1]	29 (27)

[a]VP: virtual patient.

Research Question 3: Challenges and Benefits of the Workshops

The results presented here are a summary of the field notes from the creation phase and the final round of questions, expanded by a descriptive analysis of the voting round. From a technical point of view, there were some problems due to the digital format, for example, weak network signal, low audio quality, or some participants feeling uncomfortable using Padlet for the first time. Since none of the participants were native English speakers, some struggled to find the right terms or misunderstood what others were saying due to a lack of vocabulary or the speaker's accent.

Regarding the different disciplines, it seemed that participants who had worked in their specialty for many years were somewhat biased by their daily experiences and had difficulty seeing the cases from a student's point of view. Some of the participants who were not physicians by training struggled to find the right diagnosis and expressed their uncertainty about certain medical terms or conditions. It was noticeable that topics such as didactic purpose, uncertainty (probability of differential diagnoses), or logical arrangement of nodes were hardly discussed.

All participants were cooperative and reached a consensus on the concept maps in an amicable manner. For about 10% (n=10) of the nodes, half (or more) of the participants abstained from voting or gave a thumbs down. We compared these nodes with the verbatim transcripts and found that for 6 nodes, there was no evidence in the discussion that any of the participants disagreed.

In the final round of questions, participants reported that creating a concept map with others was a complex task. On the other hand, participants found the group work helpful in stimulating their reflection and that it was constructive to create concept maps that included perspectives of different health professionals. In general, the structure of the workshops and the given timeline for the different parts were well-received. The round robin was seen as a useful introduction that helped them to understand the reasoning of other participants. Some participants mentioned that it was difficult to create a concept map for a case that they had not developed themselves or that they struggled with the fact that the case evolved over time, which made it more difficult to agree on a final version. Participants had mixed feelings regarding the usefulness of the consensus concept maps. While some were satisfied with the final concept maps and expected them to be helpful for students, others found the concept maps too messy or crowded in the end.

Research Question 4: What Students Considered as Helpful

When the medical students were asked whether they preferred the individual or consensus concept maps, there was a slight tendency toward the consensus versions as they contained more findings, which the students found helpful for their own CR process.

Regarding the content and scope of the concept maps, there was agreement that there should not be too many connections between nodes, as this was seen as more confusing than helpful. However, the students expressed contradictory opinions regarding the nodes. While some preferred the concept maps with only the most relevant nodes, others preferred those with a wider scope, as these would contain "the most details that we also agreed on while we were solving the case."

XSL•FO
RenderX

The same was true for the presentation of the content. Some students suggested having more layout features, such as “some type of colors” or dropdown functions, while others preferred a clear design and simple structure. Regarding the granularity of the nodes, some suggested that “the feedback map should be more [general.] To give us freedom” and should use broad terms such as “blood test.” Others said that in their medical school, they “can’t just say ‘do blood test,’ [but] must be very specific”; therefore, more specific terms would be helpful in the concept maps.

Discussion

Main Findings

In this study, we described the process of collaborative authoring of concept maps to serve as feedback in CR education using VPs. The participants regarded the collective process stimulating for reflection and helpful to understand the perspectives of the other health professional groups. We were able to find confirmation for this qualitative finding quantitatively by showing that the consensus concept maps contained more nodes and connections than the individual ones. This can also have negative aspects, as in the consensus workshops, participants tended to collect all nodes from the individual concept maps into the consensus version instead of selecting only the most relevant ones, paying little attention to didactic aspects. The structure of the workshops was well-received, participants appreciated working in interprofessional groups and easily reached a consensus, supporting their additions to the concept maps by high scoring of the concept map elements. However, there were some challenges, such as technical problems or participants being biased by their daily practice as specialists.

The final-year medical students in our focus groups preferred a variety of features of the concept maps, most of which were contradictory. As a result, it remains unclear which features can improve learners’ outcomes and whether consensus concept maps are more suitable for teaching CR than individual ones.

Implications of the Findings

Our research suggests that there are a few approaches to help health professionals reach a consensus on a concept map. The procedure we used for the workshops served its purpose and was well-received by the participants. Thus, the results of this study can be seen as an important step toward establishing a sound consensus concept map protocol, informing about the benefits and challenges, and leading to the following recommendations for improving the process in the future:

1. Regarding the technical aspects of the workshops, we recommend that participants be given access to the digital whiteboard prior to the workshop so that those who wish to can familiarize themselves with the tool in advance.
2. Since didactic aspects played a minor role in the creation of the consensus concept maps, we recommend that an independent person with experience in didactics and concept mapping participates at the workshop. An alternative would be to prepare a pedagogical guide or checklist to be considered when developing concept maps for teaching CR. If such an opportunity arises, addressing the pedagogical aspects of concept map development would be a helpful element of faculty development courses on VP authoring.
3. When considering concept maps for VPs that address general CR skills in medicine, such as the one in the iCoViP project repository, workshops should preferably involve only internal medicine or family medicine physicians to avoid specialty bias. This would be different if the goal of the VPs was to achieve learning objectives for specialty or interprofessional education from the outset.

In terms of real-world implementation, we consider this study an important step in providing more diverse feedback to students working on CR concept maps in the context of VPs. This study contributed to this by showing that the concept maps created by consensus groups were more elaborate, both in terms of representing many viewpoints and in terms of the number of concepts and connections. However, this study also showed that the consensus groups should be more effectively encouraged to discuss the pedagogical aspects of the concept maps, such as how to adjust the complexity to the level of knowledge or cognitive load of the students.

Limitations

Our study has several limitations. First, the number of concept maps underlying the quantitative analysis was limited, so that the corresponding results should be interpreted with caution.

Second, it is possible that the results of the workshop are not applicable to “real-world” situations, in which colleagues work together on a concept map without being observed. The participants in our workshops were very polite to each other and tended to avoid disagreements, probably because most of them did not know each other. On the other hand, we were able to include the perspectives of professionals from different disciplines.

Third, we had a limited number of students in the focus groups. Our data suggest that the effect of different features of concept maps on individual learning and preferences may vary considerably from student to student. Future research is needed to explore this in more depth.

Fourth, the sample size of VPs and workshop participants was limited, which might make our findings less generalizable. However, we did not see any new themes emerging in the subsequent workshops and focus groups, suggesting that the qualitative analysis had reached its saturation point.

Comparison With Prior Work

There is a large body of literature on the so-called “group concept mapping” [14,40], including approaches to optimizing group compositions [41] or to identifying different cognitive styles [42]. However, to the best of our knowledge, most of these studies only include undergraduate students. Therefore, our study can be considered unique in proposing a novel approach to consensus concept mapping for health professionals.

The structure of the workshops, derived from the nominal group technique and adjusted to the needs of digital education, can be seen as a major strength. First, the round robin allows participants to gain insight into one another’s CR approach.

XSL•FO
RenderX

Second, participants found the consensus creation of the concept maps useful and inspiring. Third, anonymous voting at the end facilitates the interpretation of the final concept map, as it gives the participants' view on each individual element without the need to openly criticize someone else's ideas.

When we compared the individual concept maps, we found a common tendency but also a great deal of variation, with most having only the final diagnosis and treatment in common. This is supported by the work of McGaghie et al [43,44], which shows the wide variety of approaches to a concept map of pulmonary physiology, even among experts in the same field. Therefore, the consensus process in our approach increased the universality of the feedback concept maps. Another positive aspect is that the process contributed to a rational increase in the number of connections in the concept maps. As previous research has shown, well-chosen connections are an important element of this form of knowledge representation, which is helpful in CR education [45].

We did not exclude from the study professionals without teaching experience, as we did not see clear evidence that this might be a limiting factor in creating meaningful concept maps. This is supported by a study by Charlin et al [46], who found that teaching and nonteaching physicians were similarly well suited to be part of the reference panel for concordance tests used to assess complexity and ambiguity in CR.

While some authors suggest the use of concept maps for CR assessment [47], most researchers in the field are ambivalent on this issue [8,9]. This is consistent with our findings, which suggest that the complexity and variety of the CR process make it very challenging to generate "expert concept maps" that can be used as a gold standard against which student versions can be compared.

Participants reported that they found the consensus workshops useful for reflecting on their individual concept maps and CR approach. Therefore, such workshops could also be a suitable tool for improving the concept map development of the individual participants. Further research is needed to determine the impact of the workshops on participants' ability to develop concept maps.

Our study focused on the creation of concept maps for medical students. Future research should investigate how this can be applied to other professions, as a recent meta-analysis suggests that concept maps are also an appropriate method for improving critical thinking skills in nursing students [48].

Conclusions

By providing feedback concept maps that illustrate the complexity and diversity of the CR process, we aim to support students in reflecting on their own thinking. The collaborative creation of concept maps for teaching CR is an opportunity to integrate different perspectives of health professionals and to account for individual differences in the reasoning process. In our study, we described a process for developing such collaborative concept maps and identified themes that emerged in workshops using this process. The resulting consensus concept maps tended to contain more nodes and connections than those created by individual health professionals and were well-received by students. We consider this study an important step in establishing a robust method for collaboratively creating effective concept maps in CR education. Future studies will focus on streamlining the process and identifying the most effective pedagogical features of feedback concept maps.

Acknowledgments

The creation of these resources has been partially funded by the Erasmus+ grant program of the European Union (grant 2020-1-DE01-KA226-005754). Neither the European Commission nor the project's national funding agency DAAD (German Academic Exchange Service) are responsible for the content or liable for any losses or damage resulting from the use of these resources. The authors would like to thank all the participants who took part in the workshops and student focus groups. In addition, the authors would like to thank Harriet Bergman for the language and grammar review.

Authors' Contributions

A Mayer, AAK, MS, and IH conceptualized the study. A Mayer and MS facilitated the workshops, which AAK and IH attended as observers, taking field notes. A Mayer and A Müller conducted the qualitative content analysis of the workshop transcripts, developing the coding framework and deriving themes from the data. AAK, MS, and IH reviewed the coding framework and the themes. MS and A Mayer facilitated the student focus groups. A Mayer and A Müller extracted themes from the focus group transcripts. A Mayer and AAK conducted the descriptive analysis of the questionnaire and the concept maps. A Mayer prepared all figures and the first draft of the paper. All authors reviewed and edited the draft and agreed on the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Original images of individual concept maps and consensus versions.
[PNG File, 450 KB - mededu_v11i1e57331_app1.png]

XSL•FO
RenderX

References

1. Huesmann L, Sudacka M, Durning SJ, et al. Clinical reasoning: what do nurses, physicians, and students reason about. J Interprof Care 2023 Nov 2;37(6):990-998. [doi: 10.1080/13561820.2023.2208605] [Medline: 37190790]
2. Vreugdenhil J, Somra S, Ket H, et al. Reasoning like a doctor or like a nurse? A systematic integrative review. Front Med (Lausanne) 2023;10:1017783. [doi: 10.3389/fmed.2023.1017783] [Medline: 36936242]
3. Cuthbert L, duBoulay B, Teather D, Teather B, Sharples M, duBoulay G. Expert/novice differences in diagnostic medical cognition - a review of the literature. : University of Sussex; 1999 URL: https://users.sussex.ac.uk/~bend/papers/csrp508.pdf [accessed 2025-01-09]
4. Durning SJ, Lubarsky S, Torre D, Dory V, Holmboe E. Considering “nonlinearity” across the continuum in medical education assessment: supporting theory, practice, and future research directions. J Contin Educ Health Prof 2015;35(3):232-243. [doi: 10.1002/chp.21298] [Medline: 26378429]
5. Grant J, Marsden P. Primary knowledge, medical education and consultant expertise. Med Educ 1988 May;22(3):173-179. [doi: 10.1111/j.1365-2923.1988.tb00002.x] [Medline: 3405111]
6. Charlin B, Gagnon R, Pelletier J, et al. Assessment of clinical reasoning in the context of uncertainty: the effect of variability within the reference panel. Med Educ 2006 Sep;40(9):848-854. [doi: 10.1111/j.1365-2929.2006.02541.x] [Medline: 16925634]
7. Charlin B, Lubarsky S, Millette B, et al. Clinical reasoning processes: unravelling complexity through graphical representation. Med Educ 2012 May;46(5):454-463. [doi: 10.1111/j.1365-2923.2012.04242.x] [Medline: 22515753]
8. Daley BJ, Torre DM. Concept maps in medical education: an analytical literature review. Med Educ 2010 May;44(5):440-448. [doi: 10.1111/j.1365-2923.2010.03628.x] [Medline: 20374475]
9. Daniel M, Rencic J, Durning SJ, et al. Clinical reasoning assessment methods: a scoping review and practical guidance. Acad Med 2019;94(6):902-912. [doi: 10.1097/ACM.0000000000002618] [Medline: 30720527]
10. Novak J, Cañas A. Theoretical origins of concept maps, how to construct them, and uses in education. Reflect Educ 2007;3 [FREE Full text]
11. Vink S, van Tartwijk J, Verloop N, Gosselink M, Driessen E, Bolk J. The articulation of integration of clinical and basic sciences in concept maps: differences between experienced and resident groups. Adv Health Sci Educ Theory Pract 2016 Aug;21(3):643-657. [doi: 10.1007/s10459-015-9657-2] [Medline: 26692262]
12. Torre DM, Daley B, Stark-Schweitzer T, Siddartha S, Petkova J, Ziebert M. A qualitative evaluation of medical student learning with concept maps. Med Teach 2007 Nov;29(9):949-955. [doi: 10.1080/01421590701689506] [Medline: 18158670]
13. Hay DB. Using concept maps to measure deep, surface and non‐learning outcomes. Stud Higher Educ 2007 Feb;32(1):39-57. [doi: 10.1080/03075070601099432]
14. Torre D, German D, Daley B, Taylor D. Concept mapping: an aid to teaching and learning: AMEE Guide No. 157. Med Teach 2023 May;45(5):455-463. [doi: 10.1080/0142159X.2023.2182176] [Medline: 36862077]
15. Anohina A, Grundspenkis J. Scoring concept maps. Presented at: Proceedings of the International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing; Jun 18-19, 2009; Ruse, Bulgaria.
16. Ruiz-Primo MA, Schultz SE, Li M, Shavelson RJ. Comparison of the reliability and validity of scores from two concept-mapping techniques. J Res Sci Teach 2001 Feb;38(2):260-278. [doi: 10.1002/1098-2736(200102)38:2<260::AID-TEA1005>3.0.CO;2-F]
17. Kinchin IM, Hay DB, Adams A. How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development. Educ Res 2000 Jan;42(1):43-57. [doi: 10.1080/001318800363908]
18. Morse D, Jutras F. Implementing concept-based learning in a large undergraduate classroom. CBE Life Sci Educ 2008;7(2):243-253. [doi: 10.1187/cbe.07-09-0071] [Medline: 18519616]
19. Torre DM, Durning SJ, Daley BJ. Twelve tips for teaching with concept maps in medical education. Med Teach 2013;35(3):201-208. [doi: 10.3109/0142159X.2013.759644] [Medline: 23464896]
20. Cutrer WB, Castro D, Roy KM, Turner TL. Use of an expert concept map as an advance organizer to improve understanding of respiratory failure. Med Teach 2011;33(12):1018-1026. [doi: 10.3109/0142159X.2010.531159] [Medline: 22225439]
21. McGaghie WC, McCrimmon DR, Mitchell G, Thompson JA, Ravitch MM. Quantitative concept mapping in pulmonary physiology: comparison of student and faculty knowledge structures. Adv Physiol Educ 2000 Jun;23(1):72-81. [doi: 10.1152/advances.2000.23.1.S72] [Medline: 10902530]
22. Plackett R, Kassianos AP, Kambouri M, et al. Online patient simulation training to improve clinical reasoning: a feasibility randomised controlled trial. BMC Med Educ 2020 Jul 31;20(1):245. [doi: 10.1186/s12909-020-02168-4] [Medline: 32736583]
23. Dekhtyar M, Park YS, Kalinyak J, et al. Use of a structured approach and virtual simulation practice to improve diagnostic reasoning. Diagnosis (Berl) 2021 Jul 12;9(1):69-76. [doi: 10.1515/dx-2020-0160] [Medline: 34246202]
24. Watari T, Tokuda Y, Owada M, Onigata K. The utility of virtual patient simulations for clinical reasoning education. Int J Environ Res Public Health 2020 Jul 24;17(15):5325. [doi: 10.3390/ijerph17155325] [Medline: 32722097]
25. Plackett R, Kassianos AP, Timmis J, Sheringham J, Schartau P, Kambouri M. Using virtual patients to explore the clinical reasoning skills of medical students: mixed methods study. J Med Internet Res 2021 Jun 4;23(6):e24723. [doi: 10.2196/24723] [Medline: 34085940]

XSL•FO
RenderX

26. Botezatu M, Hult H, Fors UG. Virtual patient simulation: what do students make of it? A focus group study. BMC Med Educ 2010 Dec 4;10(1):91. [doi: 10.1186/1472-6920-10-91] [Medline: 21129220]
27. Kononowicz AA, Woodham L, Georg C, Edelbring S. Virtual patient simulations for health professional education. Cochrane Database Syst Rev 2018;6:CD012194. [doi: 10.1002/14651858.CD012194.pub2]
28. Hege I, Dietl A, Kiesewetter J, Schelling J, Kiesewetter I. How to tell a patient's story? Influence of the case narrative design on the clinical reasoning process in virtual patients. Med Teach 2018 Jul;40(7):736-742. [doi: 10.1080/0142159X.2018.1441985] [Medline: 29490538]
29. Kononowicz AA, Narracott AJ, Manini S, et al. A framework for different levels of integration of computational models into web-based virtual patients. J Med Internet Res 2014 Jan 23;16(1):e23. [doi: 10.2196/jmir.2593] [Medline: 24463466]
30. Edelbring S, Dastmalchi M, Hult H, Lundberg IE, Dahlgren LO. Experiencing virtual patients in clinical learning: a phenomenological study. Adv Health Sci Educ 2011 Aug;16(3):331-345. [doi: 10.1007/s10459-010-9265-0]
31. Hege I, Kononowicz AA, Adler M. A clinical reasoning tool for virtual patients: design-based research study. JMIR Med Educ 2017 Nov 2;3(2):e21. [doi: 10.2196/mededu.8100] [Medline: 29097355]
32. Lang VJ, Kogan J, Berman N, Torre D. The evolving role of online virtual patients in internal medicine clerkship education nationally. Acad Med 2013 Nov;88(11):1713-1718. [doi: 10.1097/ACM.0b013e3182a7f28f] [Medline: 24072116]
33. Hege I, Sudacka M, Kononowicz AA, et al. Adaptation of an international virtual patient collection to the COVID-19 pandemic. GMS J Med Educ 2020;37(7):Doc92. [doi: 10.3205/zma001385] [Medline: 33364371]
34. Mayer A, Da Silva Domingues V, Hege I, et al. Planning a collection of virtual patients to train clinical reasoning: a blueprint representative of the European population. Int J Environ Res Public Health 2022 May 19;19(10):6175. [doi: 10.3390/ijerph19106175] [Medline: 35627711]
35. International Collection of Virtual Patients. URL: http://icovip.eu [accessed 2025-01-09]
36. Project results. DID-ACT. 2022. URL: https://did-act.eu/results [accessed 2025-01-09]
37. Humphrey-Murto S, Varpio L, Gonsalves C, Wood TJ. Using consensus group methods such as Delphi and Nominal Group in medical education research. Med Teach 2017 Jan;39(1):14-19. [doi: 10.1080/0142159X.2017.1245856] [Medline: 27841062]
38. McMillan SS, King M, Tully MP. How to use the Nominal Group and Delphi techniques. Int J Clin Pharm 2016 Jun;38(3):655-662. [doi: 10.1007/s11096-016-0257-x] [Medline: 26846316]
39. Kuckartz U. Qualitative Text Analysis: A Guide to Methods, Practice & Using Software: SAGE; 2002.
40. Rosas SR. Group concept mapping methodology: toward an epistemology of group conceptualization, complexity, and emergence. Qual Quant 2017 May;51(3):1403-1416. [doi: 10.1007/s11135-016-0340-3]
41. Kinchin I, Hay D. Using concept maps to optimize the composition of collaborative student groups: a pilot study. J Adv Nurs 2005 Jul;51(2):182-187. [doi: 10.1111/j.1365-2648.2005.03478.x] [Medline: 15963190]
42. Stoyanov S, Jablokow K, Rosas SR, Wopereis IGJH, Kirschner PA. Concept mapping—an effective method for identifying diversity and congruity in cognitive style. Eval Program Plann 2017 Feb;60:238-244. [doi: 10.1016/j.evalprogplan.2016.08.015]
43. McGaghie WC, Boerger RL, McCrimmon DR, Ravitch MM. Agreement among medical experts about the structure of concepts in pulmonary physiology. Acad Med 1994 Oct;69(10 Suppl):S78-S80. [doi: 10.1097/00001888-199410000-00049] [Medline: 7916837]
44. McGaghie WC, McCrimmon DR, Mitchell G, Thompson JA. Concept mapping in pulmonary physiology using pathfinder scaling. Adv Health Sci Educ Theory Pract 2004;9(3):225-240. [doi: 10.1023/B:AHSE.0000038299.79574.e8]
45. Kononowicz AA, Torre D, Górski S, Nowakowski M, Hege I. The association between quality of connections and diagnostic accuracy in student-generated concept maps for clinical reasoning education with virtual patients. GMS J Med Educ 2023;40(5):Doc61. [doi: 10.3205/zma001643] [Medline: 37881522]
46. Charlin B, Gagnon R, Sauvé E, Coletti M. Composition of the panel of reference for concordance tests: do teaching functions have an impact on examinees' ranks and absolute scores? Med Teach 2007 Feb;29(1):49-53. [doi: 10.1080/01421590601032427] [Medline: 17538834]
47. Radwan A, Abdelnasser A, Elaraby S, Talaat W. Correlation between concept maps and clinical reasoning for final year medical students at the faculty of medicine—Suez Canal University. QJM 2018 Dec 1;111(Suppl 1). [doi: 10.1093/qjmed/hcy200.108]
48. Yue M, Zhang M, Zhang C, Jin C. The effectiveness of concept mapping on development of critical thinking in nursing education: a systematic review and meta-analysis. Nurse Educ Today 2017 May;52:87-94. [doi: 10.1016/j.nedt.2017.02.018] [Medline: 28273528]

Abbreviations

CR: clinical reasoning
DID-ACT: Developing, Implementing, and Disseminating an Adaptive Clinical Reasoning Curriculum for Healthcare Students and Educators
iCoViP: International Collection of Virtual Patients

XSL•FO
RenderX

VP: virtual patient

Edited by B Lesselroth; submitted 23.02.24; peer-reviewed by N Rohani, S Jung; revised version received 18.06.24; accepted 23.11.24; published 30.01.25.

Please cite as:
Mayer A, Hege I, Kononowicz AA, Müller A, Sudacka M
Collaborative Development of Feedback Concept Maps for Virtual Patient–Based Clinical Reasoning Education: Mixed Methods Study
JMIR Med Educ 2025;11:e57331
URL: https://mededu.jmir.org/2025/1/e57331
doi:10.2196/57331

© Anja Mayer, Inga Hege, Andrzej A Kononowicz, Anja Müller, Małgorzata Sudacka. Originally published in JMIR Medical Education (https://mededu.jmir.org), 30.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Multidisciplinary Oncology Education Among Postgraduate Trainees: Systematic Review

Houman Tahmasebi[1,2], MD, MSc; Gary Ko[2], MD; Christine M Lam[1], MSc; Idil Bilgen[3], MDc; Zachary Freeman[4], BSc; Rhea Varghese[1], BHSc; Emma Reel[5], MSW; Marina Englesakis[6], HBA, MLIS; Tulin D Cil[1,2,5], MD, MEd

[1]Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[2]Department of Surgery, University of Toronto, Toronto, ON, Canada
[3]School of Medicine, Koç University, Istanbul, Turkey
[4]Faculty of Science, Wilfrid Laurier University, Waterloo, ON, Canada
[5]Sprott Department of Surgery, Princess Margaret Cancer Centre, University Health Network, 6th floor, 700 University Ave, Toronto, ON, Canada
[6]Library and Information Services, University Health Network, Toronto, ON, Canada

Corresponding Author:
Tulin D Cil, MD, MEd
Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

Abstract

Background: Understanding the roles and patient management approaches of the entire oncology team is imperative for effective communication and optimal cancer treatment. Currently, there is no standard residency or fellowship curriculum to ensure the delivery of fundamental knowledge and skills associated with oncology specialties with which trainees often collaborate.

Objective: This study is a systematic review that aims to evaluate the multidisciplinary oncology education in postgraduate medical training.

Methods: A systematic literature search was performed using MEDLINE, Embase, Cochrane Database of Systematic Reviews, Cochrane CENTRAL, APA PsycINFO, and Education Resources Information Center in July 2021. Updates were performed in February 2023 and October 2024. Original studies reporting the effectiveness of multidisciplinary oncology training among residents and fellows were included.

Results: A total of 6991 studies were screened and 24 were included. Fifteen studies analyzed gaps in existing multidisciplinary training of residents and fellows from numerous fields, including surgical, medical, and radiation oncology; geriatrics; palliative medicine; radiology; and pathology programs. Trainees reported limited teaching and knowledge of oncology outside of their respective fields and endorsed the need for further multidisciplinary oncology training. The remaining 9 studies assessed the effectiveness of educational interventions, including tumor boards, didactic sessions, clinical rotations, and case-based learning. Trainees reported significant improvements in multidisciplinary oncology knowledge and skills following the interventions.

Conclusions: These data suggest postgraduate medical trainees have limited formal multidisciplinary oncology training. Existing educational interventions show promising results in improving trainees' oncology knowledge and skills. There is a need for further research and the development of multidisciplinary oncology curricula for postgraduate medical training programs.

Trial Registration: PROSPERO CRD42022271308; https://www.crd.york.ac.uk/PROSPERO/view/CRD42022271308

(JMIR Med Educ 2025;11:e63655) doi:10.2196/63655

KEYWORDS

multidisciplinary; oncology; postgraduate medical education; resident; fellow; surgery; hematology; radiation oncology; geriatrics; palliative

Introduction

Cancer was the second leading cause of death in the United States in 2023 [1]. Cancer care often requires a team of physicians including surgical, medical, and radiation oncologists, as well as specialists in radiology and pathology [2]. Knowledge of collaborating oncologists' roles and appropriate multidisciplinary referrals may impact cancer treatment. There is evidence of improved adherence to standard treatment guidelines with multidisciplinary referrals for patients with prostate [3], lung cancer [4], and bladder cancer [5].

There is considerable potential to improve interdisciplinary communication between various oncologic specialists and to optimize psychosocial support for patient care. Therapies with different oncologists must be well coordinated and specifically selected based on the medical and social needs of each patient.

XSL•FO
RenderX

To achieve this, knowledge of other disciplines' roles, responsibilities, and treatment options is necessary for effective communication and optimal cancer care.

There is currently no standard curriculum for delivering multidisciplinary oncology education in residency and fellowship programs in the United States [6-10]. Mattes et al [11] identified that while many of the program requirements for oncology subspecialties emphasize the importance of providing multidisciplinary cancer care, how this occurs varies widely between subspecialties. Not all programs mandate multidisciplinary oncology rotations or experiential specialty training, and only a select few require attendance at multidisciplinary tumor board meetings (MTBM) [6-12]. Such a training gap may impact trainee education and, as a result, influence referral patterns and the timely access of patients to multimodal cancer therapies.

The objective of this study was to perform a systematic review of the literature to evaluate the multidisciplinary oncology education in postgraduate medical training (ie, interns, residents, and fellows). This study provides a review of literature analyzing the education of learners about the role of any collaborating physician specialty involved in oncology care, including but not limited to, medical oncology, radiation oncology, surgical oncology, and palliative care. These data summarize gaps in training programs identified across studies, the suggested educational interventions to bridge these gaps, and limitations in the literature within the field.

Methods

Research Design and Methodology

This systematic review was reported based on PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [13]. The protocol was registered and published by PROSPERO (ID: CRD42022271308).

Search Strategy

A search strategy was developed with the assistance of an information specialist using these and other related terms: "Residents or Fellows or Trainees or Medical Training" AND "Education or Training Programs" AND "Multidisciplinary" AND "Oncology." The following databases were searched from inception: MEDLINE, MEDLINE In-Process, Embase Classic + Embase, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, and APA PsycINFO (all via the Ovid platform); and Education Resources Information Center via the EbscoHost platform. The search was initially performed on July 21, 2021, and updated twice (ie, on February 26, 2023, and October 9, 2024). Table S1 in Multimedia Appendix 1 shows the number of citations identified from each database. The search strategy and the number of citations identified via MEDLINE are included in Table S2 in Multimedia Appendix 1.

Eligibility Criteria

Eligibility criteria were developed prior to the search strategy. The scope of this study was to evaluate the multidisciplinary oncology education offered by residency and fellowship programs to postgraduate medical trainees. Thus, the first eligibility criterion was the inclusion of studies investigating postgraduate medical training (ie, interns, residents, and fellows). Studies about nonphysician specialties (eg, nursing, pharmacy, or dentistry), attending or staff physicians, or those involving solely Masters, PhD, or medical students were excluded. Studies were included if their focus was specific to oncology care. Selected studies focused on multidisciplinary aspects of medical education, which included knowledge of collaborating medical specialties and their roles in cancer care (eg, surgical trainees' knowledge of radiation or medical treatments). Trainees from all specialties were included, as long as the study was assessing the multidisciplinary oncology education of trainees, and therefore, these were not necessarily restricted to oncology residency or fellowship programs (eg, medical oncology, radiation oncology, surgical oncology). Only primary research papers and studies available in English (ie, both original and translations to English) were included. Thus, all reviews, case studies, opinion papers, abstract-only papers, conference literature, and short reports were excluded.

Study Selection

There were 2 stages of review: title and abstract screening, followed by full-text screening. A total of 6 reviewers (HT, GK, CML, IB, ZF, and RV) were involved, and studies were screened by a minimum of 2 independent reviewers at each stage. Discrepancies were resolved by a third reviewer. Both screening stages were performed on Covidence [14], a web-based systematic review organization software.

Data Extraction and Synthesis

Data extraction was performed on the selected studies. Studies were divided between 3 reviewers (HT, CML, and RV) who performed data extraction. Study design, study population, outcome measures, and main results were extracted from each study.

Quality Assessment

Selected studies were independently assessed for quality by 2 independent reviewers (CL, IB, and RV) using the Mixed Methods Appraisal Tool (MMAT) version 2018 [15]. Discrepancies were resolved through discussion with a third author (HT). The MMAT was chosen due to its ability to concomitantly assess multiple study types (ie, qualitative, quantitative randomized controlled trial, quantitative nonrandomized, quantitative descriptive, or mixed methods). Each study was evaluated on a set of 5 criteria depending on the study type. For survey studies, the risk of nonresponse bias was deemed to be high if the response rate was below 70%. Studies were assigned an overall quality score ranging from 0 to 5 stars based on the number of criteria that were met.

Results

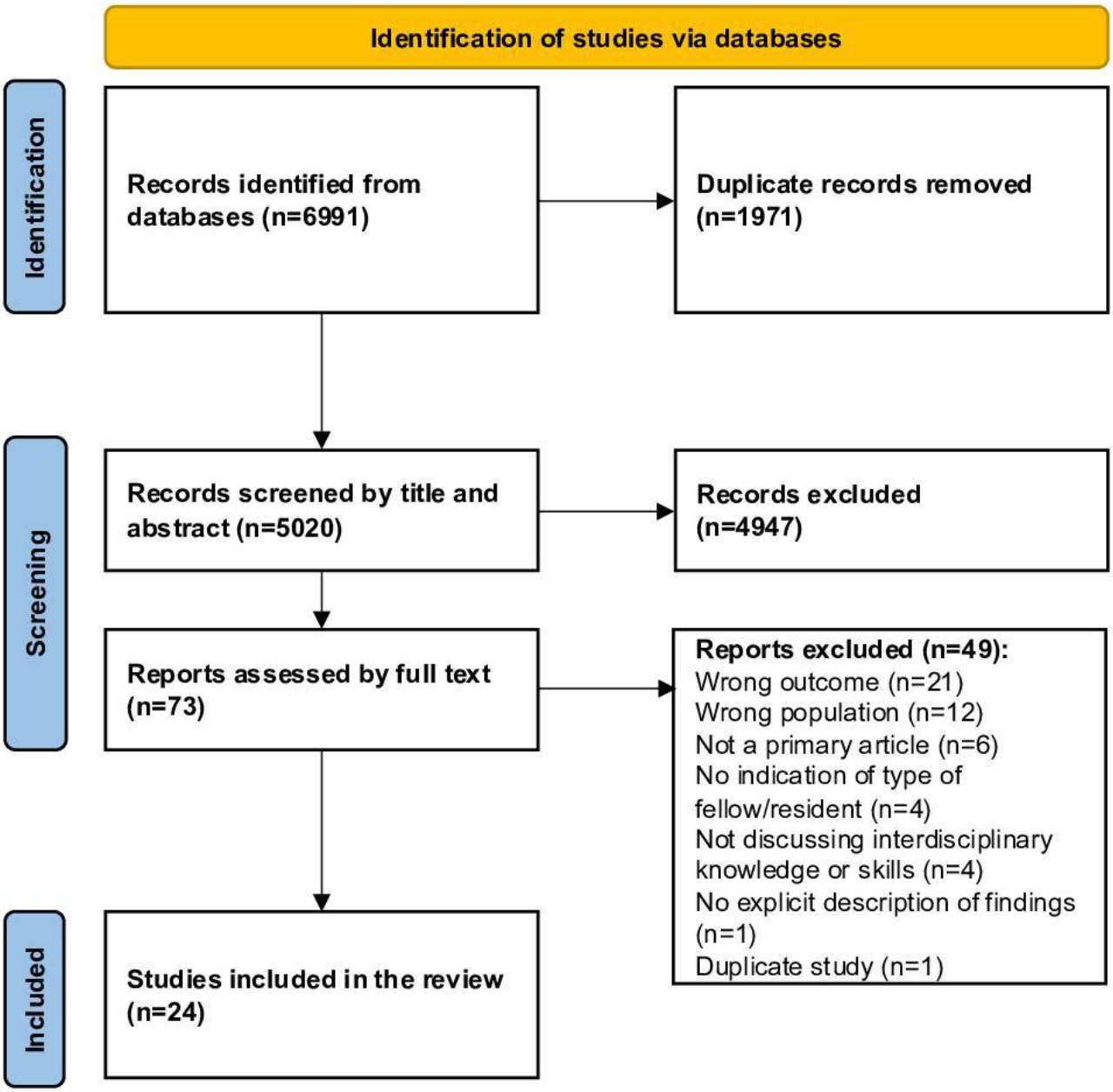
Study Characteristics

The search strategy resulted in a total of 6991 studies. After removing duplicates between databases, 5020 unique studies were identified. A total of 73 studies remained after title and abstract screening. Full-text screening excluded 49 studies, and

XSL•FO
RenderX

24 studies were therefore included in the final analysis. The PRISMA flow diagram is demonstrated in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram of the systematic review. Adapted from Page et al [13].



The remaining 24 studies were divided into 2 categories. Fifteen studies assessed the quality of existing postgraduate oncology training based on trainees' multidisciplinary knowledge. The remaining 9 assessed trainees' multidisciplinary knowledge following an educational intervention. For the latter category, all studies with educational interventions directed toward improving multidisciplinary oncology knowledge and skills among interns, residents, and fellows were included. These included studies that are part of the formal postgraduate medical training (eg, residency or fellowship program), as well as external initiatives for improving multidisciplinary oncology training. Studies involving educational interventions for medical students and staff or attending physicians were not included.

Existing Multidisciplinary Training

A summary of the 15 studies evaluating the impact of existing multidisciplinary oncology training is included in Table 1. These studies included surgical or surgical oncology fields [16-23], hematology or medical or hematology oncology [16,20-22,24,25], geriatrics or geriatric oncology [20,22,26], radiation oncology [16,20,21,23,27], palliative medicine [16,20], radiology [21], pathology [21], genetics [23], dermatology [23], pediatric specialties [28], and other medical fields (eg, internal medicine, nephrology, neurology) [22,23].

XSL•FO
RenderX

Table . Summary of studies evaluating the existing multidisciplinary education across postgraduate medical training programs.

Reference	Study design	Sample	Outcome measure	Main findings and conclusions
Akthar et al [16]	Electronic surveys completed by oncology trainees and program directors across the United States in 2013	• 557 hematology or medical oncology, surgical oncology, radiation oncology, and palliative medicine residents and fellows • 141 hematology or medical oncology, surgical oncology, radiation oncology, and palliative medicine program directors	Proportion of trainees who received formal education in oncology fields outside of their specialty	• Generally limited interdisciplinary oncology education: ≤70% of trainees reported receiving formal interdisciplinary education; highest rate of training in radiation oncology (70% of trainees) and lowest rate of training in geriatric oncology (19% of trainees) • Consistently lower rates of interdisciplinary oncology training reported by trainees compared to program directors (P<.01)
Brenner and De Donno [17]	Survey of postgraduate year 1 - 5 residents from 3 general surgery programs: Florida Atlantic University, The University of Iowa, and The University of Connecticut	• 135 general surgery residents	Proportion of residents who indicated receiving training in a specific multidisciplinary field	• Limited proportion of residents indicated receiving multidisciplinary training: radiation oncology (23%), chemotherapy (31%), and palliative medicine (53%) • Majority (82%) of residents endorsed further multidisciplinary training
David et al [25]	Survey of hematology residents (ie, postgraduate years 4 - 5) or fellows (ie, postgraduate year 6 - 7) across Canada as part of a cross-sectional study	• 29 hematology residents and 3 hematology fellows	Geriatric oncology curriculum needs assessment	• 56.3% did not receive geriatric oncology teaching • 96.9% endorsed the inclusion of geriatric training in hematology residency
Delaye et al [22]	Surveys completed by French residents and senior physicians regarding the field of onco-nephrology	• Residents (n=130) and senior physicians (n=98) from nephrology, oncology, hematology, surgery and geriatrics	Current practices in onco-nephrology, information resources, existing cooperation networks, and expectations about onco-nephrology	• Oncology residents rated their confidence in facing renal events as 5.5/10 • Nephrology residents rated their confidence in facing cancer events as 6.0/10 • 21% of residents had received onco-nephrology teaching, which was judged as insufficient
Eid et al [24]	Review of literature, expert consultation, review of fellows' rotation evaluations, and interviews with current and recently graduated fellows, as a means of needs assessment for the development of a geriatric oncology program at MD Anderson Cancer Center	• 9 current hematology-oncology fellows (years 1 - 3) at MD Anderson Cancer Center • 2 MD Anderson faculty members who recently graduated from a hematology-oncology fellowship	Geriatric oncology training program needs assessment	Top 3 identified needs for geriatric oncology programs, based on current educational gaps: • Geriatric assessment • Pharmacology knowledge • Psychosocial knowledge

XSL•FO
RenderX

Reference	Study design	Sample	Outcome measure	Main findings and conclusions
Givi et al [29]	Semistructured interviews with program directors and faculty in head and neck surgery across the United States and Canada over a 7-month period	• 58 participants including head and neck surgery program directors and faculty	Head and neck surgical oncology training needs assessment	• 38% endorsed increasing the number of head and neck surgery fellows' interactions with medical oncology, radiation oncology, and speech and language pathology • 85% view exposure to multidisciplinary teams as essential in training curricula
Le Nail and Samargandi [30]	Web-based questionnaire on various aspects of MTBMs[a] completed by French orthopedic oncology residents	• 27 orthopedic oncology residents	Residents' opinions on educational impact and areas of improvement for MTBMs	• 54% agreed that MTBM is an appropriate venue for teaching • 75% endorsed that MTBMs improved their knowledge of other specialties involved • 71% indicated opportunities to improve teaching during MTBM, the most popular suggestion being active participation of residents (voted by 46% of all residents)
Maggiore et al [26]	Web-based survey completed by program directors of geriatrics fellowship programs in the United States	• 67 geriatrics program directors	Proportion of program directors offering or endorsing future learning opportunities in the field of geriatric oncology	• Majority (81%) of program directors offered didactic teaching in the form of formal geriatric oncology lectures/seminars • Limited number of program directors offered clinical experience: 39% offered mandatory oncology clinical experience and 46% offered clinical electives • Majority (77%) endorsed oncology training as part of the geriatrics fellowship
Mäurer et al [23]	Web-based survey distributed to all junior oncology groups represented in Young Oncologists United in Germany regarding interdisciplinarity in oncology	• 294 participants including 268 physicians (staff and trainees) from internal medicine, gynecology, radiotherapy and radiation oncology, general surgery, genetics, neurosurgery, urology, neurology, and dermatology	Opinions on interdisciplinarity at clinic, educational, and research levels	• 63.1% assigned a high priority to interdisciplinary residency training • Only 18.3% had the opportunity to participate in rotations in other specialties beyond their curriculum • 71.4% were interested in participating in rotations in other specialties • 73.1% of those who completed interdisciplinary rotations benefited from them

XSL•FO
RenderX

Reference	Study design	Sample	Outcome measure	Main findings and conclusions
Morris et al [27]	Web-based survey completed by radiation oncology residents in Australia, New Zealand, and Singapore	• 61 radiation oncology residents	Proportion of residents who indicated receiving or endorsing future geriatric oncology training	• Majority (91.8%) did not receive any geriatric training • Limited number of residents (39.3%) comfortable managing complex geriatric issues • Majority (85.3%) endorsed additional geriatric training
Morris et al [20]	2-stage Delphi consensus with input from a panel of internationally recognized oncology experts, staff physicians, radiation oncology and clinical oncology trainees, allied health professionals, patients, and caregivers. Experts were from geriatrics, geriatric oncology, and radiation oncology. Staff physicians were from clinical/medical oncology, palliative care, and surgical oncology.	• A total of 103 and 54 individuals participated in rounds 1 and 2 of the modified Delphi consensus process, respectively • Majority were radiation oncologists (43%)	Establishing learning outcomes for a geriatric radiation oncology curriculum	• 33 learning outcomes identified in the areas of fundamental geriatric medicine concepts, epidemiology, geriatric screening, planning and delivery of radiation therapy, geriatric palliative care, surgery, systematic treatment, research, communication skills, and health advocacy
Park et al [18]	30 item self-efficacy survey completed by residents at Ohio State University Wexner Medical Center, in order to measure knowledge and skills in 6 breast cancer care aspects: genetics, surgery, medical oncology, radiation oncology, pathology, and radiology	• 31 general surgery residents	Residents' perceived capability (ie, self-efficacy score) in various domains of breast cancer care	• Highest self-efficacy in surgery (3.56/5) vs lowest in genetics (2.67/5), radiation oncology (2.67/5), and pathology (2.67/5) • Significant improvement of self-efficacy in surgery only (P=.002) with additional years in residency
Picca and Reed [28]	Semistructured interviews with faculty and trainees across pediatric oncology, radiology, pathology, surgical oncology, and palliative care	• 4 pediatric oncology fellows • 11 pediatric oncology, pathology, radiology, palliative care, and surgical oncology faculty physicians	Exploration of learning in tumor boards	• Trainees found tumor board presentation to be educational • Barriers to learning: competing clinical/administrative responsibilities • Facilitators to learning: learning-focused goals, faculty mentorship during presentation preparation, collaborative discussion, content tailored to learners and board exams, and supportive environment • Web-based tumor boards promoted accessibility and convenience but decreased learning due to limited engagement, discussion, and professional relationship development

XSL•FO
RenderX

Reference	Study design	Sample	Outcome measure	Main findings and conclusions
Walraven et al [21]	Semistructured interviews with Dutch residents and specialists in medical/surgical/radiation oncology, radiology, nuclear radiology, and pathology participating in MDTMs[b]	• 19 residents • 16 specialists	Residents' barriers and facilitators to participate in MDTMs	• 100% agreed that MDTMs play an important role in both education and patient care • Barriers: insufficient supervisor guidance, time constraints, meeting atmosphere and hierarchy, strict regulations, unfamiliarity, and resident's personal characteristics • Solutions: MDTM simulation training, and training courses on communication and meeting skills
Wilson et al [19]	Survey of applicants to Roswell Park Cancer Institute surgical oncology fellowship program	• 29 general surgery residents or recent general surgery graduates applying to surgical oncology fellowship	Proportion of applicants with breast surgery exposure and their comfort with medical and surgical management of breast cancer	• Majority (65%) had exposure to multidisciplinary breast cancer clinics, involving medical and surgical oncologists • Lower level of comfort (7.07/10) with breast cancer medical management compared to surgical management (7.34 - 9.10/10 depending on the type of surgery)

[a]MTBM: multidisciplinary tumor board meeting.

[b]MDTM: multidisciplinary team meetings.

Thirteen studies obtained opinions of trainees with respect to multidisciplinary oncology education within their training programs [16-25,27,28,30]. Morris et al [20] used a Delphi consensus process, and 4 studies directly interviewed trainees and faculty [21,24,28,29]. The remainder of the studies used surveys. Maggiore et al [26] surveyed geriatrics program directors, Givi et al [29] surveyed head and neck surgery program directors, and Akthar et al [16] surveyed program directors of pediatric and adult hematology oncology, surgical oncology, radiation oncology, and palliative medicine. Eid et al [24] used a combination of expert consultation, trainee interviews, review of trainee rotation evaluations, and literature review to assess their multidisciplinary educational needs.

While all studies analyzed the quality of existing multidisciplinary education, there were differences in the disciplines investigated across studies. Akthar et al [16], Delaye et al [22], Mäurer et al [23], Walraven et al [21], Picca and Reed [28], and Brenner and De Donno [17] focused on identifying broad gaps in multidisciplinary education including knowledge and skills of trainees in numerous fields, such as radiation, surgical, and medical oncology, radiology, pathology, geriatrics, palliative medicine, and other pediatric and medical fields. The remaining 8 studies focused on a more specific set of trainee skills. David et al [25], Eid et al [24], and Maggiore et al [26] assessed gaps in geriatric oncology education among hematology residents and fellows, hematology oncology fellows, and geriatrics fellows, respectively. Morris et al [20,27] assessed gaps in the radiation oncology training curriculum. Park et al [18] and Wilson et al [19] assessed the quality of general surgery residency training in breast cancer care. Le Nail and Samargandi [30] evaluated the quality of tumor boards for orthopedic oncology trainees. Finally, Givi et al [29] performed a needs assessment analysis of the head and neck surgery training curriculum.

13 studies assessed the strengths and weaknesses of oncology training programs [16-19,21-23,25-30]. Of these, 11 found that trainees had limited exposure to multidisciplinary oncology disciplines, barriers to attending multidisciplinary oncology meetings, and a low level of trainee comfort in multidisciplinary oncology knowledge [16-19,21-23,25-28]. Givi et al [29] found that 27% of interviewees indicated exposure to multidisciplinary care as a strength of the head and neck surgery training program, although 38% endorsed the need to improve fellows' multidisciplinary participation. In general, Akthar et al [16] found the least amount of multidisciplinary training in geriatric oncology, compared to palliative medicine, medical, radiation, and surgical oncology. Similarly, Morris et al [27] found that less than 10% of radiation oncology trainees received geriatrics training. Furthermore, less than half of geriatrics fellows were offered geriatric oncology rotations [26]. For multidisciplinary

XSL•FO
RenderX

breast cancer management, Park et al [18] found limited training in genetics, radiation oncology, and pathology among surgical residents, compared to rotations within surgery, radiology, and medical oncology. Brenner and De Donno [17] found that a small proportion of general surgery residents received training in the fields of radiation (23%) and medical oncology (31%), but over half (53%) received exposure to palliative care.

Additionally, 11 studies researched areas of improvement for multidisciplinary oncology education among the postgraduate programs via surveys, interviews, Delphi consensus, and literature search [17,20,21,23-30]. Maggiore et al [26] and Morris et al [27] found that 77% of geriatrics fellows and 85.3% of radiation oncology residents advocated for further geriatric oncology training. David et al [25] found that over 95% of hematology trainees endorsed geriatric training during residency. 82% of general surgery residents surveyed by Brenner and De Donno [17] agreed that additional multidisciplinary training is needed to optimize cancer care. Additionally, based on an educational needs assessment, Eid et al [24] found that the top 3 priorities for a geriatric oncology program included geriatric assessment, pharmacology, and psychosocial skills. MTBMs were found to enhance trainee experience and multidisciplinary oncology education [21,28,30]. However, some barriers to attending meetings included time constraints, clinical duties, and lack of active resident participation [21,28,30]. Residents and specialists interviewed by Walraven et al [21] suggested that the educational value of multidisciplinary team meetings could be improved through additional training such as multidisciplinary team meeting simulations and courses on effective communication and meeting skills.

Impact of Educational Interventions

A summary of the 9 studies analyzing the impact of educational interventions is included in Table 2. The majority included general surgery trainees [31-35]. Faculty and trainees from radiation oncology [32,35], medical oncology [12,35], respirology [12,36], thoracic surgery [12], gynecology [35], urology [37], and palliative medicine [38] were also included. All 9 studies demonstrated improvements in multidisciplinary oncology knowledge and skills postintervention.

XSL•FO
RenderX

Table . Summary of studies evaluating the impact of multidisciplinary educational interventions.[a]

Reference	Study design	Sample	Outcome measure	Main findings and conclusions
Cook et al [31]	Electronic surveys were sent to general surgery residents at the completion of 4-week rotations in MDB[b], USOS[c], and community-based TSR[d] at Oregon Health and Science University in 2010 - 2013. MDB included operative time, as well as half-days in pathology, radiology, medical oncology, and surgery clinic.	• Total sample size: 32 in MDB, 73 in USOS, and 51 in TSR • Operative logs of 29 residents in MDB, 11 in TSR, and 12 in USOS were obtained	• Trainee satisfaction based on surveys • Operative volume based on operative logs	• MDB rotation residents rated the opportunity to perform and learn procedures higher than those in USOS (*P*=.02) and TSR (*P*=.01) • 83% of MDB residents' operative experience included breast cancer operations, compared to 71% of USOS and 12% of TSR groups • MDB rotation residents rated higher on the quality of faculty teaching and educational materials than those on TSR (*P*=.03 and *P*=.04, respectively)
Khoshgoftar et al [37]	Short interviews were held with urology residents and faculty members regarding needs for holding web-based tumor boards prior to implementation of 20 monthly web-based tumor boards. Tumor boards were assessed through questionnaires postintervention, resident pretest and posttest scores for 5 consecutive tumor boards, and external evaluators from the faculty of urology.	• 35 urology residents • 25 urology faculty members • Panelists from pathology, radiation oncology, medical oncology, radiology, and nuclear medicine	• Needs assessment, satisfaction levels, pretest and posttest scores, recommendations from external evaluators	• Resident needs assessment was divided by level of importance and postgraduate years (ie, years 1 - 2 vs 3 - 4). An important limitation to participate was significant clinical responsibilities, particularly for lower year residents • High resident satisfaction rate (71% - 88%) based on various aspects of web-based tumor boards. The most important technical issue was the low bandwidth speed. • There was significant improvement in resident posttest scores in the majority of sessions
Mackay et al [36]	Respiratory and oncology trainees completed a 3-hour MDTM[e] simulation session and completed pre- and postsimulation questionnaires	• 19 oncology and respiratory trainees (specialty training years 3 - 7)	• Perceptions of current training programs, confidence presenting in MDTMs, use of the simulation, and impact on future clinical practice	• Trainees rated 4/10 for how well their program prepared them to present at MDTM • Trainee confidence in presenting in MDTMs increased from 5/10 to 7/10 postintervention (*P*<.01) • Trainees rated 9/10 for usefulness and 9/10 for likelihood the session will lead to changes in their practice

XSL•FO
RenderX

Reference	Study design	Sample	Outcome measure	Main findings and conclusions
Martin et al [38]	Fellows completed three 1-hour lectures in palliative radiotherapy, as well as pre- and postcourse questionnaires and objective knowledge assessment multiple-choice questions.	• 5 hospice and palliative medicine fellows at the University of California, San Diego	• Knowledge and confidence in palliative radiotherapy	• Postintervention improvement in trainee-reported confidence in discussion with patients about radiotherapy (0.009), managing its common side effects (*P*=.021), and identifying oncologic emergencies related to radiotherapy (*P*=.012) • Significant improvement in radiotherapy knowledge based on objective knowledge assessment questions (22% vs 86% pre- vs postintervention; *P*=.010) • Increased trainee-reported likelihood of collaboration with radiation oncologists postintervention (*P*=.014)
Mattes et al [12]	Faculty, fellows, and residents attended a didactic lecture on radiation therapy in lung cancer care. Knowledge was tested using multiple choice questions pre- and postintervention.	• A total of 121 faculty and trainees from pulmonology, thoracic surgery, and medical oncology • Pretest: 54 residents/fellows and 9 faculty participated • Posttest: 23 residents/fellows and 2 faculty participated	• Knowledge of radiation therapy in lung cancer treatment and comfort in appropriate referral to radiation oncology	• The majority had no didactic training (75%) or rotations (85.5%) in radiation oncology preintervention • Significant improvements in mean objective test scores postintervention (*P*<.001) • Postintervention, 100% of participants felt more knowledgeable in radiation therapy and 96% felt more comfortable making appropriate radiation oncology referrals
Meani et al [35]	Faculty and trainees completed a postintervention questionnaire following a multidisciplinary breast cancer course.	• A total of 42 participants in medical oncology, radiation oncology, gynecology, and general surgery • 11 heads of department/professors • 17 consultants/attending Physicians • 14 trainees: residents, medical fellows, PhD students, and postdoctoral fellows	• Opinions on the impact of the course	• Postintervention, 64% made changes in their clinical practice and 33% made institutional changes in breast cancer management • 95% reported increased knowledge of MDB cancer care
Sloan et al [32]		• 22 general surgery residents • 3 radiation oncology residents • 15 faculty at stations • 12 patients with breast cancer at stations	• Self-reported trainee improvement in breast cancer care–specific skills • Perception of faculty, patients, and residents of the overall quality of intervention	

XSL•FO
RenderX

Reference	Study design	Sample	Outcome measure	Main findings and conclusions
	Residents at the University of Kentucky received multidisciplinary instruction and completed 15 case-based stations about various domains of breast cancer care (ie, surgical oncology, medical oncology, radiology, radiation oncology, plastic surgery, and pathology). Surveys about the overall quality of intervention were completed by patients, faculty, and residents. Residents also completed pre- and postintervention surveys regarding specific breast cancer care-specific skills.			• Statistically significant trainee-reported improvement for all measured skills, including fine-needle aspiration, mammography interpretation, and treatment discussion with patients • Overall, intervention rated favorably by trainees, faculty, and patients
Sloan et al [33]	Residents at the University of Kentucky completed 12 case-based stations during a head and neck oncology workshop, designed by faculty from general surgery, speech pathology, dentistry, radiation therapy, otolaryngology, plastic and reconstructive surgery, pathology, anesthesiology, and cardiothoracic surgery. Surveys about the overall quality of intervention were completed by patients, faculty, and residents. Residents also completed pre- and postintervention surveys regarding head and neck-specific skills.	• 21 general surgery residents • 11 faculty at stations • 8 standardized patients at stations (including 6 patients with cancer)	• Self-reported trainee improvement in skills relevant to head and neck cancer care • Perception of faculty, patients, and residents of the overall quality of intervention	• Statistically significant trainee-reported improvement for most skills postintervention (P<.001) • Overall, intervention rated favorably by trainees, faculty, and patients • Residents generally endorsed having intervention minimum twice during residency
Sloan et al [34]	2 groups received multidisciplinary teaching in breast cancer care, including radiation oncology, radiology, surgery, and medical oncology, in the form of a 15-station workshop. The other 2 groups served as controls. 1 intervention and 1 control group were administered an 11-problem OSCE[f] assessment immediately postintervention and the remaining 2 groups were administered the same OSCE assessment 8 months later. Residents were assessed by faculty and standardized patients during OSCE assessments.	• 48 general surgery residents from the University of Kentucky, divided evenly into 4 groups • 15 faculty at stations • 12 standardized patients at stations (including 5 patients with cancer)	• Skills in diagnosis and management of breast cancer postintervention, assessed by faculty and standardized patients during OSCE assessments	• Improvement in skills of residents who attended the workshop, compared to the control group, both immediately and 8 months postintervention (P<.01) • Residents' skills diminished after 8 months, as evidence by the difference in skill set between the group tested immediately versus the one tested 8 months postintervention (P<.004)

[a]Patients who performed assessments included actual and simulated patients.

[b]MDB: multidisciplinary breast.

[c]USOS: university surgical oncology service.

[d]TSR: traditional surgical rotation.

[e]MDTM: multidisciplinary team meeting.

[f]OSCE: Objective Structured Clinical Examination.

XSL•FO
RenderX

The study by Cook et al [31] compared the impact of a multidisciplinary breast rotation to traditional oncology or community rotations using trainee self-evaluations. Martin et al [38] and Mattes et al [12] analyzed the effectiveness of didactic learning for palliative radiotherapy and lung cancer radiotherapy, respectively, using pre- and postcourse trainee evaluations. Meani et al [35] studied the impact of a multidisciplinary breast cancer course on the knowledge and practice of faculty and trainees using a questionnaire. Three studies by Sloan et al tested the quality of case-based instruction, involving workshops or Objective Structured Clinical Examination (OSCE) stations, where evaluations were completed by trainees, standardized patients, and faculty [32-34]. In the 2004 study by Sloan et al [34], faculty and standardized patient completed evaluations following the observation of trainees in OSCE stations. Patient ratings mainly included interpersonal skills, while faculty ratings included both the clinical and interpersonal skills of trainees. In the other 2 Sloan et al studies, faculty and standardized patients provided feedback on the overall quality of workshops, rather than a specific focus on trainee skills [32,33]. Many of the standardized patients were actual patients with cancer [32-34]. Two of the Sloan et al studies with breast cancer-specific stations focused on knowledge and skills in the following fields: surgical, medical, and radiation oncology; pathology; plastic surgery; and radiology [32,34]. A pilot study by the same group included a head and neck workshop in which stations were designed by faculty from general surgery, radiation oncology, cardiothoracic surgery, otolaryngology, plastic surgery, pathology, anesthesiology, speech pathology, and dentistry [33].

In 8 of these 9 studies, the benefit of educational interventions was noted by the trainees through self-assessment of knowledge or skills [12,31-33,35-38], while Sloan et al [34] demonstrated improvements in knowledge or skills, as assessed by faculty and patients following the observation of trainees in OSCE stations. In addition to reporting subjective benefits, Khoshgoftar et al [37], Mattes et al [12], and Martin et al [38] used objective assessments to demonstrate improvements in trainee knowledge postintervention. Interestingly, Sloan et al [34] showed that while the intervention benefited residents' knowledge and skill set in breast cancer management both immediately after and 8 months postintervention, it declined after 8 months. In the other 2 studies by this group [32,33], trainees, faculty, and patients rated the interventions highly.

Quality Assessment

A summary of the MMAT quality assessment is included in Table S3 in Multimedia Appendix 1. Five studies were categorized as nonrandomized, 4 as qualitative, 13 as quantitative descriptive, 1 as mixed methods, and 1 as randomized controlled. Studies were given a score out of 5, based on the number of MMAT criteria met. Two studies were given an overall MMAT quality rating of 3 stars, 14 studies were rated as 4 stars, and the remaining 8 were rated as 5 stars. Overall, all studies were deemed to be satisfactory by authors, based on MMAT quality assessment criteria.

Discussion

Principal Results

To our knowledge, this is the first systematic review of multidisciplinary oncology education in postgraduate medical training. These data summarize educational gaps and potential solutions to improve multidisciplinary education for future trainees. Of the 24 studies included in the final analysis, 15 obtained faculties' and trainees' opinions on deficiencies and areas of improvement for existing multidisciplinary oncology education [16-19,24,26,27]. They generally reported limited multidisciplinary oncology training or knowledge, barriers to multidisciplinary training, and advocated for further instruction in different areas. The remaining 9 studies studied the impact of educational interventions on trainees' oncology expertise [31-34,38]. Multidisciplinary rotations, tumor board meetings, didactic teaching, and case-based learning were found to be beneficial based on trainee self-assessments, written exams, and evaluations from faculty and patients following the observation of trainees in OSCE stations.

Filling the current gaps in multidisciplinary oncology education using the aforementioned educational interventions has the potential to improve multidisciplinary communication, appropriate referrals, and oncologic outcomes [3-5]. Studies by Mattes et al [12] and Martin et al [38] found that trainees were more likely to collaborate and make appropriate referrals to radiation oncologists after didactic teachings in lung cancer treatment and palliative radiotherapy, respectively. Several studies also found MTBMs to enhance trainee education [30,36,37]. In fact, the study by Mackay et al [36] found that tumor board simulation sessions significantly improved trainee's confidence in presenting in tumor board sessions. After all, improved communication and referral patterns are central to effective multidisciplinary collaboration among oncology specialists and ultimately improve the access of patients to evidence-based oncologic treatments.

Comparison With Prior Work

Geriatric oncology was consistently found to be an area in which trainees received limited training [16,26,27,39]. As cancer incidence increases in older adults, a population with a higher burden of comorbidities, trainees must gain sufficient knowledge and experience in geriatric oncology to optimize treatment [40]. These findings are echoed in a review by Morris et al [39] highlighting insufficient training and education in geriatric oncology among radiation oncology trainees across several different countries. This training should identify the specific needs of older patients and thereby result in a more informed and nuanced approach to this population's medical and psychosocial issues [24]. Development of these skills may be achieved through dedicated rotations or training in geriatric oncology.

Based on findings from this study, it is evident that the quality of multidisciplinary oncology education and training needs to be assessed and addressed. Implementation of benchmarks to ensure sufficient training across residency and fellowship programs commonly involved in cancer care would provide an educational quality metric [6-10]. This would encourage training

XSL•FO
RenderX

programs to develop and establish multidisciplinary oncology curricula. One approach to achieve this would be to ensure trainee participation in a variety of educational activities such as multidisciplinary case conferences, research, rotations, didactic teaching, and case-based learning led by faculty from other disciplines [11,31-34,38]. Furthermore, a review of each residency or fellowship program's curriculum by a multidisciplinary faculty committee may ensure sufficient trainee exposure to collaborating oncology areas.

Competency-based medical education is an outcome-based approach to evaluate medical trainees and ensure a high degree of graduate skill set [41]. This is often done via objective measures, such as entrustable professional activities (EPAs) and milestones. The development of standardized and program-specific EPAs, specifically for multidisciplinary oncology education, would provide training programs with a specific measure of their trainees' knowledge, skills, and progress in this area. Using EPAs would also identify areas of improvement for trainees early on in their training and would allow for additional support to improve multidisciplinary oncology competencies. Ultimately, these EPAs should mirror curriculum changes to ensure effective multidisciplinary oncology education. The benefits of using EPAs for geriatric oncology training are echoed by Eid et al [24]. They provide an example of an EPA to assess the appropriateness of chemotherapy for a geriatric patient, which includes the ability to perform a comprehensive geriatric assessment, having sufficient knowledge of chemotherapy toxicities and interactions, and assessment of suitability based on patients' comorbidities. This represents a geriatric oncology-specific EPA for medical or hematology oncology trainees. Oncology training programs may adopt similar EPAs to ensure a high quality of multidisciplinary oncology training within their residency and fellowship programs.

Despite its merits, there are potential barriers to the implementation of oncology training curricula. Several factors may prevent trainee participation in multidisciplinary education activities, including limited elective time, educational options, or available personnel. For instance, those training in the community or rural hospitals may not have access to many electives in other oncology fields. For the same reason, there may be limited available multidisciplinary faculty to either design effective oncology curricula or mentor trainees. Furthermore, many residency or fellowship programs may have strict curricula and elective requirements, and thus limit elective options for trainees. To overcome some of these challenges, studies have suggested the importance of web-based courses or teaching sessions to supplement their curriculum. As a result of the COVID-19 pandemic, web-based education has become an integral part of medical training that will likely remain used to various degrees in the future [42,43]. Data supports the effectiveness of web-based training, including web-based rotations or clinical training [44-46], tumor board meetings [28,37], surgical skills training [47], and didactic and case-based teaching [48-52].

Furthermore, local, state-wide or provincial, and national resources and programs could also be offered to trainees interested in further advancing their multidisciplinary oncology knowledge and skills outside their residency and fellowship programs. Certainly, didactic teaching [12,35,38], as well as workshops and OSCE-style evaluation sessions [32-34] are valuable in advancing trainee education in multidisciplinary oncology care. Depending on the topic, these teaching sessions could be offered in person, remotely via web-based applications, or as a prerecording to enhance trainee participation. As indicated by Mackay et al [36], tumor board simulation sessions contribute to significant improvements in trainee confidence and skills in participating in tumor boards. This is a novel educational intervention not traditionally offered by residency or fellowship programs. The addition of such resources and programs outside of the mainstream postgraduate training programs has the potential to supplement trainee education toward multidisciplinary oncology care.

Given the time constraint of residency and fellowship, it is not feasible for trainees to gain all relevant multidisciplinary knowledge and skills while also excelling in all core competencies relevant to their program. Every proposed intervention will have its own challenges to implement and needs to be balanced against other rotations within the curriculum. Yet, it is preferred that trainees obtain sufficient multidisciplinary knowledge during training rather than through experience during practice. It is crucial that training programs conduct an evaluation of any new educational intervention and prioritize selected interventions in their curricula based on outcomes and feedback.

Limitations

This study has limitations. Only 24 studies have analyzed the quality of multidisciplinary oncology education among postgraduate medical trainees. Furthermore, we limited our study to English-only and primary papers. It is possible that additional studies analyzing multidisciplinary oncology education in other languages or papers (eg, grey literature) exist that are missing from our results. Over a third of these studies were also published more than 5 years ago. Particularly, 3 of the intervention studies are by Sloan et al [32-34], published in 1997, 1999, and 2004, which could have had overlapping participants. This could limit the generalizability of the findings from these studies. There is a need for additional and more contemporary research assessing the needs of postgraduate medical trainees and the impact of newer educational interventions. It is particularly important to evaluate the use of technologies currently used in medical education such as web-based live teaching [43-47], clinical teaching tools such as case-based modules with built-in radiology software [53,54], and virtual reality surgical training [55-57]. Additionally, none of the studies on educational interventions were conducted with trainees in geriatric oncology. As previously discussed, this is an important aspect of oncology, though generally missing from oncology training curriculums. Thus, additional studies are needed within these fields. Furthermore, while a large proportion of studies solely focus on gaps in geriatric oncology education, this may not be generalizable to all multidisciplinary oncology education needs. Future research will be important in developing multidisciplinary oncology curricula for postgraduate trainees.

XSL•FO
RenderX

Conclusions

This systematic review demonstrated several gaps in the existing multidisciplinary oncology training of postgraduate medical trainees and the promising results of various educational interventions in bridging these gaps. Further studies investigating the needs of trainees at both local and national levels are needed to develop specific educational curricula and program requirements that focus on multidisciplinary oncology collaboration. Future research should also assess contemporary educational interventions to determine the most effective methods of attaining multidisciplinary oncology expertise among postgraduate medical trainees.

Authors' Contributions

The authors met all International Committee of Medical Journal Editors criteria for authorship. HT, GK, ER, ME, and TDC contributed to the study design. HT, GK, CML, IB, ZF, RV, and ME contributed to the processes of screening or data acquisition. HT, GK, CML, IB, ZF, ER, TDC, and RV participated in data analysis. All authors contributed to manuscript drafting and revision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Research data, search strategy, and assessment results.
[DOCX File, 70 KB - mededu_v11i1e63655_app1.docx]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File, 33 KB - mededu_v11i1e63655_app2.docx]

References

1. Ahmad FB, Cisewski JA, Anderson RN. Mortality in the United States—provisional data, 2023. MMWR Morb Mortal Wkly Rep 2024 Aug 8;73(31):677-681. [doi: 10.15585/mmwr.mm7331a1] [Medline: 39116025]
2. Hong NJL, Wright FC, Gagliardi AR, Paszat LF. Examining the potential relationship between multidisciplinary cancer care and patient survival: an international literature review. J Surg Oncol 2010 Aug 1;102(2):125-134. [doi: 10.1002/jso.21589] [Medline: 20648582]
3. Aizer AA, Paly JJ, Michaelson MD, et al. Medical oncology consultation and minimization of overtreatment in men with low-risk prostate cancer. J Oncol Pract 2014 Mar;10(2):107-112. [doi: 10.1200/JOP.2013.000902] [Medline: 24399853]
4. Goulart BHL, Reyes CM, Fedorenko CR, et al. Referral and treatment patterns among patients with stages III and IV non-small-cell lung cancer. J Oncol Pract 2013 Jan;9(1):42-50. [doi: 10.1200/JOP.2012.000640] [Medline: 23633970]
5. Booth CM, Siemens DR, Peng Y, Mackillop WJ. Patterns of referral for perioperative chemotherapy among patients with muscle-invasive bladder cancer: a population-based study. Urol Oncol 2014 Nov;32(8):1200-1208. [doi: 10.1016/j.urolonc.2014.05.012] [Medline: 24968946]
6. ACGME common program requirements (residency). Accreditation Council for Graduate Medical Education. 2023 Jul 1. URL: https://www.acgme.org/globalassets/pfassets/programrequirements/cprresidency_2023.pdf [accessed 2024-12-24]
7. ACGME common program requirements (fellowship). Accreditation Council for Graduate Medical Education. 2022 Jul 1. URL: https://www.acgme.org/globalassets/pfassets/programrequirements/cprfellowship_2022v3.pdf [accessed 2024-12-24]
8. ACGME program requirements for graduate medical education in complex general surgical oncology. Accreditation Council for Graduate Medical Education. 2023 Jul 1. URL: https://www.acgme.org/globalassets/pfassets/programrequirements/446_complexgeneralsurgicaloncology_2023.pdf [accessed 2024-12-24]
9. ACGME program requirements for graduate medical education in hematology and medical oncology. Accreditation Council for Graduate Medical Education. 2024 Jul 1. URL: https://www.acgme.org/globalassets/pfassets/programrequirements/2024-prs/155_hematologyandmedicaloncology_2024.pdf [accessed 2024-12-24]
10. ACGME program requirements for graduate medical education in radiation oncology. Accreditation Council for Graduate Medical Education. 2023 Jul 1. URL: https://www.acgme.org/globalassets/pfassets/programrequirements/430_radiationoncology_2023.pdf [accessed 2024-12-24]
11. Mattes MD. Multidisciplinary oncology education: going beyond tumor board. J Am Coll Radiol 2016 Oct;13(10):1239-1241. [doi: 10.1016/j.jacr.2016.06.005] [Medline: 27474420]
12. Mattes MD, Ye JC, Peters GW, et al. Pilot study demonstrating the value of interdisciplinary education on the integration of radiation therapy in lung cancer management. J Cancer Educ 2023 Apr;38(2):590-595. [doi: 10.1007/s13187-022-02158-8] [Medline: 35357645]
13. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]

XSL•FO
RenderX

14. Covidence. URL: www.covidence.org [accessed 2025-05-16]
15. Hong QN, Fàbregues S, Bartlett G, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Educ Inf 2018;34(4):285-291. [doi: 10.3233/EFI-180221]
16. Akthar AS, Hellekson CD, Ganai S, et al. Interdisciplinary oncology education: a national survey of trainees and program directors in the United States. J Canc Educ 2018 Jun;33(3):622-626. [doi: 10.1007/s13187-016-1139-6]
17. Brenner BM, De Donno MA. Assessing gaps in surgical oncology training: results of a survey of general surgery residents. J Surg Educ 2020;77(4):749-756. [doi: 10.1016/j.jsurg.2020.01.011] [Medline: 32063511]
18. Park KU, Selby L, Chen XP, et al. Development of residents' self-efficacy in multidisciplinary management of breast cancer survey. J Surg Res 2020 Jul;251:275-280. [doi: 10.1016/j.jss.2020.02.014] [Medline: 32197183]
19. Wilson JP, Miller A, Edge SB. Breast education in general surgery residency. Am Surg 2012 Jan;78(1):42-45. [Medline: 22273306]
20. Morris L, Turner S, Thiruthaneeswaran N, et al. An international expert Delphi consensus to develop dedicated geriatric radiation oncology curriculum learning outcomes. Int J Radiat Oncol Biol Phys 2022 Aug 1;113(5):934-945. [doi: 10.1016/j.ijrobp.2022.04.030] [Medline: 35500796]
21. Walraven JEW, van der Meulen R, van der Hoeven JJM, et al. Preparing tomorrow's medical specialists for participating in oncological multidisciplinary team meetings: perceived barriers, facilitators and training needs. BMC Med Educ 2022 Jun 27;22(1):502. [doi: 10.1186/s12909-022-03570-w] [Medline: 35761247]
22. Delaye M, Try M, Rousseau A, et al. Onco-nephrology: physicians' expectations about a new subspecialty. J Cancer Educ 2023 Jun;38(3):878-884. [doi: 10.1007/s13187-022-02201-8] [Medline: 35840858]
23. Mäurer M, Staudacher J, Meyer R, et al. Importance of interdisciplinarity in modern oncology: results of a national intergroup survey of the Young Oncologists United (YOU). J Cancer Res Clin Oncol 2023 Sep;149(12):10075-10084. [doi: 10.1007/s00432-023-04937-2] [Medline: 37261525]
24. Eid A, Hughes C, Karuturi M, Reyes C, Yorio J, Holmes H. An interprofessionally developed geriatric oncology curriculum for hematology-oncology fellows. J Geriatr Oncol 2015 Mar;6(2):165-173. [doi: 10.1016/j.jgo.2014.11.003] [Medline: 25487037]
25. David V, Hsu T, Mithoowani S, Fraser G, Mian H. What do hematology residents know about caring for older adults with cancer? A national survey of Canadian hematology residents' knowledge and interests. J Geriatr Oncol 2022 Nov;13(8):1236-1240. [doi: 10.1016/j.jgo.2022.08.018] [Medline: 36050270]
26. Maggiore RJ, Callahan KE, Tooze JA, Parker IR, Hsu T, Klepin HD. Geriatrics fellowship training and the role of geriatricians in older adult cancer care: a survey of geriatrics fellowship directors. Gerontol Geriatr Educ 2018;39(2):170-182. [doi: 10.1080/02701960.2016.1247070] [Medline: 27749199]
27. Morris L, Thiruthaneeswaran N, Lehman M, Hasselburg G, Turner S. Are future radiation oncologists equipped with the knowledge to manage elderly patients with cancer? Int J Radiat Oncol Biol Phys 2017 Jul 15;98(4):743-747. [doi: 10.1016/j.ijrobp.2017.01.001] [Medline: 28258899]
28. Picca A, Reed S. Off to boarding school: exploring how physicians learn in tumor board. Pediatr Blood Cancer 2023 Nov;70(11):e30632. [doi: 10.1002/pbc.30632] [Medline: 37610271]
29. Givi B, Gordon AJ, Park YS, Lydiatt WM, Tekian A. Needs assessment in head and neck surgical oncology training: a qualitative study of expert opinions. Head Neck 2022 Nov;44(11):2528-2536. [doi: 10.1002/hed.27158] [Medline: 35920353]
30. Le Nail LR, Samargandi R. Teaching potential of multidisciplinary tumor board meetings for orthopedic residents: insights from a French sarcoma reference center. Cureus 2023 May;15(5):e39783. [doi: 10.7759/cureus.39783] [Medline: 37265907]
31. Cook MR, Graff-Baker AN, Moren AM, et al. A disease-specific hybrid rotation increases opportunities for deliberate practice. J Surg Educ 2016;73(1):1-6. [doi: 10.1016/j.jsurg.2015.09.005] [Medline: 26481268]
32. Sloan DA, Donnelly MB, Schwartz RW, et al. The multidisciplinary structured clinical instruction module as a vehicle for cancer education. Am J Surg 1997 Mar;173(3):220-225. [doi: 10.1016/s0002-9610(97)89596-7] [Medline: 9124631]
33. Sloan DA, Witzke DB, Plymale MA, et al. A multidisciplinary workshop to teach head and neck oncology to residents: results of a pilot study. J Cancer Educ 1999;14(4):228-232. [doi: 10.1080/08858199909528632]
34. Sloan DA, Plymale MA, Donnelly MB, Schwartz RW, Edwards MJ, Bland KI. Enhancing the clinical skills of surgical residents through structured cancer education. Ann Surg 2004 Apr;239(4):561-566. [doi: 10.1097/01.sla.0000118568.75888.04] [Medline: 15024318]
35. Meani F, Kovacs T, Wandschneider W, Costa A, Pagani O. Multidisciplinary blended learning to build a breast cancer specialist career: survey on the perspective of the first 2 cohorts of the ESO-ULM Certificate of Competence in Breast cancer (CCB). BMC Med Educ 2022 May 5;22(1):344. [doi: 10.1186/s12909-022-03414-7] [Medline: 35513883]
36. Mackay EC, Patel KR, Davidson C, et al. Simulation as an effective means of preparing trainees for active participation in MDT meetings. Future Healthc J 2024 Mar;11(1):100017. [doi: 10.1016/j.fhj.2024.100017] [Medline: 38646046]
37. Khoshgoftar Z, Sodeifian F, Allameh F. Improving the educational gap with implementing of teaching scholarship in virtual multidisciplinary tumor boards. Int J Cancer Manag 2023;In Press(In Press). [doi: 10.5812/ijcm-137490]
38. Martin EJ, Nalawade VV, Murphy JD, Jones JA. Incorporating palliative radiotherapy education into hospice and palliative medicine fellowship training: a feasibility study. Ann Palliat Med 2019 Sep;8(4):436-441. [doi: 10.21037/apm.2019.04.02]

XSL•FO
RenderX

39. Morris L, Turner S, Thiruthaneeswaran N, Agar M. Improving the education of radiation oncology professionals in geriatric oncology: where are we and where should we be? Semin Radiat Oncol 2022 Apr;32(2):109-114. [doi: 10.1016/j.semradonc.2021.11.008] [Medline: 35307112]
40. Balducci L, Ershler WB. Cancer and ageing: a nexus at several levels. Nat Rev Cancer 2005 Aug;5(8):655-662. [doi: 10.1038/nrc1675] [Medline: 16056261]
41. Harris P, Bhanji F, Topps M, et al. Evolving concepts of assessment in a competency-based world. Med Teach 2017 Jun;39(6):603-608. [doi: 10.1080/0142159X.2017.1315071] [Medline: 28598736]
42. Dedeilia A, Sotiropoulos MG, Hanrahan JG, Janga D, Dedeilias P, Sideris M. Medical and surgical education challenges and innovations in the COVID-19 era: a systematic review. In Vivo 2020;34(3 suppl):1603-1611. [doi: 10.21873/invivo.11950]
43. Wendt S, Abdullah Z, Barrett S, et al. A virtual COVID-19 ophthalmology rotation. Surv Ophthalmol 2021;66(2):354-361. [doi: 10.1016/j.survophthal.2020.10.001] [Medline: 33058927]
44. Chandra S, Laoteppitaks C, Mingioni N, Papanagnou D. Zooming-out COVID-19: virtual clinical experiences in an emergency medicine clerkship. Med Educ 2020 Dec;54(12):1182-1183. [doi: 10.1111/medu.14266] [Medline: 32502282]
45. Haws BE, Mannava S, Schuster BK, DiGiovanni BF. Implementation and evaluation of a formal virtual medical student away rotation in orthopaedic surgery during the COVID-19 pandemic. Foot Ankle Orthopaedic 2022 Jan;7(1):2473011421S00229. [doi: 10.1177/2473011421S00229]
46. Villa S, Janeway H, Preston-Suni K, et al. An emergency medicine virtual clerkship: made for COVID, here to stay. West J Emerg Med 2021 Dec 17;23(1):33-39. [doi: 10.5811/westjem.2021.11.54118] [Medline: 35060858]
47. Harrell Shreckengost CS, Reitz A, Ludi E, Rojas Aban R, Jáuregui Paravicini L, Serrot F. Lessons learned during the COVID-19 pandemic using virtual basic laparoscopic training in Santa Cruz de la Sierra, Bolivia: effects on confidence, knowledge, and skill. Surg Endosc 2022 Dec;36(12):9379-9389. [doi: 10.1007/s00464-022-09215-9] [Medline: 35419639]
48. Wilson HC, Lim TR, Axelrod DM, et al. A multimedia paediatric cardiology assessment tool for medical students and general paediatric trainees: development and validation. Cardiol Young 2023 Mar;33(3):444-448. [doi: 10.1017/S1047951122001123] [Medline: 35411842]
49. Nozari A, Mukerji S, Lok LL, et al. Perception of web-based didactic activities during the COVID-19 pandemic among anesthesia residents: pilot questionnaire study. JMIR Med Educ 2022 Mar 31;8(1):e31080. [doi: 10.2196/31080] [Medline: 35275840]
50. Haring RS, Rydberg LK, Mallow MK, Kortebein P, Verduzco-Gutierrez M. Development and implementation of an international virtual didactic series for physical medicine and rehabilitation graduate medical education during COVID-19. Am J Phys Med Rehabil 2022 Feb 1;101(2):160-163. [doi: 10.1097/PHM.0000000000001926] [Medline: 35026777]
51. Murdock HM, Penner JC, Le S, Nematollahi S. Virtual morning report during COVID-19: a novel model for case-based teaching conferences. Med Educ 2020 Sep;54(9):851-852. [doi: 10.1111/medu.14226] [Medline: 32403168]
52. Shih KC, Chan JCH, Chen JY, Lai JSM. Ophthalmic clinical skills teaching in the time of COVID-19: a crisis and opportunity. Med Educ 2020 Jul;54(7):663-664. [doi: 10.1111/medu.14189] [Medline: 32324929]
53. El-Ali A, Kamal F, Cabral CL, Squires JH. Comparison of traditional and web-based medical student teaching by radiology residents. J Am Coll Radiol 2019 Apr;16(4 Pt A):492-495. [doi: 10.1016/j.jacr.2018.09.048] [Medline: 30449521]
54. Sugi MD, Kennedy TA, Shah V, Hartung MP. Bridging the gap: interactive, case-based learning in radiology education. Abdom Radiol 2021 Dec;46(12):5503-5508. [doi: 10.1007/s00261-021-03147-z]
55. Bernardo A. Virtual reality and simulation in neurosurgical training. World Neurosurg 2017 Oct;106:1015-1029. [doi: 10.1016/j.wneu.2017.06.140] [Medline: 28985656]
56. Bakshi SK, Lin SR, Ting DSW, Chiang MF, Chodosh J. The era of artificial intelligence and virtual reality: transforming surgical education in ophthalmology. Br J Ophthalmol 2021 Oct;105(10):1325-1328. [doi: 10.1136/bjophthalmol-2020-316845]
57. Goh GS, Lohre R, Parvizi J, Goel DP. Virtual and augmented reality for surgical training and simulation in knee arthroplasty. Arch Orthop Trauma Surg 2021 Dec;141(12):2303-2312. [doi: 10.1007/s00402-021-04037-1] [Medline: 34264380]

Abbreviations:

EPA: entrustable professional activity
MMAT: Mixed Methods Appraisal Tool
MTBM: multidisciplinary tumor board meeting
OSCE: Objective Structured Clinical Examination
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

XSL•FO
RenderX

Edited by B Lesselroth; submitted 25.06.24; peer-reviewed by N Waheed, R Bansal, S Ganesh; revised version received 26.01.25; accepted 19.03.25; published 26.05.25.

Please cite as:
Tahmasebi H, Ko G, Lam CM, Bilgen I, Freeman Z, Varghese R, Reel E, Englesakis M, Cil TD
Multidisciplinary Oncology Education Among Postgraduate Trainees: Systematic Review
JMIR Med Educ 2025;11:e63655
URL: https://mededu.jmir.org/2025/1/e63655
doi:10.2196/63655

© Houman Tahmasebi, Gary Ko, Christine M Lam, Idil Bilgen, Zachary Freeman, Rhea Varghese, Emma Reel, Marina Englesakis, Tulin D Cil. Originally published in JMIR Medical Education (https://mededu.jmir.org), 26.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Bridging Gaps in Telemedicine Education in Romania to Support Future Health Care: Scoping Review

Mircea Adrian Focsa[1], MD, PhD; Virgil Rotaru[2], MPhys; Octavian Andronic[3], MD, PhD; Marius Marginean[4], MD, PhD; Sorin Florescu[5], MD, PhD

[1]Medical Informatics and Biostatistics Department, Victor Babe University of Medicine and Pharmacy Timi oara, Timisoara, Romania
[2]Victor Babe University of Medicine and Pharmacy Timi oara, Eftimie Murgu 2 Sqr, Timi oara, Romania
[3]General Surgery Department and Innovation and eHealth Center, Carol Davila University of Medicine and Pharmacy, Bucharest, Romania
[4]Public Health Authority, Brasov, Romania
[5]Orthopedics-traumatology II, Victor Babe University of Medicine and Pharmacy Timi oara, Timisoara, Romania

Corresponding Author:
Virgil Rotaru, MPhys
Victor Babe University of Medicine and Pharmacy Timi oara, Eftimie Murgu 2 Sqr, Timi oara, Romania

Abstract

Background: Telemedicine is a key element of modern health care, providing remote medical consultations and bridging the gap between patients and health care providers. Despite legislative advancements and pilot programs, the integration of telemedicine education in Romania remains limited. Addressing these educational gaps is essential for preparing current and future medical professionals to effectively use telemedicine technologies.

Objective: This study aimed to evaluate the current state of telemedicine education for medical professionals in Romania, focusing on the integration of diagnostic and therapeutic capabilities into medical curricula, identifying the challenges and opportunities, and providing recommendations for improving telemedicine education.

Methods: A scoping review was conducted following Arksey and O'Malley's framework. Peer-reviewed articles from 2019 to 2023 were identified using databases such as PubMed and Scopus. Additional gray literature was reviewed to provide a comprehensive understanding of telemedicine education in Romania. Data were thematically analyzed to extract key findings and recommendations.

Results: The review identified significant progress in the legislative and infrastructural aspects of telemedicine in Romania, but highlighted gaps in integrating telemedicine education into curricula for medical professionals and other health care practitioners directly involved in telemedicine practices. While some universities have included telemedicine components, dedicated telemedicine courses and hands-on training remain insufficient. Barriers include a lack of infrastructure, digital literacy, and practical exposure to telemedicine technologies.

Conclusions: For telemedicine to be effectively integrated into Romania's health care system, medical education must be adapted to include comprehensive telemedicine training. Recommendations include enhancing digital literacy, fostering public-private partnerships, and incorporating telemedicine into undergraduate and continuous professional education programs. These efforts are essential for improving healthcare access and quality through telemedicine.

(JMIR Med Educ 2025;11:e66458) doi:10.2196/66458

KEYWORDS

telemedicine; digital health; healthcare education; micro-credentials; scoping review; health education; Romania; future healthcare; telehealth; healthcare providers; technologies; digital literacy; healthcare system; quality care

Introduction

Telemedicine in Romania has undergone significant development over the years, marked by early pilot projects, legislative advancements, and a growing recognition of its potential to transform health care delivery. Given the rapid advancements in digital health globally, understanding and improving telemedicine education in Romania will not only enhance national health care delivery but also offer insights for similar health care systems transitioning to more digital practices.

For the purposes of this study and in accordance with Romanian legislation, telemedicine is defined as the use of digital technologies to deliver medical acts such as diagnosis, treatment, and therapy performed exclusively by licensed medical professionals. This differs from telehealth, which encompasses broader health-related services, including health education, prevention, and administrative activities.

XSL•FO
RenderX

While telemedicine inherently involves medical acts performed by licensed clinicians, it relies on a collaborative team, including nonclinical professionals, to ensure efficient and comprehensive care delivery.

Several interdependent factors, including educational competencies, regulatory policies, technological infrastructure, and institutional readiness shape telemedicine education. This study adopts a theoretical framework that integrates competency-based medical education, digital health policies, and workforce development strategies to evaluate the current state of telemedicine training in Romania. The framework is informed by internationally recognized guidelines, such as the World Health Organization (WHO) Digital Health Competency Framework (DHCF), the International Society for Telemedicine & eHealth (ISfTeH) guidelines, and the European Health Telematics Association (EHTEL) recommendations, which outline essential skills and knowledge areas for telemedicine practitioners.

Within this framework, telemedicine competencies are influenced by both regulatory structures and digital readiness, which shape how educational programs can be effectively implemented. While Romania has made legislative progress in supporting telemedicine, educational curricula remain inconsistent, lacking standardized competencies and hands-on training opportunities. Furthermore, limited technological infrastructure and digital literacy among both professionals and patients present additional challenges. By assessing these dimensions, this study identifies the current gaps in telemedicine education and proposes targeted recommendations to improve training programs, ensuring alignment with international best practices.

The journey began in 2001 when the Romanian Space Agency (ROSA) launched the Demonstrative Pilot of Telemedicine. This project, a significant milestone in the country's health care history, focused on diagnostic, clinical, and educational applications, serving as a pioneering effort to explore the capabilities of telemedicine in the country. In 2003, further progress was made in establishing the Romanian Association for Telemedicine and Space Applications for Health (ATASS), which aimed to promote and develop telemedicine technologies [1].

A significant legislative milestone came in 2018 when telemedicine was formally incorporated into Romanian law through Government Emergency Ordinance number 8/2018, which amended Health Reform Law 95/2006. This legislative change aimed to address the chronic shortage of medical personnel, particularly in remote areas, and ensure more equitable access to healthcare services.

The COVID-19 pandemic in 2020 acted as a catalyst for telemedicine's rapid adoption. In response to the crisis, the Romanian government took swift and decisive action to establish a regulatory framework for telemedicine services. Government Decision 252/2020 and subsequent ordinances laid the groundwork for telemedicine during states of emergency and beyond. These regulations facilitated various telemedicine services, including teleconsultation, tele-expertise, teleradiology, and telemonitoring. By 2022, the regulatory framework had further evolved with Government Decision 1133/2022 [2], which approved comprehensive implementation norms for telemedicine. This decision standardized procedures for scheduling remote appointments, protecting data privacy, and setting up payment mechanisms through the National Health care Insurance House. These measures ensured that telemedicine services could be provided seamlessly and securely, enhancing their integration into the healthcare system.

In the last 2 years, significant financial investments have supported the expansion of telemedicine in Romania. The recovery and resilience plan allocated substantial funds, including approximately €100 million for telemedicine support and €300-€400 million for hospital digitalization. These investments underscored the government's commitment to advancing telemedicine as a critical component of health care delivery.

Despite these significant financial investments, the low levels of health and digital literacy among Romanian citizens present substantial barriers to the successful adoption and utilization of these technologies. Digital literacy for the public refers to the ability to access, understand, and use digital technologies for obtaining health information and services. For health care professionals, digital literacy extends to proficiency in using digital tools for clinical care, such as eHealth records, telemedicine platforms, and data privacy protocols.

Eurostat statistics [3] indicate that only 40% of Romanians use the internet to search for health information, significantly below the European Union average of 55%. This discrepancy highlights the role of education in digital engagement, with just 17% of individuals with low educational attainment using the internet for health purposes, compared with 41% of those with medium education and 66.5% of highly educated individuals.

The first cross-sectional study on health literacy in Romania [4] underscores the challenges faced by the population in processing health information. Approximately 21.6% of respondents found it difficult to protect themselves from illness based on health information provided by the media. Moreover, 7.5% of participants demonstrated inadequate health literacy, and 33.2% had problematic health literacy, leaving a majority (59.2%) with sufficient health literacy. Key determinants of health literacy included age, gender, education, and self-reported health status, while the residential area did not appear to influence health literacy levels. These findings underscore the considerable gaps in both health and digital literacy among the Romanian population.

Moreover, significant gaps still need to be addressed in integrating telemedicine education effectively within medical curricula, particularly in ensuring that current and future medical professionals are adequately prepared to leverage these technologies in practice. Most medical universities and medical schools have started incorporating telemedicine into their curricula, mainly as part of the medical informatics discipline, aiming to familiarize future health care professionals with digital health tools and focusing on both telemedicine's technical and ethical aspects. Professional development opportunities have also expanded, with continuous education programs incorporating modules on telemedicine. Online training

XSL•FO
RenderX

platforms and workshops have become vital resources for health care providers, helping them stay updated with the latest advancements and best practices in telemedicine. Beyond serving as a means of continuing education, these platforms often provide essential initial training for health care professionals new to telehealth, equipping them with foundational knowledge and skills.

Telehealth competency, encompassing knowledge, skills, and attitudes essential for effectively delivering care via telemedicine, is increasingly recognized as a critical aspect of modern health care practice. However, in Romania, health care professionals often lack structured and standardized training in telemedicine, resulting in gaps in areas such as conducting teleconsultations, ensuring data security, and communicating effectively with patients in virtual settings. These gaps highlight the need for targeted educational interventions to prepare health care professionals for the demands of telemedicine.

The scope of this study is confined to telemedicine, which involves clinical activities performed by physicians and other licensed medical professionals, ensuring a clear distinction from the broader concept of telehealth.

The aim of this scoping review is to evaluate the current state of telemedicine education in Romania, identify the challenges and opportunities associated with its integration into medical curricula, and provide recommendations for improving telemedicine education. Specifically, this study maps existing telemedicine education initiatives, assesses barriers to implementation, and proposes strategies to enhance training programs for health care professionals.

Methods

Study Design

This study used a scoping review methodology to comprehensively explore the current landscape of telemedicine education in Romania. The study specifically evaluates educational approaches for medical professionals performing telemedicine, addressing clinical activities such as diagnosis, therapy, and patient management. Telemedicine education evaluated in this study encompasses competencies applicable to multiple specialties, including primary care, chronic disease management, and specialized services such as telemonitoring and telerehabilitation. A scoping review was chosen for its ability to map key concepts, types of evidence, and gaps in research related to a defined area or field of interest, particularly in complex or under-reviewed topics. This approach is especially suitable for telemedicine education in Romania, given its rapidly evolving nature and the need to synthesize diverse sources of information

Research Question

The primary research question guiding this scoping review was: “What is the current state of telemedicine education in Romania, and what are the key challenges and opportunities for its integration into medical curricula?”. This question was formulated to encompass the broad scope of telemedicine education, including formal educational programs, professional development initiatives, and digital literacy efforts.

Literature Search Strategy

A systematic search of peer-reviewed journal papers was conducted across multiple databases, including PubMed, Scopus, Web of Science, and Google Scholar. The search covered articles published between January 2019 and December 2023. The following keywords and their combinations were used: telemedicine, telehealth, digital health, medical education, telemedicine education, Romania, eHealth, and digital literacy.

Additional filters were applied to include only papers available in English or Romanian, with a focus on education, telemedicine implementation, and health care policy in Romania. Papers were limited to those that addressed telemedicine within the context of health care education, the use of digital tools in clinical training, and barriers or facilitators to telemedicine adoption.

In addition to peer-reviewed papers, relevant gray literature was included in the scoping review. Sources of gray literature encompassed reports from governmental and nongovernmental organizations, policy briefs, and institutional documents related to telemedicine education in Romania. These documents were identified through searches of online repositories and institutional websites, and they provided critical context on legislative developments, pilot projects, and educational initiatives that may not have been extensively covered in academic databases.

Study Selection

The study selection process involved 2 independent researchers (MF and VR) who screened titles, abstracts, and full texts to ensure alignment with the inclusion and exclusion criteria. Any disagreements were resolved through discussion, and when consensus could not be reached, a third researcher reviewed the articles to make the final decision. The process was facilitated using Rayyan platform, allowing efficient tracking and documentation of the selection process.

The initial search yielded 105 journal papers, screened for relevance based on titles and abstracts. Articles that focused on telemedicine in clinical practice without addressing educational aspects were excluded. After this preliminary screening, 35 papers remained for full-text review. A further screening based on inclusion and exclusion criteria led to the identification of 19 papers deemed highly relevant to the study’s aims.

Eligibility Criteria

The inclusion and exclusion criteria are listed in Textbox 1.

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria

Inclusion criteria

- Studies published between 2019 and 2023.
- Studies addressing telemedicine education or digital literacy training in medical or health care fields.
- Studies conducted in or relevant to the Romanian context.

Exclusion criteria

- Articles focusing exclusively on telemedicine in clinical practice without reference to education.
- Studies not available in English or Romanian.

Data Extraction

For each of the 19 selected peer-reviewed papers, data were extracted using a standardized data charting form, which included the following variables: (1) author(s), publication year, and country of origin; (2) study design (eg, qualitative, quantitative, and mixed methods); (3) the focus of the study (eg, telemedicine education, barriers to digital health adoption, and telemedicine curriculum development); (4) key findings relevant to telemedicine education, digital literacy, and health care training; and (5) recommendations for telemedicine integration into medical education.

Data Analysis

Two independent researchers conducted a thematic analysis to identify recurring patterns and themes within the selected literature. An inductive approach allowed themes to emerge naturally from the data. The analysis involved coding and categorizing data using manual methods, followed by a review of the themes to ensure consistency and relevance. The themes were grouped into broad categories, such as the current status of telemedicine education in Romania, barriers to telemedicine education, opportunities for development, and digital literacy challenges.

Ethical Considerations

Since this study was based on a review of publicly available literature, no ethical approval was required. However, all articles were reviewed with a commitment to academic integrity and transparency, and any potential conflicts of interest were disclosed.

Results

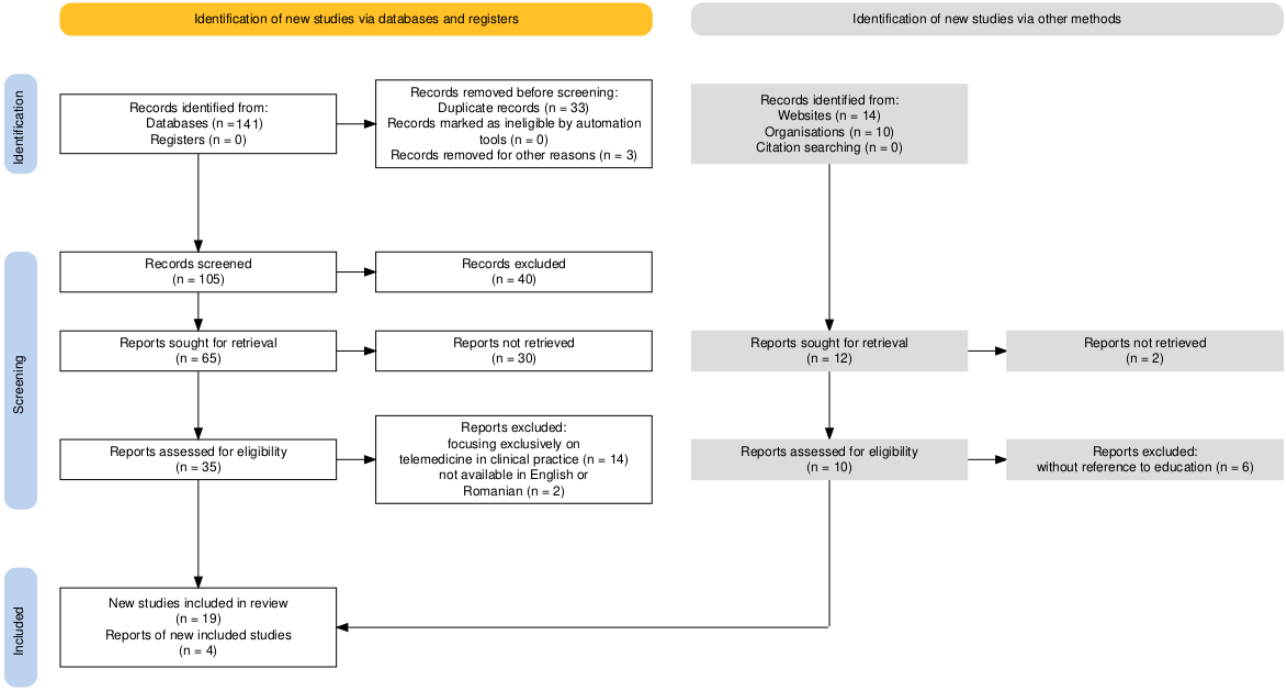
Included Studies

The initial search yielded 105 journal papers, screened for relevance based on titles and abstracts. Articles that focused on telemedicine in clinical practice without addressing educational aspects were excluded. After this preliminary screening, 35 papers remained for full-text review. A further screening based on inclusion and exclusion criteria led to the identification of 19 papers deemed highly relevant to the study's aims (Figure 1 and Checklist 1) .

The results presented are based on themes identified through inductive thematic analysis, which highlighted key areas such as the integration of telemedicine into curricula, barriers to adoption, and digital literacy as a critical enabler for telemedicine education.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart for identification and selection of studies.



Current State of Telemedicine Education in Romania

The data collected and synthesized in Table 1 offers a comprehensive overview of recent research in telemedicine, eHealth, and artificial intelligence (AI)-based health care interventions in Romania. It includes various study designs such as cross-sectional, prospective, case-control, and system architecture studies and spans diverse health care topics, including telemedicine-driven rehabilitation, the impact of virtual communities on telemedicine adoption, and the integration of AI and Internet of Things (IoT) technologies in health care. The analysis covers key aspects of telemedicine's role during and post–COVID-19 and its relevance for both chronic diseases and mental health management. The studies underscore the positive reception of telemedicine in various medical fields, from diabetes management to cardiac rehabilitation, with the technology being especially vital during the COVID-19 pandemic. However, challenges such as digital literacy, accessibility, and infrastructure remain.

While Table 1 provides a summary of the reviewed studies, the specific components of telemedicine education and digital literacy training are discussed in greater detail within the following subsection, to align with the thematic approach of this scoping review.

XSL•FO
RenderX

Table . Overview of recent research in telemedicine, eHealth, and artificial intelligence–based health care interventions in Romania [5-23].

PMID	The focus	Key findings	Recommendations
35879930 [5]	Openness of medical students to telemedicine	Moderate-to-high acceptance, need for practical exposure	Integrate telemedicine into medical curricula, enhance digital literacy
35935629 [6]	Evaluating telemedicine benefits for cardiovascular patients during COVID-19.	Telemedicine facilitated patient management, including medication adjustments, but barriers like data security and reimbursement need addressing.	Improve telemedicine frameworks and regulations for cardiovascular care beyond the pandemic.
37297692 [7]	Perceptions of telemedicine among health care professionals	Positive outlook, concerns over digital literacy	Increase digital education for health care professionals
37510969 [8]	Telemedicine-driven pulmonary rehabilitation for post–COVID-19	Significant improvement in physical and mental health	Integrate telerehabilitation into postacute COVID-19 management
36141685 [9]	AI[a] in eHealth, telemedicine, and remote monitoring	AI advances in eHealth and telemedicine	Enhance integration of AI into eHealth for better health care services
36141899 [10]	Home-based robotic cardiac telerehabilitation system	RoboTeleRehab system is feasible, positive feedback	Further testing on cardiac patients, integration into rehabilitation programs
33923842 [11]	Blockchain-enabled framework for mHealth[b] systems	Improves security, transparency, and immutability	Implement blockchain for secure patient data management in mHealth
35062129 [12]	AI in primary care and telemedicine	AI aids in primary care diagnosis, treatment, and decision-making	Enhance AI systems to support primary care and telemedicine workflows
36141297 [13]	Impact of virtual communities on telemedicine usage	Virtual communities influence patient satisfaction and usage intention	Use virtual communities to promote telemedicine services
35954526 [14]	Adoption of eHealth and mHealth for mental health	Accessibility and data security are critical for adoption	Improve accessibility and data security in digital mental health tools
36556027 [15]	Management of dilated cardiomyopathy during COVID-19 using telemedicine	Telemedicine maintains clinical stability in patients with home monitoring	Use multiparametric home monitoring to manage dilated cardiomyopathy during crises
33953605 [16]	Perception of Romanian family doctors on telemedicine during COVID-19	Positive view, but tele-diagnostic challenges and time constraints	Training for family doctors and continued telemedicine reimbursement
34898981 [17]	Assessing patient adherence to telemedicine in diabetes	Developed reliable and valid tool for telemedicine adherence in diabetes care	Use the instrument to optimize telemedicine platforms based on patient needs
35150518 [18]	Attitudes toward eHealth during the COVID-19 pandemic	Negative attitudes in Greece due to forced usage	Address technical skills gaps and improve ease of access
37372846 [19]	Telemonitoring for cardiovascular disease during and post-COVID-19	Telemedicine improved cardiovascular prevention during the pandemic	Universal access to home telemonitoring for high-risk cardiovascular patients
36900948 [20]	Virtual assistant in cardiac rehabilitation	Similar results for virtual versus in-person rehabilitation	Optimize virtual assistants for cardiac rehab
33800728 [21]	IoT[c]-based biometric monitoring system for elders	Scalable solution for monitoring and cognitive assessment	Expand to predictive analysis for cognitive and physiological data
35979169 [22]	Telerehabilitation for Parkinson disease	Improved walking performance using telerehabilitation	Expand telerehabilitation for other neurorehabilitation settings
31407668 [23]	Use of telemedicine for rare diseases in Romania	Telemedicine improves access to rare disease care	Increase telemedicine use for remote consultations in rare diseases

[a]AI: artificial intelligence.

[b]mHealth: mobile health.

[c]IoT: Internet of Things.

XSL•FO
RenderX

University Programs and Curricula

In Romania, prominent medical universities in Timisoara, Cluj-Napoca, Iasi, and Bucharest have incorporated telemedicine and digital health into their educational programs. These changes were particularly accelerated following the COVID-19 pandemic, which highlighted the need for digital health competencies among health care professionals. New curricula encompass various aspects, such as telemedicine applications, electronic health records (EHRs), and data management strategies. Despite these inclusions, the development of dedicated disciplines and laboratory classes specific to telemedicine remains ongoing.

While telemedicine education in Romania has introduced theoretical foundations and basic digital health tools, it remains limited in scope and practical application. Key gaps include the lack of hands-on training, comprehensive and stand-alone telemedicine courses, and interprofessional education initiatives. Furthermore, advanced topics such as AI and patient engagement strategies remain underexplored, emphasiing the need for structured and innovative approaches to telemedicine education.

On October 14, 2023, the Center for Innovation and e-Health at the University of Medicine and Pharmacy “Carol Davila” in Bucharest hosted a specialized course titled “Telemedicine - Current Information and Skills.” [24]. This initiative covers fundamental theoretical concepts of telemedicine, including applicability and legal frameworks, alongside practical skills for using associated technologies.

Many health care professionals participate in continuing medical education (CME) programs to stay updated on telemedicine. These programs focus on skills for telemedicine platforms, data privacy laws, and improving remote patient interaction. The Digital Innovation Zone Association [25], affiliated with the North-East Regional Development Agency, offers comprehensive 1- to 2-month programs delivered in a hybrid online and on-site format.

A significant milestone in telemedicine education within the country was marked by the launch of the Erasmus+ project “TEAM: Supporting Innovation in Telemedicine Education with Cross-European Collaboration” in November 2023. The University of Medicine and Pharmacy “Victor Babes” in Timisoara is a principal participant in this consortium, which also includes partners from Belgium, Slovenia, Croatia, Greece, and Ukraine. The project aims to develop adaptable learning pathways that conform to international best practices, focusing on surmounting challenges such as limited digital literacy and fostering cross-sectoral cooperation. Targeting higher education students in health care and IT, educators, and institutions, the project also extends its reach to health care and IT professionals and policymakers. One of the primary outcomes of this initiative is the establishment of flexible microcredentials in Telemedicine designed to enhance proficiency in digital health among students and professionals alike.

Professional Development Workshops and Seminars

Universities and health care institutions in Romania proactively conduct workshops and seminars to offer practical training in telemedicine technologies. These educational sessions frequently use case studies and practical simulations to enrich the learning experience. They cover a spectrum of topics, from fundamental telemedicine principles to advanced applications such as the integration of AI tools in clinical settings.

Across Romania, hospitals and medical associations also organize workshops and seminars aimed at practicing health care professionals. These sessions foster proficiency in telemedicine platforms, digital communication skills, and data security. Participants, including doctors, nurses, and ancillary staff, benefit from the self-paced learning environment these workshops provide. They typically include hands-on sessions where attendees can interact with telemedicine software, acquire best practices for conducting virtual consultations, and gain insights into the legal and ethical dimensions of remote health care.

A notable initiative was the establishment of the ROHEALTH cluster in 2015, which brought together various entities within the health and bioeconomy sectors to enhance their competitive edge. This cluster supports an online platform offering diverse courses and webinars, including specialized offerings such as “eTELEDOC, emergency telemedicine” [26]. These resources aim to bolster the competencies of health care professionals in the evolving landscape of telemedicine.

Digital Literacy Programs

Several nongovernmental organizations offer training programs and workshops to enhance digital literacy among health care providers. Initiated in 2022 by PALMED, the Patronage of Private Medical Service Providers, the project titled “Digital Skills for Employees - Support for SMEs in the Health Sector to Assimilate Technologies and Develop Telemedicine Services (TELMed)” is a notable endeavor under the Human Capital Operational Program for 2014–2020 [27]. This project specifically aims to augment the digital competencies of personnel across 35 small and medium-sized enterprises in the health care sector, including hospitals, clinics, offices, and laboratories. By enhancing these skills, the initiative not only supports the adaptation and expansion of telemedicine activities but also prepares the workforce for advancements related to Industry 4.0 and smart specialization areas.

Challenges in Telemedicine Education

The challenges identified in this section were derived from thematic analysis of the 19 peer-reviewed papers and relevant gray literature, including institutional reports and policy briefs. These sources highlighted common barriers such as limited digital literacy among health care professionals, the lack of standardized curricula, and inadequate infrastructure for telemedicine training.

The pandemic propelled telemedicine into prominence, shedding light on Romanian family doctors’ diverse experiences and perceptions [16]. Over a quarter of general practitioners reported that remotely addressing patients’ health care needs was more manageable, demonstrating adaptability to telemedicine modalities. Nevertheless, challenges such as the time-intensive nature of teleconsultations, diagnostic uncertainties, and patients’ difficulties with technology have surfaced. These issues highlight the critical need for specialized training programs in

XSL•FO
RenderX

telemedicine for both health care professionals and patients to mitigate disruptions in health care delivery effectively.

Furthermore, the moderate-to-high acceptance of telemedicine among Romanian medical students emphasizes the necessity of integrating telemedicine education early in their medical training. Telemedicine fundamentally transforms the patient-physician relationship, requiring physicians to develop new communication skills, known as "webside manner." The methodological norms for telemedicine services in Romania also emphasize the importance of well-trained professionals who can navigate legal, ethical, business, and practical challenges. Training should start at the undergraduate level and continue through all professional stages for medical staff in all specialties.

While the primary objective of this paper is to explore telemedicine education for health care professionals, patient experiences with telemedicine offer indirect yet critical insights. These experiences highlight areas where health care providers may require additional training, such as building virtual rapport, managing technological issues, and addressing patient concerns about telemedicine efficacy and privacy. Integrating these considerations into training programs can better align telemedicine education with real-world practice.

In 2021, a study [17] evaluating the desirability, acceptability, and adherence to telemedicine among diabetes patients underscores the need for educational programs. Such initiatives targeting patients are essential to foster a positive perception and readiness to use telemedicine services, particularly in chronic conditions like diabetes, where continuous care and monitoring are essential. Patients, particularly those managing chronic diseases, require thorough education to use virtual technology effectively, including understanding how to operate software and hardware, such as mobile communication devices and other digital interfaces essential for remote health care.

Discussion

Principal Findings

This scoping review identified significant gaps in telemedicine education in Romania, despite recent legislative and infrastructural advancements supporting telemedicine adoption. While some medical universities have incorporated telemedicine content into their curricula, there remains a lack of structured, hands-on training and dedicated telemedicine courses. The study also highlighted key barriers, including limited digital literacy among health care professionals, insufficient policy support for mandatory telemedicine training, and a lack of standardized competency frameworks. These findings underscore the need for formalized telemedicine education initiatives, integrated into both undergraduate and CME programs, to enhance digital health preparedness among health care providers.

The findings of this scoping review highlight the significant progress and persistent challenges in telemedicine education in Romania. This discussion will interpret these results, explore their implications, and propose strategies for advancing telemedicine education in the country. To advance telemedicine in Romania, several strategic directions need to be pursued.

Medical schools must develop comprehensive programs that include hands-on training with telemedicine platforms and technologies, such as conducting teleconsultations or using telemonitoring devices in simulated or real clinical environments. The launch of specialised courses and international collaborations, such as the Erasmus+ TEAM project, demonstrates a commitment to enhancing telemedicine education.

In Romania, health care professionals are required by law to participate in CME programs to retain their licenses. While telemedicine-specific education is not yet a mandatory component of these requirements, its growing integration into CME programs highlights the recognition of telemedicine as an essential skill set for modern medical practice. This approach aligns with recommendations from the EHTEL, which emphasizes the importance of ongoing training in digital health for all health care providers [28]. The involvement of industry clusters like ROHEALTH in providing specialized webinars indicates a promising collaboration between academia and industry.

However, the review also suggests that these efforts may not be sufficient to meet the rapidly evolving needs of the health care system. There appears to be a need for more structured, comprehensive, and widely accessible professional development programs in telemedicine. While beneficial for updating specific skills, self-directed CME is often insufficient in providing a holistic understanding of telehealth practice. Without structured education, health care professionals may lack critical knowledge of professional telehealth standards, guidelines, and best practices, placing them at risk of learning by trial and error. Comprehensive, formalized training programs are essential to equip professionals with the competencies needed to deliver high-quality, safe, and effective telehealth care.

The identified gaps in digital literacy among both health care providers and patients represent a significant barrier to the effective implementation of telemedicine. The initiatives aimed at enhancing digital skills, such as PALMED's "Digital Skills for Employees" project, are steps in the right direction. These efforts align with European Union–wide initiatives like the Digital Education Action Plan (2021‑2027) [29], which emphasizes the importance of digital skills across all sectors, including health care. However, these efforts need to be scaled up and integrated more systematically into both medical education and public health initiatives. The review also highlights the need for patient education in telemedicine, particularly for managing chronic conditions. In addition, robust policy support, increased public awareness, and education are crucial for the effective implementation of telemedicine, ultimately improving health care access and outcomes across Romania.

To fully realize the benefits of telemedicine, regulatory bodies in Romania should consider introducing mandatory education requirements for telemedicine practice. While technological proficiency is an important component of telemedicine education, health care professionals must also develop a broader range of competencies to provide effective telehealth care. These include clinical decision-making, patient communication, and

XSL•FO
RenderX

understanding the ethical and legal dimensions of telemedicine. Incorporating recognized telemedicine competency frameworks into educational programs can ensure comprehensive preparation for future telehealth practitioners.

Telemedicine is pivotal in advancing integrated care by fostering coordination among health care professionals and enabling patient-centered approaches. To fully realize its potential, telemedicine education must include interprofessional education and training for collaborative care, equipping health care teams with the skills to work cohesively in delivering seamless and effective telehealth services [30].

The results highlight a diverse range of technologies relevant to telemedicine education, including mobile health platforms, IoT devices, EHR systems, and continuous monitoring technologies. Although only a minority of reviewed studies explicitly addressed AI, its inclusion underscores the importance of preparing professionals for future technological advancements. While AI is not yet pervasive in all areas of health care, equipping professionals with foundational knowledge will prepare them for its growing integration into clinical practice.

Creating a thorough telemedicine education program necessitates teamwork between health care professionals, educators, technology specialists, and policymakers. Collaborating can help ensure the curriculum meets health care system needs, includes cutting-edge telemedicine technologies, and prepares providers and patients for future health care delivery.

Implications of Findings

To close the skills gaps, it is essential to develop educational programs that address the points mentioned below.

Improve Digital Literacy

Training must concentrate on enhancing the digital skills of health care professionals and patients so they can successfully navigate eHealth and mobile health platforms. While improving, digital literacy among Romanian health care professionals remains largely dependent on voluntary initiatives such as webinars, workshops, and seminars. This fragmented approach highlights the need for structured and comprehensive education programs to ensure that professionals are fully equipped to leverage telemedicine technologies effectively.

Integrates AI Technologies

Educational initiatives must cover foundational AI and machine learning (ML) concepts relevant to health care applications, including data privacy, ethical considerations, and interpreting AI-generated insights.

Develop New Skills

Training for telemedicine should include soft skills such as remote patient interaction, digital communication etiquette, and managing online patient relationships to adapt to remote health care dynamics. It also includes educating patients on how to use telemedicine services, getting ready for virtual visits, and handling their health data online.

Promote Trust in the Efficacy of Telemedicine

The research suggests the importance of increasing trust among health care providers and patients regarding telemedicine's effectiveness. For that purpose, telemedicine training should cover not only the technical aspects but also the clinical relevance and influence on patient outcomes, especially in managing chronic conditions.

Deal With Ethical and Privacy Concerns

Telemedicine training should incorporate adherence to recognized telehealth and telemedicine standards and guidelines, such as GDPR (General Data Protection regulations) and those created by the International Society for Telemedicine and eHealth [31]. Training should address ethical issues, privacy concerns, legal regulations, patient privacy protection strategies, and ethical guidelines for virtual patient interactions, ensuring alignment with global best practices.

Limitations and Future Directions

This scoping review, while comprehensive, has certain limitations. The focus on English and Romanian language publications may have excluded relevant studies in other European languages. In addition, the rapid evolution of telemedicine, particularly in response to the COVID-19 pandemic, means that some recent developments may not be fully captured in the published literature.

Future research should focus on (1) longitudinal studies assessing the long-term impact of telemedicine education on health care delivery and outcomes in Romania; (2) comparative analyses of telemedicine education approaches across different European countries, particularly comparing Eastern and Western European contexts; and (3) in-depth qualitative studies exploring the experiences and perspectives of medical students, health care providers, and patients regarding telemedicine education and implementation in the Romanian and broader international context.

Conclusion

Despite significant advancements, telemedicine in Romania still faces challenges. Infrastructure deficiencies, digital literacy gaps, and regulatory hurdles remain significant obstacles. However, ongoing investments in infrastructure, education, and regulatory frameworks are expected to address these issues, paving the way for broader adoption of telemedicine and improved health care access across the country.

The integration of telemedicine into medical education in Romania is crucial for the future of health care delivery. By addressing the current challenges and learning from successful global models, Romania can enhance its telemedicine capabilities and ensure that health care providers are well-prepared to leverage telemedicine technologies to improve health care access and quality. Moving forward, efforts should focus on enhancing digital health literacy, optimizing telemedicine systems, and expanding the use of AI and IoT for more integrated health care.

The future of telemedicine in Romania looks promising. As the country continues to invest in telemedicine education and infrastructure, health care providers will be better prepared to

XSL•FO
RenderX

leverage digital health technologies, ultimately enhancing the quality and accessibility of health care services for all Romanians.

Acknowledgments

We would like to acknowledge Victor Babes University of Medicine and Pharmacy Timisoara for their support in covering the costs of publication for this research paper. Part of the activities described in this manuscript were supported by funding from the EU Erasmus+ program under the TEAM project.

Conflicts of Interest

None declared.

Checklist 1
PRISMA-ScR checklist.
[PDF File, 349 KB - mededu_v11i1e66458_app1.pdf]

References

1. Panait L, Doarn CR, Saftoiu A, Popovici C, Valeanu V, Merrell RC. A review of telemedicine in Romania. J Telemed Telecare 2004;10(1):1-5. [doi: 10.1258/135763304322764103] [Medline: 15006207]
2. Romanian Government. Decision Nr. 1133 from 14 September 2022. URL: https://legislatie.just.ro/Public/DetaliiDocument/259367 [accessed 2023-12-15]
3. Eurostat. Individuals – internet activities: seeking health information. URL: https://ec.europa.eu/eurostat/databrowser/view/ISOC_CI_AC_I__custom_4264326/default/table?lang=en [accessed 2023-12-15]
4. Coman MA, Forray AI, Van den Broucke S, Chereches RM. Measuring health literacy in Romania: validation of the HLS-EU-Q16 survey questionnaire. Int J Public Health 2022;67:1604272. [doi: 10.3389/ijph.2022.1604272] [Medline: 35185446]
5. Cretu S, Gorzko AM, Salavastru CM. Openness toward the use of telemedicine among medical students in Romania: A cross-sectional study. JAAD Int 2022 Sep;8:139-141. [doi: 10.1016/j.jdin.2022.06.011] [Medline: 35879930]
6. Ghilencea LN, Chiru MR, Stolcova M, et al. Telemedicine: benefits for cardiovascular patients in the COVID-19 Era. Front Cardiovasc Med 2022;9:868635. [doi: 10.3389/fcvm.2022.868635] [Medline: 35935629]
7. Andronic O, Petrescu GED, Artamonov AR, et al. Healthcare professionals' specialists' perception of telemedicine in romania-a quantitative study of beliefs, practices, and expectations. Healthcare (Basel) 2023 May 25;11(11):1552. [doi: 10.3390/healthcare11111552] [Medline: 37297692]
8. Pescaru CC, Crisan AF, Marc M, et al. A systematic review of telemedicine-driven pulmonary rehabilitation after the acute phase of COVID-19. J Clin Med 2023 Jul 24;12(14):4854. [doi: 10.3390/jcm12144854] [Medline: 37510969]
9. Pap IA, Oniga S. A review of converging technologies in eHealth pertaining to artificial intelligence. Int J Environ Res Public Health 2022 Sep 10;19(18):11413. [doi: 10.3390/ijerph191811413] [Medline: 36141685]
10. Mocan B, Mocan M, Fulea M, Murar M, Feier H. Home-based robotic upper limbs cardiac telerehabilitation system. Int J Environ Res Public Health 2022 Sep 15;19(18):11628. [doi: 10.3390/ijerph191811628] [Medline: 36141899]
11. Taralunga DD, Florea BC. A blockchain-enabled framework for mHealth systems. Sensors (Basel) 2021 Apr 16;21(8):2828. [doi: 10.3390/s21082828] [Medline: 33923842]
12. Turcian D, Stoicu-Tivadar V. Artificial intelligence in primary care: an overview. Stud Health Technol Inform 2022 Jan 14;289:208-211. [doi: 10.3233/SHTI210896] [Medline: 35062129]
13. Priescu I, Oncioiu I. Measuring the impact of virtual communities on the intention to use telemedicine services. Healthcare (Basel) 2022 Sep 4;10(9):1685. [doi: 10.3390/healthcare10091685] [Medline: 36141297]
14. Tudor AIM, Nichifor E, Litră AV, Chi u IB, Brătucu TO, Brătucu G. Challenges in the adoption of eHealth and mHealth for adult mental health management-evidence from Romania. Int J Environ Res Public Health 2022 Jul 27;19(15):9172. [doi: 10.3390/ijerph19159172] [Medline: 35954526]
15. Iliuta L, Andronesi AG, Panaitescu E, Rac-Albu ME, Scafa-Udri te A, Moldovan H. Challenges for management of dilated cardiomyopathy during COVID-19 pandemic-a telemedicine application. J Clin Med 2022 Dec 14;11(24):7411. [doi: 10.3390/jcm11247411] [Medline: 36556027]
16. Florea M, Lazea C, Gaga R, et al. Lights and shadows of the perception of the use of telemedicine by Romanian family doctors during the COVID-19 pandemic. Int J Gen Med 2021;14:1575-1587. [doi: 10.2147/IJGM.S309519] [Medline: 33953605]
17. Patrascu R, Albai A, Braha A, et al. Instrument for assessing patients' desirability, acceptability, and adherence to telemedicine in diabetes: development, validity, and reliability. Patient Prefer Adherence 2021;15:2705-2713. [doi: 10.2147/PPA.S343869] [Medline: 34898981]

XSL•FO
RenderX

18. Giannouli V, Stoyanova S, Drugas M, Ivanova D. Attitudes towards eHealth during the COVID-19 pandemic: untangling the gordian Knot in Greece, Bulgaria and Romania in healthcare professionals and students? Psychiatr Danub 2021 Dec;33(Suppl 13):415-419. [Medline: 35150518]
19. Iliu ă L, Andronesi AG, Rac-Albu M, et al. Challenges in caring for people with cardiovascular disease through and beyond the COVID-19 pandemic: the sdvantages of universal access to home telemonitoring. Healthcare (Basel) 2023 Jun 12;11(12):1727. [doi: 10.3390/healthcare11121727] [Medline: 37372846]
20. Lăcraru AE, Busnatu S, Pană MA, et al. Assessing the efficacy of a virtual assistant in the remote cardiac rehabilitation of heart failure and ischemic heart disease patients: case-control study of Romanian adult patients. Int J Environ Res Public Health 2023 Feb 22;20(5):3937. [doi: 10.3390/ijerph20053937] [Medline: 36900948]
21. Vizitiu C, Bîră C, Dinculescu A, Nistorescu A, Marin M. Exhaustive description of the system architecture and prototype implementation of an IoT-based eHealth biometric monitoring system for elders in independent living. Sensors (Basel) 2021 Mar 6;21(5):1837. [doi: 10.3390/s21051837] [Medline: 33800728]
22. Anghelescu A. Telerehabilitation: a practical remote alternative for coaching and monitoring physical kinetic therapy in patients with mild and moderate disabling Parkinson's disease during the COVID-19 Pandemic. Parkinsons Dis 2022;2022:4370712. [doi: 10.1155/2022/4370712] [Medline: 35979169]
23. Giunti G, Guisado-Fernandez E, Belani H, Lacalle-Remigio JR. Mapping the access of future doctors to health information technologies training in the European union: cross-sectional descriptive study. J Med Internet Res 2019 Aug 12;21(8):e14086. [doi: 10.2196/14086] [Medline: 31407668]
24. Center of Innovation and e-Health – UMFCD. Telemedicine - current information and skills. 2023. URL: https://cieh.umfcd.ro/2023/09/22/Telemedicina-info [accessed 2025-05-01]
25. Digital Innovation Zone. Telemedicine specialist curriculum. 2024. URL: https://digital-innovation.zone/cultura-si-competente-digitale/telemedicine-specialist/ [accessed 2025-05-01]
26. RoHEALTH webinar. 2024. URL: https://rohealth.ro/detalii-curs/37 [accessed 2025-05-01]
27. Telemed. 2024. URL: https://www.telemed.org.ro/comunicat-de-presa-lansare-proiect.html [accessed 2025-05-01]
28. EHTEL European health telematics association. Vision and mission. 2025. URL: https://ehtel.eu/about/vision-and-mission.html [accessed 2025-05-01]
29. European Commission. Digital education action plan. URL: https://education.ec.europa.eu/focus-topics/digital-education/action-plan [accessed 2025-05-01]
30. Interprofessional education collaborative. IPEC core competencies for interprofessional collaborative practice: version 3. : Interprofessional Education Collaborative; 2023.
31. ISfTeH education program. URL: https://www.isfteh.org/education/isfteh_education_program [accessed 2025-01-25]

Abbreviations

AI: artificial intelligence
ATASS: Association for Telemedicine and Space Applications for Health
CME: Continuing Medical Education
DHCF: Digital Health Competency Framework
EHR: electronic health record
EHTEL: European Health Telematics Association
GDPR: General Data Protection regulations
IoT: Internet of Things
ISfTeH: International Society for Telemedicine & eHealth
ROSA: Romanian Space Agency
WHO: World Health Organization

Edited by B Lesselroth; submitted 23.09.24; peer-reviewed by K Drude, T Vagg; revised version received 01.02.25; accepted 25.02.25; published 14.05.25.

Please cite as:
Focsa MA, Rotaru V, Andronic O, Marginean M, Florescu S
Bridging Gaps in Telemedicine Education in Romania to Support Future Health Care: Scoping Review
JMIR Med Educ 2025;11:e66458
URL: https://mededu.jmir.org/2025/1/e66458
doi:10.2196/66458

XSL•FO
RenderX

© Mircea Adrian Focsa, Virgil Rotaru, Octavian Andronic, Marius Marginean, Sorin Florescu. Originally published in JMIR Medical Education (https://mededu.jmir.org), 14.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Virtual Simulation Tools for Communication Skills Training in Health Care Professionals: Literature Review

Manuel Fernández-Alcántara[1,2*], PhD; Silvia Escribano[2,3*], PhD; Rocío Juliá-Sanchis[2,3*], PhD; Ana Castillo-López[3*]; Antonio Pérez-Manzano[4*], PhD; M Macur[5*], PhD; Sedina Kalender-Smajlović[5*], PhD; Sofía García-Sanjuán[2,3*], PhD; María José Cabañero-Martínez[2,3*], PhD

[1]Department of Health Psychology, Faculty of Health Sciences, University of Alicante, Alicante, Spain
[2]Institute of Health and Biomedical Research of Alicante, Alicante, Spain
[3]Department of Nursing, Faculty of Health Sciences, University of Alicante, Carretera San Vicente del Raspeig s/n, Alicante, Spain
[4]University of Murcia, Murcia, Spain
[5]Angela Boškin Faculty of Health Care, Spodnji Plavž 3, Jesenice, Slovenia
[*]all authors contributed equally

Corresponding Author:
Sofía García-Sanjuán, PhD
Institute of Health and Biomedical Research of Alicante, Alicante, Spain

Abstract

Background: Quality clinical care is supported by effective patient-centered communication. Health care professionals can improve their communication skills through simulation-based training, but our knowledge about virtual simulation and its effectiveness and use in training health professionals and students is still growing rapidly.

Objective: The objective of this study was to review the current academic literature to identify and evaluate the virtual simulation tools used to train communication skills in health care students and professionals.

Methods: This review was carried out in June 2023 by collecting data from the MEDLINE/PubMed and Web of Science electronic databases. Once applicable studies were identified, we recorded data related to type of technology used, learning objectives, degree of learning autonomy, outcomes, and other details.

Results: We found 35 articles that had developed and/or applied a virtual environment for training communication skills aimed at patients, in which 24 different learning tools were identified. Most had been developed to independently train communication skills in English, either generally or in the specific context of medical history (anamnesis) interviews. Many of these tools used a virtual patient that looked like a person and had the ability to vocally respond. Almost half of the tools analyzed allowed the person being trained to respond orally using natural language. Of note, not all these studies described the technology they had used in detail.

Conclusions: Many different learning tools with very heterogeneous characteristics are being used for the purposes of communication skills training. Continued research will still be required to develop virtual tools that include the most advanced features to achieve high-fidelity simulation training.

(JMIR Med Educ 2025;11:e63082) doi:10.2196/63082

KEYWORDS

communication skills; virtual patient; virtual simulation; health care professionals; virtual simulation tool; skill training; communication; heterogeneous; heterogeneous characteristics; virtual tool; patient-centered; patient-centered communication; implementation

Introduction

Effective patient-centered communication is one of the key components of quality clinical care [1]. Thus, it is vital that health care professionals adequately manage their communication skills. This involves mastering the transmission of information; listening and comprehensively understanding all the issues related to the health of each patient [2]; and responding appropriately to the physical and emotional needs of patients [3]. Better communication when supporting decision-making means that patients are better able to understand their situation, feel better informed, and are more active in the decision-making process [4]. Hence, acquiring good communication skills has been related to improved health outcomes, general patient satisfaction [5], better adherence to treatment plans [6], and positive effects on health care costs and length of hospital stay [7].

XSL•FO
RenderX

However, despite recognizing the importance of communication, health professionals are not always sufficiently skilled in this area [8]. Therefore, it is advisable that both health and educational institutions introduce different means of supporting the development of communication skills into their training plans as a priority objective. Furthermore, this training must also be implemented through effective educational strategies [9]. It has previously been shown that simulation-based learning is an effective means of acquiring communication skills [9]. Specifically, simulation with a standardized or simulated patient, which consists of using trained people to realistically portray a patient within learning contexts [10], is widely used to train communication skills [1].

Nonetheless, although the use of simulation methodologies has greatly advanced training in communication skills, its implementation also has limitations. For example, in terms of the human resources used in this type of training, it is particularly difficult to recruit actors able to simulate patients precisely and consistently in a completely standardized way [11,12]. Other difficulties include temporal–spatial issues because the availability of simulations with standardized patients is limited to a specific physical space and time [13]. A training alternative that could overcome these limitations is the use of standardized virtual patient programs that use computerized characters rather than real actors [14].

Indeed, compared to standardized patients, there are significant advantages to the use of virtual patients, including the need for fewer staff and resources once developed [15], unlimited availability, and the fact that they are highly customizable [14]. Additionally, these tools provide highly interactive, engaging, and more standardized experiences because educators can control their design, programming, delivery, and use [14]. It is also worth noting that these solutions can be personalized according to specific individual needs, given that they are not limited by time or space, so students can repeatedly engage in training in more clinical scenarios than is possible through traditional methods [15]. In addition, this technology also allows students to learn in a safe environment with low levels of risk and anxiety, which encourages them to gain greater personal awareness of their learning processes [16].

Virtual simulation has gained attention in recent years as a promising tool for training both undergraduate and graduate students, as well as health care professionals, in various competencies, including nontechnical skills. This growing interest is evident in an increasing number of studies focused on its potential applications in health care education [17]. However, despite this expanding body of research, it is advisable to continue researching with the aim of fully exploring and understanding which technological and technical skills are more suitable to train in virtual simulation [17]. Some reviews on virtual simulation and the learning of nontechnical skills such as communication are available [17-19]. For example, in their integrative review, Peddle et al [19] examined how interactions with virtual patients impacted nontechnical skills in general, without exclusively focusing on communication skills or technical and instructional design characteristics. Subsequently, both the systematic review by Lee et al [18] and the literature review by Battegazzorre et al [17] examined the technical characteristics of virtual learning applications aimed at improving communication skills. However, it is noteworthy that these reviews include studies published only up to December 2018 and May 2020, respectively, which highlights a gap in the literature regarding recent advancements in virtual simulation technologies.

The development of communication skills is fundamental for the effective clinical practice of health care professionals. However, the increasing diversity of virtual simulation tools and the rapid pace of technological innovation pose significant challenges to understanding which tools are most effective for training these skills. This raises the following key questions: what are the characteristics of the current virtual simulation tools used for training communication skills, and how effective are they in fostering realistic and immersive learning experiences? To address these questions, we conducted a systematic review of the virtual simulation tools available to train communication skills in health care professionals, analyzing their design, degree of immersion, and autonomy to identify their strengths and limitations.

Therefore, the objective of this study was to review the current academic literature to identify and evaluate the virtual simulation tools used to train communication skills in health care students and professionals and to assess their effectiveness and limitations in training health care personnel.

Methods

Design

We completed a literature review to identify virtual simulation tools designed to train communication skills in health care professionals, including students in training and practicing professionals. The inclusion criteria were studies that examined (1) virtual simulation tools and/or those based on artificial intelligence (AI), (2) tools used to train communication skills in health professionals, and (3) tools targeting training in communication skills and/or therapeutic relationships with patients. Studies were excluded if (1) the tools were designed to train interprofessional communication, (2) the objective was noneducational, and (3) the tool was designed to train patients in social and/or communication skills. This systematic review was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 [20] guidelines to ensure comprehensive and transparent reporting of the methodology and findings.

Search Strategy

The search for studies was conducted in June 2023 in the MEDLINE/PubMed and Web of Science electronic databases. As part of the search strategy, we consulted the PubMed thesaurus using the following Medical Subject Headings (MeSH) terms: “Artificial Intelligence,” “Machine learning,” “virtual reality,” and “social skills.” The natural language search terms included in the title and/or abstract fields were “artificial intelligence,” “machine learning,” “virtual reality,” “e-simulation,” “web-based simulation,” “virtual simulation,” “virtual patient,” “social skills,” “interpersonal skills,” “social ability,” “social competences,” and “communication skills.”

XSL•FO
RenderX

The complete search strategy was as follows: (((“Artificial Intelligence”[MeSH Terms] OR “Machine Learning”[MeSH Terms] OR “Artificial Intelligence”[Title/Abstract] OR “Machine Learning”[Title/Abstract])) OR ((“Virtual Reality”[MeSH Terms] OR “Virtual Reality”[Title/Abstract] OR “e-simulation”[Title/Abstract] OR “web-based simulation”[Title/Abstract] OR “virtual simulation”[Title/Abstract]) OR (“virtual patient”[Title/Abstract]))) AND ((“Social Skills”[MeSH Terms] OR “Social Skills”[Title/Abstract] OR “interpersonal skills”[Title/Abstract] OR (“social ability”[Title/Abstract] OR “social abilities”[Title/Abstract]) OR (“social competence”[Title/Abstract] OR “social competences”[Title/Abstract]) OR “communication skills”[Title/Abstract])).

No temporal restrictions were applied in any of these cases. Despite previous reviews focusing on similar topics [17-19], it was decided not to base the current review on them. This decision was due to differences in the search strategy used, which did not account for the wide range of synonyms associated with each term established for this review. Furthermore, it is important to note that Lee et al [18] focused their strategy exclusively on communication among medical students, while Peddle et al [19] directed their attention to all nontechnical skills, not just communication skills.

The eligibility of the studies was independently assessed by 2 of the authors (MJCM and RJS) and any discrepancies were resolved by another author (SE).

Data Extraction

Data related to the characteristics of the studies (publication year, country, language, objective, and type) as well as data related to the outcome of the use of the digital/virtual training tool for improving communication skills in health care professionals were recorded. Specifically, we noted the tool name, training language, learning objective, degree of learning autonomy (fully autonomous vs instructor-mediated training), patient type (avatar/doll, virtual patient with a human-like appearance, real person, etc), type of answers given by the trainee (written or oral conversation), and type of technology used.

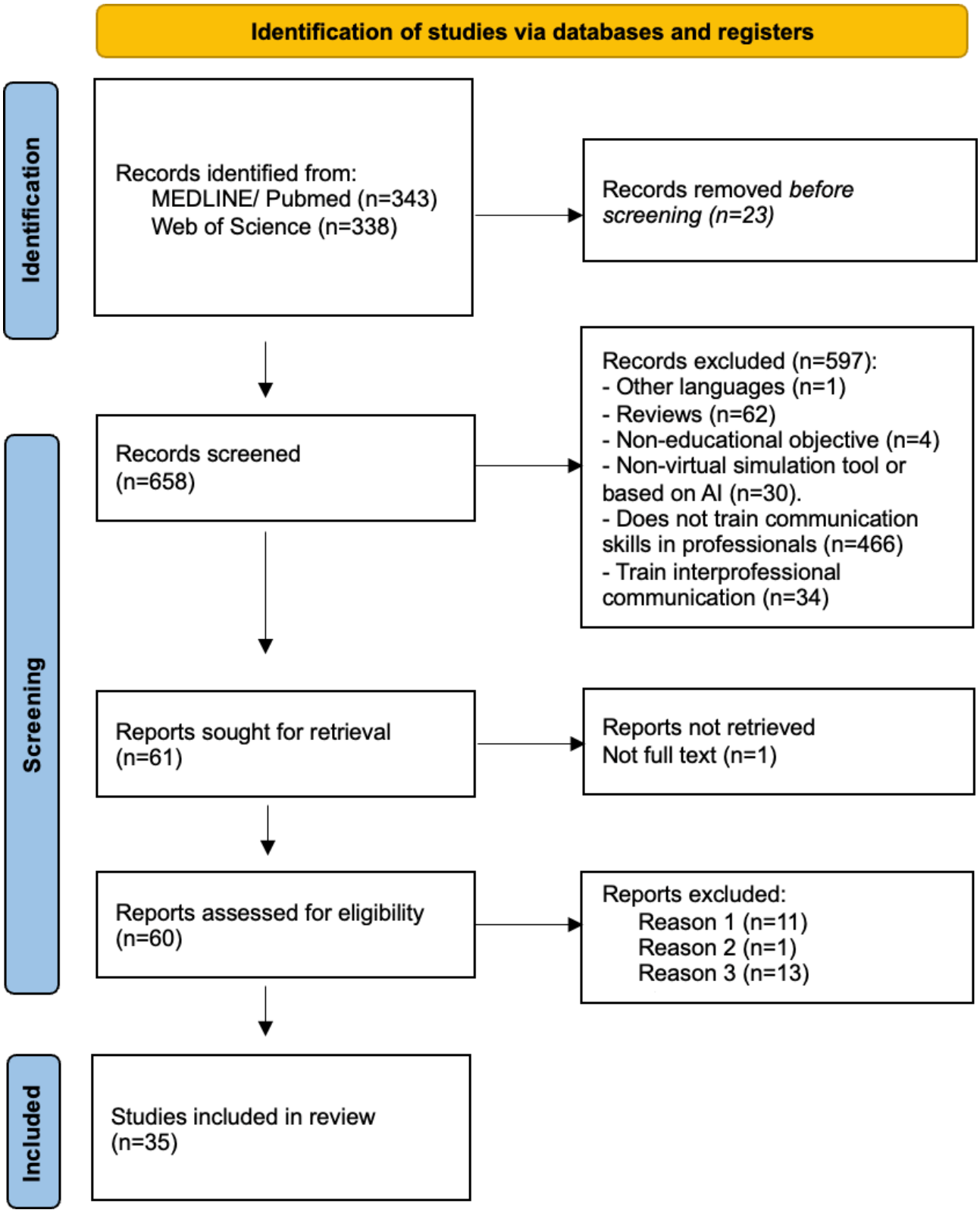
Results

Overview

The studies were manually screened and coded. Our search of PubMed and the Web of Science produced 681 records, of which 23 duplicates were eliminated. During the screening process, 2 of the authors independently analyzed 658 studies based on their titles and abstracts (Figure 1). After this initial screening, the full texts of 61 records were obtained for analysis. We requested the full texts of a further 2 articles from the corresponding authors by email and through ResearchGate; of these, we included 1 in this review. Of these 60 studies, 25 were excluded because they did not meet the inclusion criteria. Specifically, 11 articles had not directly trained clinical communication skills with patients (criterion 1), 1 had not studied virtual training (criterion 2), and 13 had not used a tool designed for training purposes (criterion 3). Therefore, a total of 35 articles were included in the review. Finally, one of the authors extracted the relevant data from these 35 studies and entered them into a database following the coding manual we had prepared for this purpose.

XSL•FO
RenderX

Figure 1. PRISMA flowchart. Reason 1: articles not directly related to training clinical communication skills with patients; reason 2: did not study virtual training; reason 3: did not use a tool designed for training purposes. AI: artificial intelligence; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



Characteristics of the Studies Included

A total of 35 articles were obtained that had developed and/or applied a virtual environment for training communication skills that would be directed toward patients; overall, 43% (n=15) were articles published in the United States and 17% (n=6) were from Australia, with the remaining articles having been published in Europe and Asia (Table 1). All the articles had been published in English and their objectives are shown in Table 1.

XSL•FO
RenderX

Table . Description of the studies (N=35).

Articles	Country	Language	Objective
Ali et al [21], 2020	United States	English	Describe the iterative participatory design of SOPHIE, an online virtual patient for "practice" based on feedback from sensitive conversations between patients and clinicians and discuss an initial qualitative evaluation of the system by professional end users.
Bánszki et al [22], 2018	Australia	English	Explore a novice clinical educator's experience in training essential communication and interpersonal skills using a virtual patient.
Bearman and Cesnik [23], 2001	Australia	English	Assess students' attitudes toward learning communication skills through a virtual patient; compare the acceptability of the 2 distinct types of virtual patient designs.
Bearman et al [24], 2001	Australia	English	Compare 2 types of virtual patients to understand how different virtual patient designs affect the student learning experience.
Bearman [25], 2003	Australia	English	Explore the students' experience with the virtual patient.
Borja-Hart et al [26], 2019	United States	English	Assess students' confidence and impressions when using their communication skills with a virtual patient and evaluate their competencies in the use of this technology.
Chae et al [27], 2023	Korea	English	The purpose of this study was to describe the development of SimCARE and evaluate the feasibility of its use in nursing education.
Courteille et al [28], 2014	Sweden	English	To investigate the dynamics and congruence of interpersonal behaviors and socioemotional interaction exhibited during the learning experience in a virtual patient, and to evaluate which interaction design features contribute most to behavioral and affective engagement in the medical student.
Deladisma et al [29], 2008	United States	English	Develop a virtual training environment system that can be accessed independently.
Dickerson et al [30], 2006	United States	English	Provide information about the advantages and disadvantages of using synthesized speech and evaluate the fidelity necessary for the training of communication skills.
Du et al [31], 2022	China	English	To evaluate the history-taking skills of nursing undergraduates using a virtual standardized patient, and to explore its independent influencing factors.
Guetterman et al [32], 2019	United States	English	To investigate the differential effects of a virtual patient–based simulation developed to train health care professionals in empathetic patient-provider and interprofessional communication.

XSL•FO
RenderX

Articles	Country	Language	Objective
Hwang et al [33], 2022	Taiwan, Japan	English	A virtual patient–based social learning approach is proposed to enhance nursing students' performance and clinical judgment in education programs.
Jacklin et al [34], 2018	United Kingdom	English	Create a virtual patient that simulates a primary care consultation, offering the opportunity to practice decision-making. A second objective was to involve patients in the design of a virtual patient simulation and inform the design process.
Jacklin et al [35], 2021	United Kingdom	English	This study aims to evaluate a virtual patient workshop for medical students aimed at developing the communication skills required for shared decision-making.
Kleinsmith et al [2], 2015	United States	English	Develop an understanding of whether students can respond empathically to expressions of concern from a virtual patient.
Lok [36], 2006	United States	English	Teach communication skills using virtual humans.
Maicher et al [37], 2019	United States	English	Describe a virtual standardized patient system that allows students to practice their history-taking skills and receive immediate feedback.
Mayor Silva et al [38], 2023	Spain	English	The objective was to develop a virtual reality simulator to improve communication skills and compare its results with a traditional workshop based on cases and theoretical content explained through video.
Nakagawa et al [39], 2022	Japan	English	The objective structured clinical examination is among validated approaches used to assess clinical competence through structured and practical evaluation.
Ochs et al [40], 2019	France	English	Evaluate the virtual reality training platform in which the user experience is analyzed based on the virtual environment.
Perez et al [41], 2022	United States	English	The purpose of this study was to explore the use of virtual simulation to experience difficult conversations and to evaluate differences in perceptions between nurse educator, family nurse practitioner, and nurse anesthesia students.
Plass et al [42], 2022	Germany	English	The purpose of this study is to evaluate the effectiveness of a brief virtual role-play motivational interviewing training program on motivational interviewing knowledge and skills in first-year undergraduate medical students, making use of both a pre-test and a then-test (retrospective pre-test) to check for response shift in evaluating the educational intervention.

XSL•FO
RenderX

Articles	Country	Language	Objective
Quail et al [12], 2016	Australia	English	Investigate students' communication skills, knowledge, confidence, and empathy in simulated and traditional learning environments.
Real et al [43], 2017	United States	English	Develop an immersive virtual reality curriculum on addressing flu vaccine hesitancy using Kern's 6-step approach to curriculum design. The goal of the program was to teach best communication practices in cases of questions about the flu vaccine.
Real et al [44], 2017	United States	English	Create an immersive virtual reality curriculum to teach pediatric residents communication skills when discussing flu vaccination. Compare effectiveness with a control group.
Real et al [45], 2022	United States	English	Examined the acceptability and tolerability of the approach and the impact of deliberate practice using virtual reality simulations on clinicians' confidence related to shared decision-making communication skills.
Rouleau et al [46], 2022	Canada	English	This study aimed to assess the acceptability of a virtual patient simulation to improve nurses' relational skills in a continuing education context.
Sapkaroski et al [47], 2022	Australia	English	The aim of this study was to establish whether the mode of delivery, virtual reality simulated learning environments versus clinical role-play, could have a measurable effect on clinical empathic communication skills for magnetic resonance imaging scenarios.
Sezer and Sezer [48], 2019	Turkey	English	Design, develop, and evaluate a 3D virtual patient application that can move, has voice and lip synchronization, allows written communication, and is supported by a solid scenario to improve students' communication skills.
Şimşek Çetinkaya et al [49], 2022	Turkey	English	This study aimed to determine the effectiveness of 2 simulation types used for family planning consultation of midwifery students and to compare these methods.
Shorey et al [50], 2019	Singapore	English	Develop and evaluate the use of virtual patients to better prepare undergraduate nursing students to communicate with real-life patients, their families, and other health care professionals during their clinical stays.
Shorey et al [51], 2020	Singapore	English	To examine user attitudes and experiences and clinical facilitators' perspectives on student performance in the clinical environment following virtual patient training.

XSL•FO
RenderX

Articles	Country	Language	Objective
Shorey et al [52], 2023	United States	English	This study aimed to evaluate the effectiveness of this theory-based virtual intervention on nursing students' learning attitudes, communication self-efficacy, and clinical performance.
Stevens et al [53], 2006	United States	English	Create an interactive virtual clinical scenario of a patient with acute abdominal pain to teach medical students history-taking and communication techniques.

Features of the Virtual Tools

After reading the full text of the 35 articles, we identified 24 different learning tools that had been developed to train communication skills in students or health professionals (Table 2). Most of them (n=15; 62%) had provided training in English [2,21,22,24,26,28,29,32,34,37,41,43,46,47,52]. Regarding the learning objective of the virtual environment, 42% (n=10) aimed to train communication skills in the specific contexts of a clinical history and/or anamnesis interview [2,29,31,33,35,37,42,46,48,52], 42% (n=10) taught general communication skills [22,24,26-28,38,39,41,47,49], and 8% (n=2) covered giving bad news [21,40]. There was also a tool that had been specifically developed to train communication skills to address flu vaccination hesitancy [43-45]. Another tool that had been used to train communication skills focused on empathy is also worth highlighting [32].

XSL•FO
RenderX

Table . Virtual tools and their characteristics (n=24 tools).

Articles	Tool name	Language	Study purpose	Degree of learning autonomy	Patient type	Type of student responses during training	Type of technology used
Ali et al [21], 2020	SOPHIE	English	Train communication skills for the delivery of bad news. Aimed at health professionals.	Autonomous	Virtual patient with the appearance of a person. Responded with a voice. The entire transcript can be seen.	Oral conversation	Artificial intelligence
Bánszki et al [22], 2018; Quail et al [12], 2016	Not specified	English	Training communication skills. Aimed at health care students.	An instructor mediated the training.	Virtual patient with the appearance of a person. Responded with a voice.	Oral conversation	The instructor was in another room where they controlled everything and responded in the simulated interaction.
Bearman and Cesnik [23], 2001; Bearman et al [24], 2001; Bearman [25], 2003	Not specified	English	Training in communication skills. Aimed at medical students.	Autonomous	Real person speaking. Viewing of recorded videos.	Written. Choice of 3 or 4 written response options available after each video. The authors developed 2 types of responses to compare which was more effective: narrative (detailed communicative structures) or problem-solving (labels with possible actions).	A total of 154 recorded videos. The next video shown was adjusted depending on the response given. Therefore, the virtual patient became satisfied according to responses chosen by the student.
Borja-Hart et al [26], 2019	Used *Shadow Health* from Elsevier	English	Training in communications skills. Aimed at pharmacy students.	Autonomous	Virtual patient with the appearance of a person. Responded with a voice.	Natural language (written and spoken). Students could choose the interaction they would carry out: ask, empathize, or educate.	*Shadow Health* is simulation software that generates different scenarios. The article did not explain any more about the technology used.
Chae et al [27], 2023	SimCARE	Korean	Training in intercultural communication skills. Aimed at nursing students.	Autonomous	Virtual patient with the appearance of a person. Responded with a voice.	They selected a written response from among those on offer.	A virtual reality headset. The authors described the technology used to generate the 3D graphics (Unity 2019.4.0f1 game engine), avatars (DAZ 3D software), and avatar animation (iClone 7).

XSL•FO
RenderX

Articles	Tool name	Language	Study purpose	Degree of learning autonomy	Patient type	Type of student responses during training	Type of technology used
Courteille et al [28], 2014	Not specified	English and Swedish	Training in communication skills. Aimed at medical students.	Autonomous	Real person speaking. Viewing of recorded videos.	Written. Students replied in text written in natural language.	Interactive Simulation of Patients. A database with 200 videos for each case, allowing the simulator to respond according to the question posed by the student.
Deladisma et al [29], 2008; Dickerson et al [30], 2006; Lok [36], 2006; Stevens et al [53], 2006	Not specified	English	Training in communication skills and anamnesis techniques. Aimed at medical students.	Autonomous but with availability of additional resources. The technology that drives this interaction largely consisted of commodity hardware and software: 2 desktop computers, 2 cameras, a data projector, and a wireless microphone.	Virtual patient with the appearance of a person (an avatar called Diana) who spoke and produced natural gestures. The authors developed 2 types of communication for the avatar to study which one was more effective: real recorded communication or virtual communication.	Oral conversation. The students could speak using natural language. The software also detected various gestures.	The speech recognition worked using *Dragon Naturally Speaking* by Scansoft, which is a database developed with content organized in semantic categories to detect the communicative structures used by the students.
Du et al [31], 2022	University A Virtual Patient (UA-VP, 2021)	Chinese	Training in communication skills to carry out a nursing evaluation by following Gordon's Functional Patterns.	Autonomous	A virtual patient with the appearance of a person. Responded with text based on a predefined chat.	Written and oral conversation	Recognizes structures and offers feedback based on the uploaded chat scripts (as bullet points and not reflecting the most important part of the interaction). Used WeChat, a social media app.
Guetterman et al [32], 2019	Used MPathic-VR	English	Trained empathic communication skills. Aimed at medical students.	Autonomous	Virtual patient with the appearance of a person. Responded with a voice.	Oral conversation. It also detected gestures and movements.	Artificial intelligence.
Hwang et al [33], 2022	Not specified	Chinese	Trained students in diagnosis and treatment and has a specific medical history module which trains communication skills.	Autonomous	Virtual patient with the appearance of a person. Responded with voice and text.	Did not specify	Learning system designed as a decision tree.

XSL•FO
RenderX

Articles	Tool name	Language	Study purpose	Degree of learning autonomy	Patient type	Type of student responses during training	Type of technology used
Jacklin et al [34], 2018; Jacklin et al [35], 2021	Not specified	English	Training in communication skills for shared decision-making during clinical interviews. Aimed at medical and/or pharmacy students.	Autonomous	Virtual patient with the appearance of a person. Responded through a voice and with gestures.	Written text. Choice of 3 answer options.	A web-based virtual patient simulator.
Kleinsmith et al [2], 2015	Neurological Examination Rehearsal Virtual Environment	English	Trained communication skills for use during clinical interviews. Aimed at nursing students.	Autonomous	Virtual patient with the appearance of a person. A virtual patient responded with a voice and through text.	Written. The student inserted text written in natural language.	Virtual People Factory. A database used by the simulator to respond based on the student's question.
Maicher et al [37], 2019	Not specified	English	Trained skills for performing an anamnesis (to collect medical information). It does not address communicative listening strategies such as empathy. Aimed at medical students.	Autonomous	Virtual patient with the appearance of a person. Responded with voice and text.	Oral conversation. Text could also be written.	Artificial intelligence. The open-source natural language processing engine ChatScript is used for the conversion element. Unity gaming platform.
Mayor Silva et al [38], 2023	Not specified	Spanish	Training in communication skills. Aimed at nursing students.	An instructor mediated the evaluation.	Not specified	Not specified	A virtual reality headset.
Nakagawa et al [39], 2022	Not specified	Japanese	Trained communication skills such as desire suppression, expectation acceptance, facial expression, emotional communication, dominance, maintaining relationships, and dealing with disagreements. Aimed at pharmacy students.	Autonomous	A chatbot. Written and oral.	Oral conversation in natural language	Artificial intelligence. If the artificial intelligence did not detect the keywords, the conversation did not continue. There was no direct feedback.
Ochs et al [40], 2019	ACORFORMed	French	Training in the delivery of bad news. Aimed at medical practitioners (students and professionals).	Autonomous in some functions (eg, dialogue generator). In others (eg, categorizing the response and sending it to the simulator), the instructor mediated the learning.	Virtual patient with the appearance of a person. Responded with a voice.	Oral conversation	A virtual reality headset. The instructor categorized the response using a previously coded database and sent that information to the simulator.

XSL•FO
RenderX

Articles	Tool name	Language	Study purpose	Degree of learning autonomy	Patient type	Type of student responses during training	Type of technology used
Perez et al [41], 2022	Used the Mursion tool	English	Trained communication skills for use in difficult conversations. Aimed at nursing students.	Autonomous	Virtual patient with the appearance of a person. Responded with a voice.	Oral conversation in natural language.	Artificial intelligence (using the Mursion tool).
Plass et al [42], 2022	Used the Kognito Conversarion Platform	German	Training in person-centered communication skills for motivational interviewing. Aimed at medical students.	Autonomous	Virtual patient with the appearance of a person. Responded with a voice.	Select between different answer options.	Artificial intelligence (using the Kognito Conversation Platform).
Real et al [43], 2017; Real et al [45], 2022; Real et al [44], 2017	Not specified	English	Training in communication skills to inform patients about vaccination. Aimed at medical residents.	An instructor mediated the training.	Virtual patient with the appearance of a person. Responded through a voice and with gestures.	Oral conversation and natural language.	Unity gaming platform. A virtual reality headset.
Rouleau et al [46], 2022	Not specified	French, English	Training in nursing relational skills for use in motivational interviews.	Autonomous	Virtual patient with the appearance of a person. Responded with a voice.	Select between different answer options	Used the MedicActiV platform
Sapkaroski et al [47], 2022	Not specified	English	Training in communication skills. Aimed at medical students.	Autonomous	Virtual patient with the appearance of a person. Responded with voice and text.	Select from among answer options. This part of the case simulation was mandatory. It was also capable of natural language oral conversation and the ability to ask alternative questions was optional.	Clinical Education Training Solution virtual reality clinic software using the Oculus Rift CV1 virtual reality headset.
Sezer and Sezer [48], 2019	Not specified	Turkish	Training in basic communication skills for use in a medical interview. Aimed at training health care students.	Autonomous	Virtual patient with the appearance of a person. Responded with a voice and in writing.	Natural written text	Virtual People Factory for avatar and simulation generation. The scenario was created in Unity 3DTM. Different variations of the simulation interventions the students could apply at each stage were included and these answer combinations were compared to the closest preprogrammed scenario to give an answer.

XSL•FO
RenderX

Articles	Tool name	Language	Study purpose	Degree of learning autonomy	Patient type	Type of student responses during training	Type of technology used
Şimşek Çetinkaya et al [49], 2022	Not specified	Turkish	Training in communication skills for use in a family planning consultation. Aimed at midwifery students.	The instructor offered feedback after watching the simulation.	The patient type was not specified. Responded with a voice.	Oral conversation	Not specified
Shorey et al [50], 2019; Shorey et al [51], 2020; Shorey et al [52], 2023	Virtual Counselling Application using Artificial Intelligence	English	Trained basic communication skills for use in an interview. Aimed at nursing students.	Autonomous	Virtual patient with the appearance of a person. Responded with a voice and in writing.	Oral conversation in natural language	Artificial intelligence. Used the Dialogflow chatbot from Google Cloud to store and process natural language. The scenario was created in Unity 3D.

Several major virtual tools were identified in this review for training communication skills in health care professionals. SOPHIE [21] is a tool designed to train the delivery of bad news using a virtual patient that interacts through oral conversations, leveraging AI. Shadow Health [26] focuses on communication skills for pharmacy students, allowing both written and spoken interactions with a virtual patient. SimCARE [27] is a virtual reality–based tool aimed at nursing students, training intercultural communication skills through animated avatars. MPathic-VR [32] trains medical students in empathic communication, featuring virtual patients that respond with voice and detect nonverbal cues like gestures. ACORFORMed [40] trains medical practitioners in delivering bad news through virtual reality interactions with a virtual patient. Mursion [41] is designed for nursing students to practice difficult conversations using natural language processing for realistic interactions, while the Kognito Conversation Platform [42] supports motivational interviewing through person-centered communication training with virtual patients. VCAAI [50-52] trains basic communication skills in nursing interviews. These tools highlight the diversity of approaches in the use of virtual patients for communication training. Finally, 14 virtual tools did not specify their name.

Some (n=19, 79%) of the tools allowed students to train completely autonomously, whereas 21% (n=5) required an online instructor to mediate the training and respond during the interactions [22,39,40,44,49]. One of the tools could be defined as partially autonomous because a trained instructor had to perform some of the functions [40]. Regarding the patient type used for the training, the vast majority of the tools used virtual patients (n=19; 79%) with the appearance of a real person [2,21,22,26,29,31-33,35,37,40-42,44,46-48,51]. Of these, 95% (18/19) responded with a voice (18/24, 75%), except for the tool published by Du et al [46]. Two tools (8%) used videos recorded with real people [24,28].

Regarding the types of responses the user could give during the training, almost half of the tools analyzed (n=11, 45%) allowed the user to respond orally using natural language [21,22,26,29,31,32,37,39,41,44,49,51]. Shadow Health [26], for example, offers both written and spoken interactions, while SOPHIE [21] focuses solely on oral communication.

Discussion

This study reviews and analyzes the 24 virtual simulation tools available for training communication skills in health care professionals, assessing their characteristics, levels of immersion, and the autonomy they provide in learning processes. Although virtual simulation tools have shown significant growth in recent years, driven by technological advances, the review identified a high degree of heterogeneity in the approaches, technologies, and interaction methods used. This variety has made it challenging to standardize and effectively integrate these tools into consistent training plans. Most tools rely on virtual patients with a limited range of interaction capabilities, and very few offer fully immersive experiences that mimic real-world clinical communication. Furthermore, limited accessibility to tools in languages other than English, as well as a lack of high-fidelity technologies for simulating realistic, natural language–based conversations, continue to pose significant challenges. Considering these challenges, this review highlights several key findings regarding the applications of virtual environments to enhance communication skills training that will be detailed in the following paragraphs.

First, it is important to highlight the large number of different applications we identified that have been used to improve communication skills (either in basic or more specific situations) through virtual environments. Similarly, other reviews have also concluded that the use of virtual patients for clinical communication training has grown exponentially over the last decade [17,18], which has been driven by rapid technological advances [54], also providing further evidence of the benefits associated with this type of resource [18]. In fact, this work has included 13 new virtual simulation environments developed based on the published review by Battegazzorre et al [17].

XSL•FO
RenderX

Most of the applications we considered in this review used English, which could represent an obstacle for professionals and students who do not know this language. Indeed, only one of the tools identified used Spanish and in this case, it was also mediated by an instructor, thereby making it difficult for students to use it autonomously and independently [38] Therefore, there is still a long way to go to make these tools highly accessible at an international level. Regarding the more technical characteristics, we observed visible heterogeneity in the types of technologies used, including in the different types of patients used for training—for example, the use of chatbots, images, and/or recordings of real people and virtual patients. However, our results showed that almost all the applications we identified had designed virtual environments using virtual patients that looked like a person and could vocally respond to and receive oral responses to simulate a real conversation [21,22,26,29,32,37,39,41,44,49,51]. A key implementation across the tools was the use of natural language processing to simulate realistic conversations.

Training in simulation environments that assume an appropriate level of fidelity (a 3D term that includes physical/environmental, psychological, and conceptual elements) increases realism [55] and influences learning engagement [56]. For example, in their systematic review, Kaplonyi et al [1] reflected how simulations with the use of standardized patients are considered realistic environments and an effective means for learning communication skills. Indeed, the academic literature proposes that virtual patients can be used as a complementary alternative to working with standardized patients [57] and can represent patients in a realistic clinical environment [17] to effectively help students to acquire or improve their communication skills [18]. Nonetheless, it will be important for future lines of research to use standardized tests to evaluate the beneficial effects of training with this type of virtual tool before fully integrating them into training plans [18,54].

In terms of the fidelity of these tools, increasing the immersion of virtual simulations—defined as the psychological state of the perception of being inside or surrounded by something [58]—by using virtual patients with natural language processing and auditory and visual behavior [17,59] is positively related to better communication skills performance [17,19]. However, we must not forget that realism and authenticity, which are both relevant factors in design, are not only achieved through physical resemblance (physical fidelity) but also require other fidelity factors [19]. Hence, future research in this field should be designed to also consider conceptual fidelity (scenarios and cases consistent with reality) and psychological fidelity (the ability to provoke emotional responses like reality) in the design of virtual simulations [19], factors that were not considered in this review.

Nevertheless, we identified 2 tools that had specifically used recordings of real people in the clinical situations being trained, which could have generated a greater feeling of immersion among students because of the increased physical, auditory, and visual fidelity of these tools. However, in the interactions with the simulation developed by Bearman et al [23], users had to respond from a pool of pre-established options, limiting the immersion experience because the participant was unable to develop their own communication skills in the way they would have to when facing real situations. In a tool developed by Courteille et al [28], although the user had been allowed to issue a natural language response, this had to be done in writing, which also reduced the degree of reality and spontaneity one would expect from a real conversation. Therefore, highly immersive technologies must be designed to overcome these ongoing technological challenges, such as how to integrate effective natural language processing systems and natural conversation flows into these tools [60] and how to best capture nonverbal communication [17,18]. For example, in this review, we only identified 2 applications that could detect gestures and/or emotions [29,32].

Of note, most of the tools we identified were based on autonomous learning and therefore represented promising applications with potential great benefits such as high accessibility levels, the possibility of repeating the experience multiple times, and cost reduction once running [16,17]. In this sense, technological advances that can integrate systems that provide feedback to participants—such as AI and machine learning (ML)—without the need for an instructor/teacher to mediate the learning stand out in particular [60]. For example, compared to a previous literature review [18], we found more tools in which the feedback was provided by the virtual system itself. However, as discussed, despite cataloging the existence of various patient simulation tools with interesting characteristics, we did not identify any that simultaneously integrated the use of a real person (a standardized patient) with the objective of increasing the environmental fidelity to allow the user to train through an oral conversation using natural language and using complex technology, such as AI and ML, with the ability to detect, encode, and respond to complex communication structures [60].

Finally, it is important to note that there were several limitations to this review. First, we only consulted 2 medical databases—MEDLINE/PubMed and the Web of Science. Despite being a health science–specific database and a multidisciplinary database, respectively, having replicated the search in more technological databases may have provided some additional studies for consideration. Therefore, it is possible we did not recover all the relevant records on virtual simulation tools to train communication skills in health care professionals registered in the academic literature. Second, there is still inadequate standardization in academic and scientific fields regarding the term “virtual simulation” [16,55,61]. Thus, different terms in the academic literature are all used to refer to the concept of virtual simulation including “serious games,” “virtual worlds,” “virtual patients,” and “virtual reality,” [55] which may have also caused us to miss certain relevant records.

In conclusion, this review identified and analyzed the 24 main virtual tools described in the academic literature that have been used to date to train communication skills in the context of health sciences. The high heterogeneity in terms of their characteristics means that tools based on AI and ML that contribute to training both students and practicing health professionals with as high a fidelity as possible to real life remain to be developed. Although many tools offer a degree of realism, few incorporate advanced features like AI-driven

XSL•FO
RenderX

conversational flows or nonverbal cue detection, limiting the immersive experience. This highlights a need for further development to create more effective training environments. Addressing these gaps requires future innovations that integrate natural language processing and other advanced capabilities to enhance both the realism and educational value of virtual simulations.

Acknowledgments

This study is an Erasmus project funded by the European Union (Strategic Partnerships in Higher Education [KA203], in the Call for proposals Cooperation for Innovation and exchange of good practice 2020; grant agreement 2020--1-ES01-KA203-082566).

Conflicts of Interest

None declared.

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File, 90 KB - mededu_v11i1e63082_app1.pdf]

References

1. Kaplonyi J, Bowles KA, Nestel D, et al. Understanding the impact of simulated patients on health care learners' communication skills: a systematic review. Med Educ 2017 Dec;51(12):1209-1219. [doi: 10.1111/medu.13387] [Medline: 28833360]
2. Kleinsmith A, Rivera-Gutierrez D, Finney G, Cendan J, Lok B. Understanding empathy training with virtual patients. Comput Human Behav 2015 Nov 1;52:151-158. [doi: 10.1016/j.chb.2015.05.033] [Medline: 26166942]
3. Stehr P, Reifegerste D, Rossmann C, Caspar K, Schulze A, Lindemann AK. Effective communication with caregivers to prevent unintentional injuries in children under seven years. A systematic review. Patient Educ Couns 2022 Aug;105(8):2721-2730. [doi: 10.1016/j.pec.2022.04.015] [Medline: 35537900]
4. Stacey D, Légaré F, Lewis K, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Apr 12;4(4):CD001431. [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
5. Boissy A, Windover AK, Bokar D, et al. Communication skills training for physicians improves patient satisfaction. J Gen Intern Med 2016 Jul;31(7):755-761. [doi: 10.1007/s11606-016-3597-2] [Medline: 26921153]
6. Ammentorp J, Graugaard LT, Lau ME, Andersen TP, Waidtløw K, Kofoed PE. Mandatory communication training of all employees with patient contact. Patient Educ Couns 2014 Jun;95(3):429-432. [doi: 10.1016/j.pec.2014.03.005] [Medline: 24666773]
7. Agarwal R, Sands DZ, Schneider JD. Quantifying the economic impact of communication inefficiencies in U.S. hospitals. J Healthc Manag 2010;55(4):265-281. [Medline: 20812527]
8. Synnot A, Bragge P, Lowe D, et al. Research priorities in health communication and participation: international survey of consumers and other stakeholders. BMJ Open 2018 May 8;8(5):e019481. [doi: 10.1136/bmjopen-2017-019481] [Medline: 29739780]
9. Gutiérrez-Puertas L, Márquez-Hernández VV, Gutiérrez-Puertas V, Granados-Gámez G, Aguilera-Manrique G. Educational interventions for nursing students to develop communication skills with patients: a systematic review. Int J Environ Res Public Health 2020 Mar 26;17(7):2241. [doi: 10.3390/ijerph17072241] [Medline: 32225038]
10. Lewis KL, Bohnert CA, Gammon WL, et al. The Association of Standardized Patient Educators (ASPE) Standards of Best Practice (SOBP). Adv Simul (Lond) 2017;2:10. [doi: 10.1186/s41077-017-0043-4] [Medline: 29450011]
11. Nestel D, Tabak D, Tierney T, et al. Key challenges in simulated patient programs: an international comparative case study. BMC Med Educ 2011 Sep 25;11:69. [doi: 10.1186/1472-6920-11-69] [Medline: 21943295]
12. Quail M, Brundage SB, Spitalnick J, Allen PJ, Beilby J. Student self-reported communication skills, knowledge and confidence across standardised patient, virtual and traditional clinical learning environments. BMC Med Educ 2016 Feb 27;16(72):73. [doi: 10.1186/s12909-016-0577-5] [Medline: 26919838]
13. Padilha JM, Machado PP, Ribeiro AL, Ramos JL. Clinical virtual simulation in nursing education. Clin Simul Nurs 2018 Feb;15:13-18. [doi: 10.1016/j.ecns.2017.09.005]
14. Yang H, Xiao X, Wu X, et al. Virtual standardized patients versus traditional academic training for improving clinical competence among traditional Chinese medicine students: prospective randomized controlled trial. J Med Internet Res 2023 Sep 20;25:e43763. [doi: 10.2196/43763] [Medline: 37728989]
15. Urresti-Gundlach M, Tolks D, Kiessling C, Wagner-Menghin M, Härtl A, Hege I. Do virtual patients prepare medical students for the real world? Development and application of a framework to compare a virtual patient collection with population data. BMC Med Educ 2017 Sep 22;17(1):174. [doi: 10.1186/s12909-017-1013-1] [Medline: 28938884]

XSL•FO
RenderX

16. Plotzky C, Lindwedel U, Sorber M, et al. Virtual reality simulations in nurse education: a systematic mapping review. Nurse Educ Today 2021 Jun;101:104868. [doi: 10.1016/j.nedt.2021.104868] [Medline: 33798987]
17. Battegazzorre E, Bottino A, Lamberti F. Training medical communication skills with virtual patients: literature review and directions for future research. In: Intelligent Technologies for Interactive Entertainment: Springer International Publishing; 2021:207-226. [doi: 10.1007/978-3-030-76426-5_14]
18. Lee J, Kim H, Kim KH, Jung D, Jowsey T, Webster CS. Effective virtual patient simulators for medical communication training: a systematic review. Med Educ 2020 Sep;54(9):786-795. [doi: 10.1111/medu.14152] [Medline: 32162355]
19. Peddle M, Bearman M, Nestel D. Virtual patients and nontechnical skills in undergraduate health professional education: an integrative review. Clin Simul Nurs 2016 Sep;12(9):400-410. [doi: 10.1016/j.ecns.2016.04.004]
20. Fuentes A. Reseña de sitio web: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). Declaración PRISMA 2020 [Article in Spanish]. R Est Inv Psico y Educ 2022;9(2):323-327. [doi: 10.17979/reipe.2022.9.2.9368]
21. Ali MR, Sen T, Kane B, et al. Novel computational linguistic measures, dialogue system and the development of SOPHIE: standardized online patient for healthcare interaction education. IEEE Trans Affective Comput 2020;14(1):223-235. [doi: 10.1109/TAFFC.2021.3054717]
22. Banszki F, Beilby J, Quail M, Allen P, Brundage S, Spitalnick J. A clinical educator's experience using a virtual patient to teach communication and interpersonal skills. AJET 2018;34(3):60-73. [doi: 10.14742/ajet.3296]
23. Bearman M, Cesnik B. Comparing student attitudes to different models of the same virtual patient. Stud Health Technol Inform 2001;84(Pt 2):1004-1008. [Medline: 11604882]
24. Bearman M, Cesnik B, Liddell M. Random comparison of "virtual patient" models in the context of teaching clinical communication skills. Med Educ 2001 Sep;35(9):824-832. [doi: 10.1046/j.1365-2923.2001.00999.x] [Medline: 11555219]
25. Bearman M. Is virtual the same as real? Medical students' experiences of a virtual patient. Acad Med 2003 May;78(5):538-545. [doi: 10.1097/00001888-200305000-00021] [Medline: 12742794]
26. Borja-Hart NL, Spivey CA, George CM. Use of virtual patient software to assess student confidence and ability in communication skills and virtual patient impression: a mixed-methods approach. Curr Pharm Teach Learn 2019 Jul;11(7):710-718. [doi: 10.1016/j.cptl.2019.03.009] [Medline: 31227094]
27. Chae D, Kim J, Kim K, Ryu J, Asami K, Doorenbos AZ. An immersive virtual reality simulation for cross-cultural communication skills: development and feasibility. Clin Simul Nurs 2023 Apr;77:13-22. [doi: 10.1016/j.ecns.2023.01.005]
28. Courteille O, Josephson A, Larsson LO. Interpersonal behaviors and socioemotional interaction of medical students in a virtual clinical encounter. BMC Med Educ 2014 Apr 1;14(1):64. [doi: 10.1186/1472-6920-14-64] [Medline: 24685070]
29. Deladisma AM, Johnsen K, Raij A, et al. Medical student satisfaction using a virtual patient system to learn history-taking communication skills. Stud Health Technol Inform 2008;132:101-105. [Medline: 18391266]
30. Dickerson R, Johnsen K, Raij A, et al. Virtual patients: assessment of synthesized versus recorded speech. Stud Health Technol Inform 2006;119:114-119. [Medline: 16404028]
31. Du J, Zhu X, Wang J, et al. History-taking level and its influencing factors among nursing undergraduates based on the virtual standardized patient testing results: cross sectional study. Nurse Educ Today 2022 Apr;111:105312. [doi: 10.1016/j.nedt.2022.105312] [Medline: 35287063]
32. Guetterman TC, Sakakibara R, Baireddy S, et al. Medical students' experiences and outcomes using a virtual human simulation to improve communication skills: mixed methods study. J Med Internet Res 2019 Nov 27;21(11):e15459. [doi: 10.2196/15459] [Medline: 31774400]
33. Hwang GJ, Chang CY, Ogata H. The effectiveness of the virtual patient-based social learning approach in undergraduate nursing education: a quasi-experimental study. Nurse Educ Today 2022 Jan;108:105164. [doi: 10.1016/j.nedt.2021.105164] [Medline: 34627030]
34. Jacklin S, Maskrey N, Chapman S. Improving shared decision making between patients and clinicians: design and development of a virtual patient simulation tool. JMIR Med Educ 2018 Nov 6;4(2):e10088. [doi: 10.2196/10088] [Medline: 30401667]
35. Jacklin S, Maskrey N, Chapman S. Shared decision-making with a virtual patient in medical education: mixed methods evaluation study. JMIR Med Educ 2021 Jun 10;7(2):e22745. [doi: 10.2196/22745] [Medline: 34110299]
36. Lok B. Teaching communication skills with virtual humans. IEEE Comput Graph Appl 2006;26(3):10-13. [doi: 10.1109/mcg.2006.68] [Medline: 16711211]
37. Maicher KR, Zimmerman L, Wilcox B, et al. Using virtual standardized patients to accurately assess information gathering skills in medical students. Med Teach 2019 Sep;41(9):1053-1059. [doi: 10.1080/0142159X.2019.1616683] [Medline: 31230496]
38. Mayor Silva LI, Caballero de la Calle R, Cuevas-Budhart MA, Martin Martin JO, Blanco Rodriguez JM, Gómez Del Pulgar García Madrid M. Development of communication skills through virtual reality on nursing school students: clinical trial. Comput Inform Nurs 2023 Jan 1;41(1):24-30. [doi: 10.1097/CIN.0000000000000866] [Medline: 35363632]
39. Nakagawa N, Odanaka K, Ohara H, Kisara S. Communication training for pharmacy students with standard patients using artificial intelligence. Curr Pharm Teach Learn 2022 Jul;14(7):854-862. [doi: 10.1016/j.cptl.2022.06.021] [Medline: 35914846]

XSL•FO
RenderX

40. Ochs M, Mestre D, de Montcheuil G, et al. Training doctors' social skills to break bad news: evaluation of the impact of virtual environment displays on the sense of presence. J Multimodal User Interfaces 2019 Mar;13(1):41-51. [doi: 10.1007/s12193-018-0289-8]
41. Perez A, Gaehle K, Sobczak B, Stein K. Virtual simulation as a learning tool for teaching graduate nursing students to manage difficult conversations. Clin Simul Nurs 2022 Jan;62:66-72. [doi: 10.1016/j.ecns.2021.10.003]
42. Plass AM, Covic A, Lohrberg L, Albright G, Goldman R, Von Steinbüchel N. Effectiveness of a minimal virtual motivational interviewing training for first years medical students: differentiating between pre-test and then-test. Patient Educ Couns 2022 Jun;105(6):1457-1462. [doi: 10.1016/j.pec.2021.09.020] [Medline: 34598801]
43. Real FJ, DeBlasio D, Ollberding NJ, et al. Resident perspectives on communication training that utilizes immersive virtual reality. Educ Health (Abingdon) 2017;30(3):228-231. [doi: 10.4103/efh.EfH_9_17] [Medline: 29786025]
44. Real FJ, DeBlasio D, Beck AF, et al. A virtual reality curriculum for pediatric residents decreases rates of influenza vaccine refusal. Acad Pediatr 2017;17(4):431-435. [doi: 10.1016/j.acap.2017.01.010] [Medline: 28126612]
45. Real FJ, Hood AM, Davis D, et al. An immersive virtual reality curriculum for pediatric hematology clinicians on shared decision-making for hydroxyurea in sickle cell anemia. J Pediatr Hematol Oncol 2022 Apr 1;44(3):e799-e803. [doi: 10.1097/MPH.0000000000002289] [Medline: 35319512]
46. Rouleau G, Gagnon MP, Côté J, Richard L, Chicoine G, Pelletier J. Virtual patient simulation to improve nurses' relational skills in a continuing education context: a convergent mixed methods study. BMC Nurs 2022 Jan 4;21(1):1. [doi: 10.1186/s12912-021-00740-x] [Medline: 34983509]
47. Sapkaroski D, Mundy M, Dimmock MR. Immersive virtual reality simulated learning environment versus role-play for empathic clinical communication training. J Med Radiat Sci 2022 Mar;69(1):56-65. [doi: 10.1002/jmrs.555] [Medline: 34706398]
48. Sezer B, Sezer TA. Teaching communication skills with technology: creating a virtual patient for medical students. AJET 2019;35(5):183. [doi: 10.14742/ajet.4957]
49. Şimşek Çetinkaya Ş, Gümüş Çalış G, Kıbrıs Ş, Topal M. Effectiveness of virtual patient simulation versus peer simulation in family planning training in midwifery students: a comparative educational intervention. Interactive Learning Environments 2022 Aug 28;32(3):942-951. [doi: 10.1080/10494820.2022.2105897]
50. Shorey S, Ang E, Yap J, Ng ED, Lau ST, Chui CK. A virtual counseling application using artificial intelligence for communication skills training in nursing education: development study. J Med Internet Res 2019 Oct 29;21(10):e14658. [doi: 10.2196/14658] [Medline: 31663857]
51. Shorey S, Ang E, Ng ED, Yap J, Lau LST, Chui CK. Communication skills training using virtual reality: a descriptive qualitative study. Nurse Educ Today 2020 Nov;94:104592. [doi: 10.1016/j.nedt.2020.104592] [Medline: 32942248]
52. Shorey S, Ang ENK, Ng ED, et al. Evaluation of a theory-based virtual counseling application in nursing education. Comput Inform Nurs 2023 Jun 1;41(6):385-393. [doi: 10.1097/CIN.0000000000000999] [Medline: 36728150]
53. Stevens A, Hernandez J, Johnsen K, et al. The use of virtual patients to teach medical students history taking and communication skills. Am J Surg 2006 Jun;191(6):806-811. [doi: 10.1016/j.amjsurg.2006.03.002] [Medline: 16720154]
54. Mendez KJW, Piasecki RJ, Hudson K, et al. Virtual and augmented reality: implications for the future of nursing education. Nurse Educ Today 2020 Oct;93:104531. [doi: 10.1016/j.nedt.2020.104531] [Medline: 32711132]
55. Cant R, Cooper S, Sussex R, Bogossian F. What's in a name? Clarifying the nomenclature of virtual simulation. Clin Simul Nurs 2019 Feb;27:26-30. [doi: 10.1016/j.ecns.2018.11.003]
56. Watts PI, McDermott DS, Alinier G, et al. Healthcare Simulation Standards of Best PracticeTM simulation design. Clin Simul Nurs 2021 Sep;58:14-21. [doi: 10.1016/j.ecns.2021.08.009]
57. Maicher K, Danforth D, Price A, et al. Developing a conversational virtual standardized patient to enable students to practice history-taking skills. Sim Healthcare 2017;12(2):124-131. [doi: 10.1097/SIH.0000000000000195]
58. Witmer BG, Singer MJ. Measuring presence in virtual environments: a presence questionnaire. Presence (Camb) 1998 Jun;7(3):225-240. [doi: 10.1162/105474698565686]
59. Zielke MA, Zakhidov D, Hardee G, et al. Developing virtual patients with VR/AR for a natural user interface in medical teaching. In: 2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH): IEEE; 2017. [doi: 10.1109/SeGAH.2017.7939285]
60. Stamer T, Steinhäuser J, Flägel K. Artificial intelligence supporting the training of communication skills in the education of health care professions: scoping review. J Med Internet Res 2023 Jun 19;25(8):e43311. [doi: 10.2196/43311] [Medline: 37335593]
61. Foronda CL, Fernandez-Burgos M, Nadeau C, Kelley CN, Henry MN. Virtual simulation in nursing education: a systematic review spanning 1996 to 2018. Sim Healthcare 2020;15(1):46-54. [doi: 10.1097/SIH.0000000000000411]

Abbreviations

AI: artificial intelligence
MeSH: Medical Subject Headings
ML: machine learning

XSL•FO
RenderX

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by B Lesselroth; submitted 10.06.24; peer-reviewed by A Grilo, JD Ramos Pichardo; revised version received 29.10.24; accepted 02.01.25; published 06.05.25.

Please cite as:
Fernández-Alcántara M, Escribano S, Juliá-Sanchis R, Castillo-López A, Pérez-Manzano A, Macur M, Kalender-Smajlović S, García-Sanjuán S, Cabañero-Martínez MJ
Virtual Simulation Tools for Communication Skills Training in Health Care Professionals: Literature Review
JMIR Med Educ 2025;11:e63082
URL: https://mededu.jmir.org/2025/1/e63082
doi:10.2196/63082

© Manuel Fernández-Alcántara, Silvia Escribano, Rocío Julia-Sanchis, Ana Castillo-López, Antonio Pérez-Manzano, M Macur, Sedina Kalender-Smajlovic, Sofía García-Sanjuán, Maria José Cabañero-Martínez. Originally published in JMIR Medical Education (https://mededu.jmir.org), 6.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

AI in the Health Sector: Systematic Review of Key Skills for Future Health Professionals

Javier Gazquez-Garcia[1], MSc; Carlos Luis Sánchez-Bocanegra[2], PhD; Jose Luis Sevillano[3], PhD

[1]Servicio Andaluz de Salud, Distrito Sanitario Almeria, Almeria, Spain

[2]Faculty of Health Sciences, Universidad Oberta de Catalunya (UOC), Barcelona, Spain

[3]Universidad de Sevilla, ETS Ingenieria Informatica, Avda Reina Mercedes s/n, Sevilla, Spain

Corresponding Author:
Jose Luis Sevillano, PhD
Universidad de Sevilla, ETS Ingenieria Informatica, Avda Reina Mercedes s/n, Sevilla, Spain

Abstract

Background: Technological advancements have significantly reshaped health care, introducing digital solutions that enhance diagnostics and patient care. Artificial intelligence (AI) stands out, offering unprecedented capabilities in data analysis, diagnostic support, and personalized medicine. However, effectively integrating AI into health care necessitates specialized competencies among professionals, an area still in its infancy in terms of comprehensive literature and formalized training programs.

Objective: This systematic review aims to consolidate the essential skills and knowledge health care professionals need to integrate AI into their clinical practice effectively, according to the published literature.

Methods: We conducted a systematic review, across databases PubMed, Scopus, and Web of Science, of peer-reviewed literature that directly explored the required skills for health care professionals to integrate AI into their practice, published in English or Spanish from 2018 onward. Studies that did not refer to specific skills or training in digital health were not included, discarding those that did not directly contribute to understanding the competencies necessary to integrate AI into health care practice. Bias in the examined works was evaluated following Cochrane's domain-based recommendations.

Results: The initial database search yielded a total of 2457 articles. After deleting duplicates and screening titles and abstracts, 37 articles were selected for full-text review. Out of these, only 7 met all the inclusion criteria for this systematic review. The review identified a diverse range of skills and competencies, that we categorized into 14 key areas classified based on their frequency of appearance in the selected studies, including AI fundamentals, data analytics and management, and ethical considerations.

Conclusions: Despite the broadening of search criteria to capture the evolving nature of AI in health care, the review underscores a significant gap in focused studies on the required competencies. Moreover, the review highlights the critical role of regulatory bodies such as the US Food and Drug Administration in facilitating the adoption of AI technologies by establishing trust and standardizing algorithms. Key areas were identified for developing competencies among health care professionals for the implementation of AI, including: AI fundamentals knowledge (more focused on assessing the accuracy, reliability, and validity of AI algorithms than on more technical abilities such as programming or mathematics), data analysis skills (including data acquisition, cleaning, visualization, management, and governance), and ethical and legal considerations. In an AI-enhanced health care landscape, the ability to humanize patient care through effective communication is paramount. This balance ensures that while AI streamlines tasks and potentially increases patient interaction time, health care professionals maintain a focus on compassionate care, thereby leveraging AI to enhance, rather than detract from, the patient experience.

(JMIR Med Educ 2025;11:e58161) doi:10.2196/58161

KEYWORDS

artificial intelligence; healthcare competencies; systematic review; healthcare education; AI regulation

Introduction

Technological advancements have transformed health care, improving diagnostics and patient care through solutions such as diagnostic support systems and telemedicine. These technologies reduce costs and enhance care quality, but their adoption faces challenges, particularly in terms of infrastructure, training, and education [1,2]. To address these challenges, the European Commission's DigComp framework [3] outlines key competencies needed to adapt to new technologies, with a particularly challenging issue being the integration of artificial intelligence (AI) into health care.

AI systems use complex algorithms to analyze large datasets, identify patterns, and improve decision-making [4]. AI

XSL•FO
RenderX

classification is multifaceted. The European Commission distinguishes between AI software, such as virtual assistants, and AI embedded in physical devices, such as robots [3]. Alternatively, Russell and Norvig's [5] taxonomy assesses AI based on cognitive and behavioral capabilities, differentiating systems that emulate human or rational thought and action.

Each AI approach is designed to refine its problem-solving abilities, whether by mimicking human behavior or optimizing logical decisions [4]. AI improves clinical decision-making by addressing human limitations in data processing, supporting evidence-based practices through technologies such as machine learning and deep learning [6]. Various AI applications, such as image processing and convolutional neural networks, enhance clinical outcomes [6].

While human interaction remains central to health care, AI mitigates cognitive biases and provides faster, more precise outcomes [7]. For example, AI can process millions of medical images far faster than a human radiologist, improving accuracy through continual learning [6].

The integration of AI into health care is not just a technological advancement but a profound transformation of the operational, cultural, and ethical frameworks of health care organizations. This shift requires professionals to develop specialized knowledge in AI disciplines, such as machine learning, deep learning, and natural language processing, to use these tools effectively and ethically [8].

AI's integration demands expertise and regulatory oversight to ensure ethical use, bias mitigation, and privacy protection [9]. The US Food and Drug Administration is exploring regulatory frameworks for AI-based algorithms in medicine, though a definitive process has yet to be established [10]. A recent study in *JAMA Ophthalmology* [11,12] highlighted the risks of AI misuse, including sophisticated models such as ChatGPT. Professionals are essential in identifying data falsifications that may not be evident to untrained individuals [12].

Addressing these challenges requires a concerted effort to enhance AI literacy and redesign clinical processes, fostering synergy between human judgment and AI-augmented decision-making [13,14]. This transition demands rigorous training in technical, procedural, and collaborative skills to ensure health care professionals can effectively integrate AI into practice, enabling them to adapt to the evolving technological landscape and improve clinical practice [15].

Numerous studies have explored the integration of AI in medical education, demonstrating its potential to enhance practical skills and personalize the learning experience for students. However, most of these reviews focus on the application of AI to improve medical education [16], which examines the use of various AI methods to enhance training in different medical domains.

Some studies [17] propose that training health care professionals in AI requires specialized roles, such as health information management professionals, to manage and adapt AI technologies in clinical settings. This approach ensures that AI integration occurs safely and efficiently, considering data quality, ethical, and legal aspects.

A review of how training programs for health care professionals deal with AI shows both the relatively low number of programs available and their significant limitations [18]. The authors recommend future curricula be designed with AI-related core competencies in mind.

This review aims to identify and highlight the critical competencies necessary to guide the development of educational programs designed to optimize the use of AI in clinical settings, thus addressing a growing need at the intersection of medicine and technology. By analyzing and synthesizing the existing literature on AI training for health care professionals, our goal is to provide a comprehensive framework that informs continuous education and specialized training in AI, ensuring the safe and effective implementation of these advanced technologies in daily clinical practice.

The integration of AI into health care presents several critical challenges. This study aims to consolidate and articulate the specific skills and knowledge required for health care professionals to effectively implement AI in routine clinical practice. Our research question is as follows: "In healthcare professionals, what specific skills and competencies are necessary for the effective implementation and use of AI technologies in daily clinical practice compared to their current skill set?" This question focuses on identifying and defining the critical competencies required for health care professionals to effectively use AI, providing a solid foundation for the development of educational and training programs in this emerging field.

Methods

Study Design

In conducting a systematic review between November and December 2023, we adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [19] (Multimedia Appendix 1 and Checklist 1).

Data Sources and Search Strategy

Comprehensive searches were performed across databases including PubMed, Scopus, and Web of Science, using Health Science Descriptors and Medical Subject Headings descriptors pertinent to AI in health care and associated skills. The search strategy was refined using Boolean and truncation operators. The search queries included combinations such as ("Artificial Intelligence" OR "AI") AND "Healthcare Professionals" AND ("Skills" OR "Competencies" OR "Education"), ("Artificial Intelligence" OR "AI") AND ("data analysis" OR "ethical considerations").

Inclusion and Exclusion Criteria

We established specific selection criteria, prioritizing peer-reviewed original articles, review articles, editorials, and commentaries that directly explored the required skills for health care professionals to integrate AI into their practice. Regarding the inclusion criteria, studies were considered if they: were published in English or Spanish, taking advantage of the linguistic accessibility for the research team in order to reflect the possible specific characteristics of the Spanish-speaking

XSL•FO
RenderX

community; published from 2018 onward, to ensure the relevance and timeliness of the research given the rapid advancements in AI technologies in recent years; focused on the necessary skills for the effective use of AI by health care professionals, encompassing both technical and management competencies; and included aspects related to training in digital health, highlighting the importance of specific training in the use of emerging technologies.

Conversely, studies were excluded if they did not meet these criteria: studies written in languages other than English or Spanish, as they could not be accurately analyzed by the research team; studies published before 2018, to focus on the most recent trends in AI in health care; studies not specifically related to the skills or needs of health care professionals for the use of AI tools, excluding research that did not directly address this focus; and studies that did not refer to specific skills or training in digital health, discarding those that did not directly contribute to understanding the competencies necessary to integrate AI into health care practice.

This selection methodology was designed to identify studies that provided significant evidence on the key competencies health care professionals need to develop for effectively integrating AI into their clinical practice, thus, ensuring that the systematic review focused on research offering practical and applicable insights.

Data Extraction and Study Quality Assessment

The search strategy was developed iteratively to optimize the retrieval of relevant studies. An initial search used the aforementioned databases and descriptors. Titles and abstracts underwent rigorous review for relevance by 2 independent researchers (JG-G and CLS-B), with duplicates removed and articles not meeting inclusion criteria or fitting exclusion criteria discarded. Mendeley (Elsevier Ltd) served as the reference management tool, not involved in the data extraction process.

Subsequently, full-text articles were retrieved for an in-depth content review based on the inclusion criteria specified earlier. However, 3 articles were not retrievable despite repeated efforts. Two articles were inaccessible due to subscription restrictions, and attempts to obtain them via interlibrary loans were unsuccessful. The third article had a broken link, and the authors were unresponsive after multiple contact attempts. These articles have been documented in the “reports not retrieved” section of the PRISMA flow diagram in the Results section.

To assure methodological integrity, discrepancies were resolved through consultation with a third researcher (JLS), reaching consensus on study admissibility. The screening process was manual, without the use of automation tools, and each study was carefully evaluated. Mendeley was used again for reference management, without affecting the data extraction process. The PRISMA flow diagram is provided in the Results section.

Data were extracted and synthesized from eligible studies, encompassing details such as authors, publication year, study design, identified skills, and methodological quality assessment. Tables served as the primary method for tabulating and visually presenting the results and their synthesis.

The data search spanned all pertinent dimensions of these outcomes, including variables relevant to the analyzed studies. These variables covered aspects such as humanization, social skills, and participants’ AI usage experience, offering a comprehensive perspective on the competencies necessary for integrating AI into health care practices.

The GRADE (Grading of Recommendations Assessment, Development, and Evaluation) framework [20] was applied to assess the evidence quality of the included works, considering the quality of studies, result consistency, imprecision, potential bias, and other pertinent factors. Mixed methods or qualitative studies were appraised using the mixed methods appraisal tool [21].

Bias in the examined works was evaluated following Cochrane’s domain-based recommendations [22], considering five types of bias, each with its domains. Bias risk was independently assessed by at least 2 reviewers, with discrepancies resolved via discussion or consultation with a third reviewer when necessary.

Synthesis Method

The research question guided the synthesis groupings, focusing the analysis on the skills and competencies necessary for the effective implementation of AI technologies by health care professionals. No standardization metric was applied. The synthesis method involved extracting relevant sections of the studies related to the identified skills and competencies. Overall, the risk of bias assessment of the studies showed no critical results. Consequently, the included studies were synthesized with equal weight.

The categorization of skills and competencies was based on a qualitative synthesis approach, where 2 independent reviewers extracted and coded data from each study. The categories were developed iteratively by grouping competencies that were conceptually aligned. The final domains were determined based on the frequency with which competencies appeared across the studies, with higher-frequency competencies categorized as key domains. For example, “AI fundamentals” and “ethical and legal considerations” were identified as critical due to their frequent mention in the selected studies, whereas others such as “data governance” and “programming” appeared less often but were still considered relevant.

Study design was not restricted, so a meta-analysis was not performed due to the heterogeneity of the studies and the differing amounts and qualities of information reported on skills and competencies. The reported effectiveness of the competencies was synthesized based on each study’s findings. The identified competencies are presented in the results section. Tables and figures aggregate information about this study’s characteristics and focal areas of this review (skills and competencies).

This review aimed to identify and delineate the skills and competencies essential for health care professionals to use AI in their clinical practices effectively. The objective encompassed various outcome domains, not limited to interpreting results from machine learning models, managing AI biases, ethical

XSL•FO
RenderX

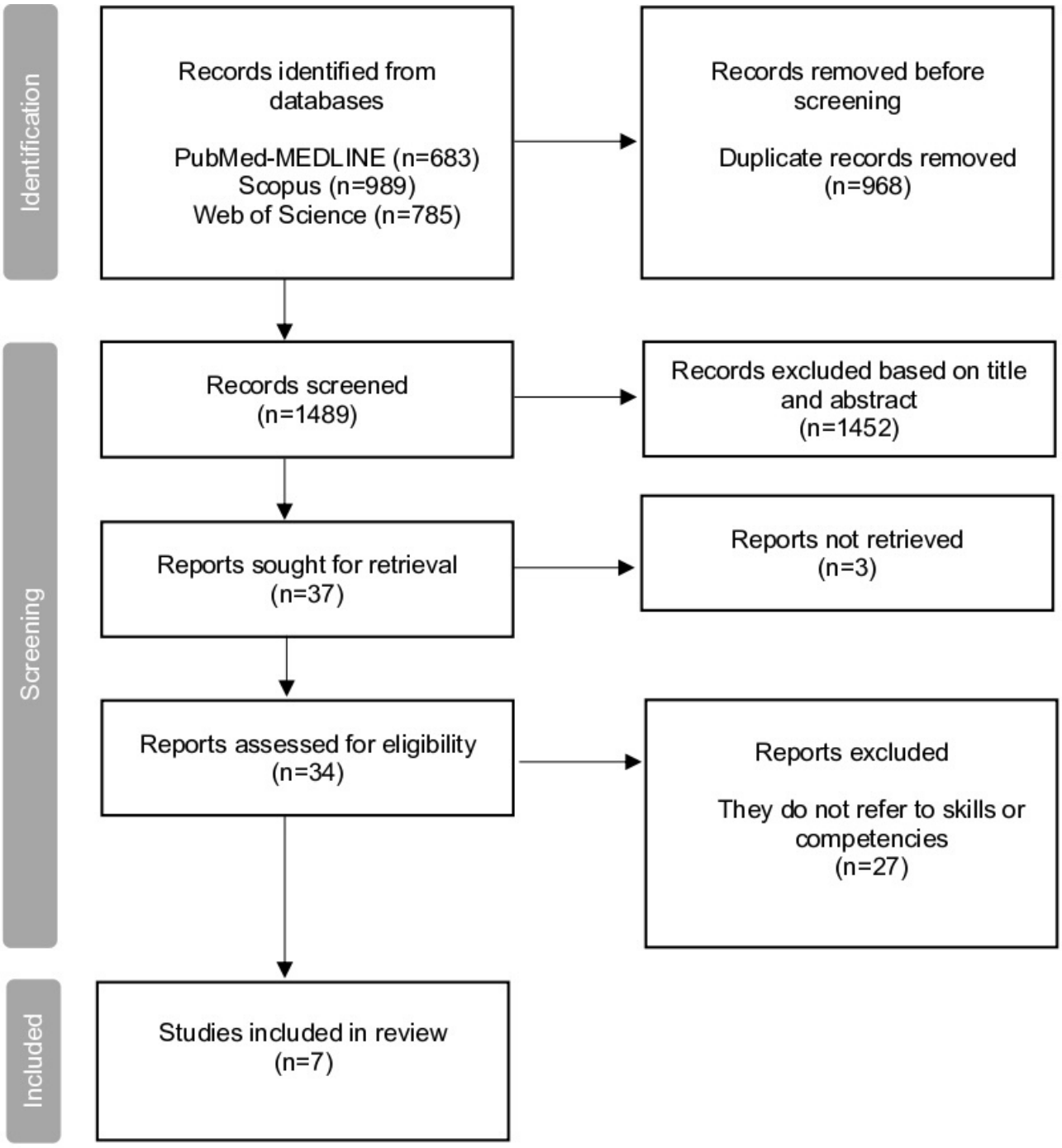
considerations in AI-assisted decision-making, and the technical skills required for effective AI tool use.

Results

Overview

The initial database searches yielded a total of 2457 articles (Figure 1). After deleting duplicates and screening titles and abstracts, 37 articles were selected for full-text review. Out of these, only 7 met all the inclusion criteria for this systematic review [23-29]. Each selected article specifically concentrated on the competencies essential for the integration of AI into routine clinical practice. A substantial number of the excluded articles (n=28) made reference to, but did not directly engage with, the competencies in question. Table S1 in Multimedia Appendix 1 delineates the characteristics of these studies and the principal competencies identified therein.

Figure 1. PRISMA flow diagram. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



The review identified a diverse range of skills and competencies necessary for the effective implementation of AI in health care. These competencies were categorized into several key areas based on their frequency of appearance in the selected studies. The identified skills, definitions, and the number of studies that reported each are summarized in Table S2 in Multimedia Appendix 1.

XSL•FO
RenderX

Identified Skills

The analysis of the selected studies highlights several key competencies deemed essential for the effective integration of AI into routine clinical practice.

AI fundamentals were identified as an essential competency by 86% of the studies [23-27,29]. This indicates a strong consensus on the importance of foundational AI knowledge for health care professionals. The studies emphasize that a solid understanding of AI principles is crucial for effectively implementing and using AI technologies in clinical settings. This foundational knowledge forms the bedrock upon which more advanced competencies are built, ensuring that professionals can confidently engage with AI tools and integrate them into their practice.

Ethical and legal considerations were emphasized by 71% of the studies [23-26,29]. The studies underscored the importance of having a solid knowledge base in various aspects of ethics, including patient privacy, data security, biases in algorithms, and transparency and explainability.

Data analysis and management skills were highlighted by 43% of the studies [24,26,28]. These studies emphasize several secondary skills crucial for effective data handling. For instance, McCoy et al [26] stress the importance of data acquisition, cleaning, and visualization as foundational steps in preparing data for AI applications. Singh et al [24] underline the need for robust data management practices to ensure the integrity and reliability of AI outputs. Wiljer and Hakim [28] highlight the necessity of developing capabilities in data governance, which includes the secure storage and regulatory compliance of health care data.

Communication and teamwork were identified as important competencies by 43% of the studies [23,25,27]. These studies underscore the critical need for effective communication skills to convey complex AI-related information to both colleagues and patients. Çalışkan et al [23] emphasize the role of interdisciplinary teamwork in the development and implementation of AI applications, highlighting the necessity for seamless collaboration between health care professionals and AI experts. Liaw et al [25] point out the importance of clear and empathetic communication with patients regarding the use of AI in their care, ensuring transparency and maintaining trust. Sujan et al [27] stress the need for health care teams to work cohesively to monitor and supervise AI systems, ensuring their safe and ethical use.

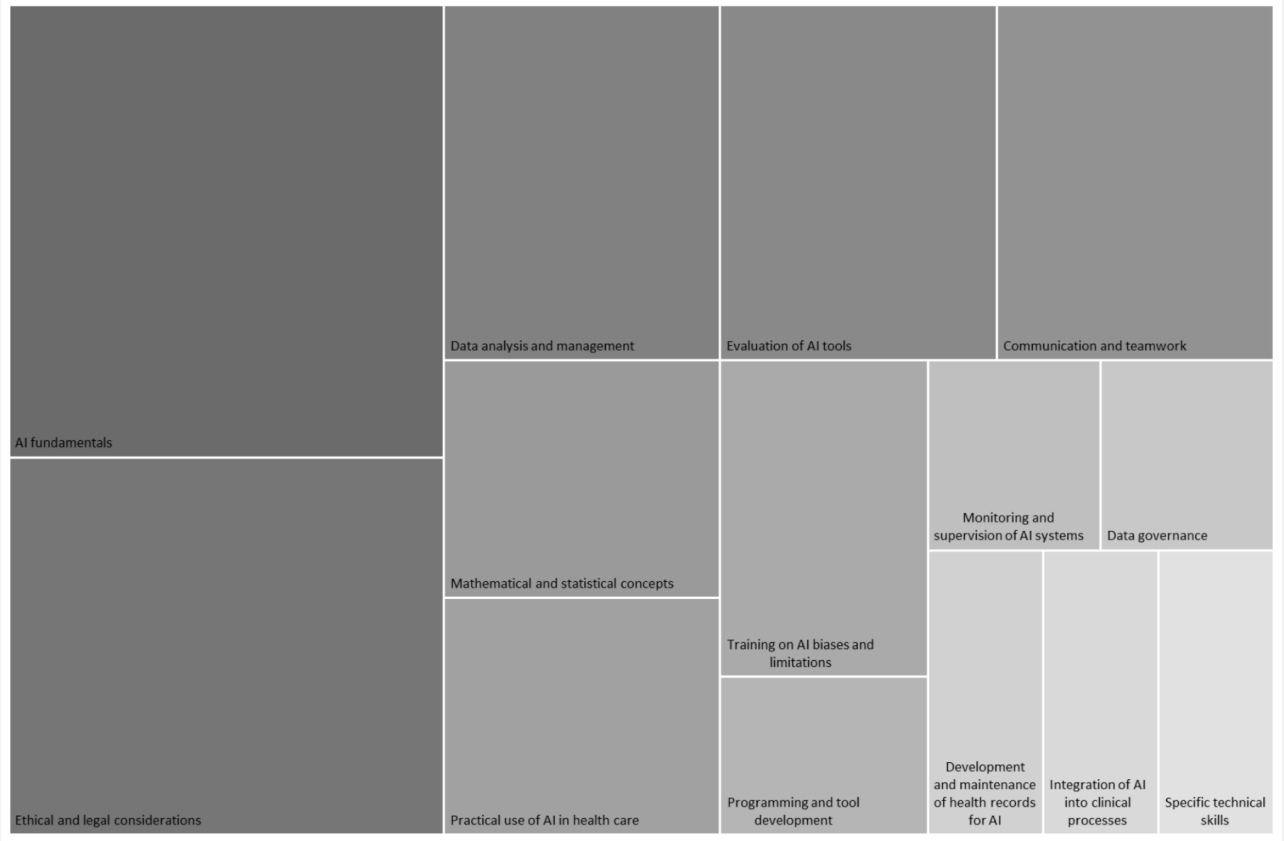
Evaluation of AI tools was mentioned by 43% of the studies [25,26,29]. These studies highlight several essential secondary skills necessary for the rigorous assessment of AI technologies. Liaw et al [25] emphasize the importance of understanding evidence-based evaluation methods to critically assess the performance and utility of AI tools in clinical practice. McCoy et al [26] discuss the need for health care professionals to be proficient in performing critical evaluations, including assessing the accuracy, reliability, and validity of AI algorithms. Sapci and Sapci [29] stress the significance of ongoing evaluation and monitoring of AI tools to ensure they consistently meet clinical standards and enhance patient outcomes.

Some investigations highlight more technical abilities, such as programming [26], or a deeper mathematical acumen [24,28]. These skills, while crucial, were not as frequently cited as the previously mentioned competencies. Conversely, a capability deemed highly significant across the literature is the aptitude for evaluating AI tools to ascertain their quality and rationalize their application, in addition to scrutinizing potential biases and limitations [24,25].

Figure 2 presents a mosaic plot displaying the distribution of identified competencies based on the percentage of studies that reported each skill. The size of each tile corresponds to the proportion of studies mentioning that competency, providing a visual representation of the relative importance of each skill across the reviewed literature.

XSL•FO
RenderX

Figure 2. Distribution of identified competencies. AI: artificial intelligence.



Discussion

General Aspects

This review underscores the critical need for health care professionals to acquire competencies related to the effective use of AI in clinical practice. Five key areas of competence were identified as essential for AI integration in clinical practice: AI fundamentals, ethical and legal considerations, data analysis and management, communication and teamwork, and the evaluation of AI tools. These competencies are essential to ensure the safe and reliable integration of AI technologies into routine health care workflows. Furthermore, ethical and legal considerations, such as data security and transparency, are crucial for building trust in AI-driven decisions (Table S2 in Multimedia Appendix 1).

In line with our findings, recent studies have emphasized the growing need to define and standardize AI competencies within clinical settings. A 2022 exploratory review [30] highlighted this gap, calling for more structured educational programs to support the integration of AI into health care practice. Additionally, a survey among nursing staff [31] indicated that most professionals had acquired their AI skills independently, further reinforcing the need for formalized training curricula.

Skills

AI Fundamentals

The consensus across the literature highlights the essential nature of AI fundamentals [23-27,29]. This foundational knowledge is crucial for health care professionals to effectively implement and use AI technologies, providing the bedrock upon which more advanced competencies are built. AI fundamentals encompass a basic understanding of machine learning, neural networks, and deep learning. While health care professionals are not expected to have deep expertise in these technical areas, a solid grasp of how AI models function—how they are trained, how they process data, and how they support clinical decision-making—is necessary. This understanding allows professionals to effectively integrate AI into their workflows and ensure the appropriate use of AI-driven tools in patient care.

A key competency is the ability to critically assess AI algorithms, identify biases, and interpret the results. Studies, such as those by Faes et al [32], provide guidelines to help clinicians evaluate AI outputs without needing advanced computational skills, allowing them to focus on patient care. For example, clinicians should recognize when AI models may be limited by biased data or poor generalizability. Similar to using ultrasound technology—where clinicians do not fully understand the physics but can accurately interpret images—the same principle applies to AI. Health care professionals must develop the skills to interpret AI-generated information and effectively communicate it to patients, fostering trust and transparency in AI-driven care [33].

Ethical and Legal Considerations

Ethical and legal considerations were emphasized in 71% of the reviewed studies [23-26,29], highlighting the importance of health care professionals understanding and adhering to the ethical principles and legal frameworks governing the use of AI in clinical practice. These considerations include ensuring

XSL•FO
RenderX

patient privacy, data security, and compliance with relevant regulations, which are critical for maintaining the trust and safety of patients.

Stöger et al [34] emphasized the crucial role of managing high-quality data and the risks associated with its mismanagement, underscoring both the ethical and legal implications for health care professionals using AI systems. In this context, health care providers must be familiar with the laws and regulations governing AI to ensure that the technology is implemented in compliance with legal and ethical standards.

In the European Union, the recent approval of the AI Act introduces comprehensive regulations [35] aimed at safeguarding fundamental rights and democracy by controlling AI systems based on their potential risks and impact. The AI Act establishes stringent requirements for high-risk AI systems, including mandatory impact assessments and prohibitions on certain AI practices that threaten fundamental rights. For health care professionals, this underscores the need for a thorough understanding of these legal frameworks to ensure that AI is used responsibly and safely in clinical practice.

Moreover, the opacity of machine learning algorithms, often referred to as “black boxes” due to their complex and abstract problem-solving methods, represents a significant challenge in ensuring accountability and transparency in AI-driven health care solutions [36]. Although efforts to make these algorithms more interpretable are still developing [37], health care professionals must rely on regulated AI tools that meet established standards for safety and efficacy. Regulatory bodies such as the US Food and Drug Administration have already approved specific AI tools for medical applications [38], providing a level of assurance that these technologies meet rigorous safety and legal standards. This regulatory oversight is critical for fostering trust in AI among clinicians and patients, while also mitigating liability concerns for health care providers.

Data Analysis and Management

Data analysis and management were identified as key competencies in 43% of the reviewed studies [24,26,28], emphasizing the critical role these skills play in the effective implementation of AI in health care. This domain encompasses the ability to handle large volumes of health care data, including data acquisition, cleaning, visualization, and governance. Health care professionals need to be proficient in organizing and managing these datasets to ensure the integrity and reliability of AI outputs.

A core competency in this area includes the preparation of clean and structured datasets for AI models, as highlighted by McCoy et al [26], who stress the importance of data preparation steps such as cleaning and visualization to optimize AI performance. Singh et al [24] underline the need for robust data management practices to guarantee that the data used by AI systems is accurate, reliable, and compliant with health care regulations.

Wiljer and Hakim [28] further highlight the significance of data governance, which includes not only managing data but also ensuring its secure storage and adherence to regulatory requirements. Proficient data handling is crucial, as health care professionals are often the primary data generators. Their ability to manage and interpret large datasets ensures that AI methodologies are applied effectively in clinical practice, supporting both the accuracy of AI predictions and the quality of patient care [39].

Evaluate AI Tools

The ability to evaluate AI tools emphasises the need for health care professionals to critically assess the performance, accuracy, and reliability of AI technologies in clinical practice. Effective evaluation is closely tied to a foundational understanding of AI principles, which allows professionals to determine whether AI tools are suitable for their specific clinical settings.

A key competency in this domain involves assessing AI tools with the same rigor applied to new drugs, diagnostic tests, or treatment protocols. AI-based tools must undergo extensive testing for accuracy, generalizability, efficacy, and fairness to ensure they meet clinical standards [25]. This evaluation process is crucial for ensuring that AI technologies deliver reliable outcomes and can be integrated safely into health care workflows.

Additionally, health care professionals must develop skills for the ongoing evaluation and monitoring of AI tools to detect any performance shifts or biases that may arise over time [40]. Given the continuous learning nature of AI systems, regular reassessment is necessary to maintain their reliability across diverse patient populations and clinical scenarios.

Communication and Teamwork

Communication and teamwork were identified as key competencies in 43% of the reviewed studies [23,25,27], focusing on the dual importance of effectively conveying AI-related insights to patients and facilitating collaboration among health care professionals. Health care providers must clearly explain AI-generated results, addressing patient concerns with empathy, ensuring transparency, and maintaining trust in the use of AI technologies in clinical care.

Equally important is the interdisciplinary collaboration required for integrating AI into clinical workflows. Health care professionals need to work closely with AI specialists, data scientists, and other colleagues to ensure the safe and effective use of AI tools. Studies such as those by Çalışkan et al [23] stress the necessity of seamless teamwork between health care providers and AI experts, while Liaw et al [25] underscore the need for clear and compassionate communication with patients, ensuring they understand how AI is applied in their care and the potential implications.

This competency involves not only understanding AI outputs but also contextualizing them for patients in a way that fosters trust and humanizes AI-assisted care. At the same time, health care teams must collaborate cohesively to monitor and manage AI systems. As AI streamlines routine tasks, health care providers will have more time for patient interaction, making effective and empathetic communication even more critical [25]. Ensuring that both patients and professional teams feel supported and understood is essential for the ethical and successful integration of AI in health care practice.

XSL•FO
RenderX

Competencies Based on Identified Skills

Based on the previously identified skills, and following an approach similar to that outlined by the Association of American Medical Colleges in 2021 [41], we propose a set of competencies that health care professionals should acquire to effectively integrate AI into clinical practice. These competencies, derived from the literature reviewed, encompass the key domains and provide a framework to guide educational programs and ongoing training in AI-driven tools in health care settings.

Table S3 in Multimedia Appendix 1 outlines the competencies for each skill, offering a structured approach to developing AI proficiency. This framework ensures that professionals not only grasp AI fundamentals but also apply these technologies ethically, manage health care data effectively, rigorously evaluate AI tools, and communicate insights to both patients and interdisciplinary teams. However, these competencies should be considered only a first proposal; this list may be modified depending on the specific profile of the training program, on foreseeable technological developments, as well as on the evaluation of the academic results of the new curricula.

AI in Health Care Training

Recent data from Rock Health, a venture capital firm specializing in digital health, have underscored an exponential increase in investments directed toward digital health enterprises and related technologies [42]. This trend distinctly signals an evolving health care landscape, increasingly reliant on novel technologies, thereby accentuating the imperative for health care professionals to proficiently integrate such advancements into their clinical practices [43].

Despite widespread agreement on the necessity for comprehensive AI training from the outset of medical education, there is a lack of consensus on the specific content and approach of such training [15]. The discussion around this issue is abundant, yet concrete resolutions are rare. The findings of this review aim to provide an approach to the possible competencies required, offering a structured framework for developing comprehensive AI training programs for health care professionals.

Integrating AI education into health care curricula presents several challenges due to the variability and lack of standardization of required competencies. To address this need, we propose general guidelines for the development of essential AI competencies for health care professionals. These guidelines provide a reference framework that educators and course coordinators can use to design training programs, assess student progress, and establish performance criteria. For example, health care professionals should acquire the ability to assess the quality of algorithms and their interpretations, as well as identify and mitigate potential biases. This approach facilitates the integration of AI into clinical practice and provides a clear guide for curriculum development and performance evaluation.

Augmenting the education and training of health care professionals is posited to elevate their confidence in using these tools. Although concerns persist regarding AI's potential to supplant human roles, a more discerning view proposes that AI will primarily alleviate the burden of mundane tasks. This reallocation of time and resources is anticipated to enhance patient interactions and elevate the quality of health care services provided [39].

Several authors argue that the existing academic infrastructure is ill-equipped to incorporate AI education, citing time constraints and a lack of teaching expertise as significant obstacles. An alternative proposed involves the use of specific AI tools not only for clinical applications but also to elucidate the underlying algorithms, focusing on their practical use and ethical implications [44].

A solution for integrating AI education into health care curricula, addressing the shortage of instructors with expertise in clinical AI applications, involves leveraging established training programs from other institutions [45]. Programs by Stanford University [46] and Harvard University [47] serve as examples, providing access to high-quality educational content. These programs offer a comprehensive curriculum that covers essential AI concepts, practical applications, and ethical considerations, enabling health care professionals to gain a deep understanding of AI technologies.

The integration of AI into clinical practice is expected to augment, not replace, the roles of health care professionals. It calls for a workforce proficient in digital health and communication, capable of leveraging AI's benefits while recognizing its limitations and ethical considerations [48]. This paradigm shift offers an opportunity to enhance patient care, delegating computational tasks to AI and focusing on the human aspects of health care delivery [37].

Limitations

This systematic review encounters several challenges, primarily due to the limited availability of literature specifically addressing the competencies required for integrating AI into health care practice. The scarcity of targeted studies can be attributed to the nascent and rapidly evolving nature of AI applications in health care.

To mitigate this issue, the search criteria were broadened to include general terms related to AI in health care and competencies required by health care professionals. While necessary, this broadening may have introduced studies that do not exclusively focus on AI competencies, potentially affecting the homogeneity of the findings. The review also faced potential language bias, as it primarily focused on literature in English and Spanish. Pertinent studies in other languages might have been excluded. Despite rigorous and independent review processes, the selection could still be influenced by subjective interpretation of the inclusion and exclusion criteria.

Limiting the review to studies published after 2018 aimed to capture the most recent advancements, but this restriction might omit emerging research. Despite these efforts, there remains a clear knowledge gap in the specific skills required for effective AI use in health care. The literature predominantly addresses AI applications and benefits, but lacks detailed research on the precise skills health care professionals need. Current research is often fragmented and varies significantly in scope and depth. Few studies offer comprehensive models or curricula for AI competency training in health care. This inconsistency highlights

XSL•FO
RenderX

the need for more robust research to identify essential AI skills and explore effective methods for integrating these skills into health care education and practice.

Conclusions and Future Works

This systematic review has identified essential competencies for health care professionals to effectively integrate AI into clinical practice. The analysis reveals a consensus on the importance of five key areas: AI fundamentals, ethical and legal considerations, data analysis and management, communication and teamwork, and evaluation of AI tools. These competencies are crucial for leveraging AI technologies to enhance patient care and health care delivery. A consensus within the scholarly discourse suggests the necessity for health care professionals to attain proficiency in these domains to ensure the judicious application of AI tools, thereby accruing benefits for both patients and the health care ecosystem.

AI fundamentals form the backbone of necessary knowledge, enabling health care professionals to understand and use AI technologies effectively. Ethical and legal considerations ensure that AI applications adhere to patient privacy, data security, and transparency standards, maintaining trust and compliance within health care settings. Data analysis and management skills are vital for handling large datasets, ensuring accurate AI outputs, and supporting informed clinical decisions.

Communication and teamwork are also critical, facilitating the clear conveyance of AI-related information among health care professionals and to patients, thereby promoting transparency and trust. The ability to evaluate AI tools is essential for assessing the performance and reliability of AI technologies, ensuring they meet clinical standards and deliver safe, effective patient care.

Augmenting the education and training of health care professionals is posited to elevate their confidence in using these tools. Although concerns persist regarding AI's potential to supplant human roles, a more discerning view proposes that AI will primarily alleviate the burden of mundane tasks. This reallocation of time and resources is anticipated to enhance patient interactions and elevate the quality of health care services provided.

In this context, the importance of communication skills becomes increasingly paramount. The introduction of AI tools is expected to afford health care professionals additional time per patient encounter, potentially heightening patient satisfaction and care quality.

The ambition extends beyond merely acquiring proficiency in disciplines ancillary to traditional health care paradigms. Considering the already intricate and comprehensive nature of health care education, particularly in medicine, the emphasis is placed on fostering an in-depth comprehension of AI's functionalities, inherent biases, pragmatic utility, and cost-effectiveness compared to abstaining from AI applications.

The integration of AI into health care is indispensable for advancing patient care but requires a concerted effort to develop and standardize competencies among health care professionals. Regulatory oversight and enhanced educational frameworks are essential for overcoming existing barriers and leveraging AI's full potential in clinical settings.

Despite the progress highlighted in this review, significant gaps remain in the literature, particularly concerning the specific educational frameworks and training programs needed to develop these competencies. Most existing research focuses on the potential applications and benefits of AI, with less emphasis on the precise skills required for effective implementation.

Future research should prioritize the development and validation of standardized AI competency frameworks tailored for health care professionals. These frameworks should cover technical skills, ethical and legal aspects, and data management practices. Collaborative efforts between academic institutions, health care organizations, and AI experts can create comprehensive training programs to address these competencies.

Additionally, longitudinal studies are necessary to evaluate the long-term effectiveness of AI training programs. Research should explore how health care professionals apply their AI training in clinical settings, assessing the impact on patient outcomes, clinical decision-making, and health care efficiency.

Acknowledgments

This work has been partially supported by the Telefonica Chair on "Intelligence in Networks" of the Universidad de Sevilla.

Data Availability

All data generated or analyzed during this study are included in this published article. This encompasses detailed descriptions of the databases consulted, the search criteria used, the selection process for the included studies, and the analytical methods applied. Specifically, this paper delineates the comprehensive search strategy, including the exact search terms, the databases accessed (PubMed, Scopus, and Web of Science), and the filters used (such as publication date ranges and language restrictions). Additionally, the criteria for study selection—both inclusion and exclusion criteria—are explicitly outlined to ensure reproducibility and transparency of the review process. The methodologies used for data extraction and analysis are also described, providing insight into how the findings were synthesized and interpreted. Through this approach, we aim to ensure that our systematic review process is fully transparent, enabling other researchers to replicate this study or to conduct further analysis based on the procedures and datasets detailed within this article.

XSL•FO
RenderX

Authors' Contributions

All authors contributed significantly to this literature review. Specifically, each author participated in one or more of the following: conceptualization and design of the review, data acquisition, analysis, interpretation, or critically revising this work. All authors approved the final paper for publication and agree to be accountable for all aspects of this work, ensuring the integrity and accuracy of their contribution.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional tables.
[DOCX File, 109 KB - mededu_v11i1e58161_app1.docx]

Checklist 1
PRISMA checklist.
[DOCX File, 31 KB - mededu_v11i1e58161_app2.docx]

References

1. do Nascimento IJB, Abdulazeem HM, Vasanthan LT, et al. The global effect of digital health technologies on health workers' competencies and health workplace: an umbrella review of systematic reviews and lexical-based and sentence-based meta-analysis. Lancet Digit Health 2023 Aug;5(8):e534-e544. [doi: 10.1016/S2589-7500(23)00092-4] [Medline: 37507197]
2. do Nascimento IJB, Abdulazeem H, Vasanthan LT, et al. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. NPJ Digit Med 2023 Sep 18;6(1):161. [doi: 10.1038/s41746-023-00899-4] [Medline: 37723240]
3. Vuorikari R, Kluzer S, Punie Y. DigComp 22: The Digital Competence Framework for Citizens - With New Examples of Knowledge, Skills and Attitudes: Publications Office of the European Union; 2022. [doi: 10.2760/490274]
4. Qué es la inteligencia artificial. Gobierno de España, Plan de Recuperación Transformación y Resiliencia. URL: https://planderecuperacion.gob.es/noticias/que-es-inteligencia-artificial-ia-prtr [accessed 2025-01-14]
5. Russell S, Norvig P. Artificial Intelligence: A Modern Approach, 4th edition: Pearson – prentice hall.
6. Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. Minim Invasive Ther Allied Technol 2019 Apr;28(2):73-81. [doi: 10.1080/13645706.2019.1575882] [Medline: 30810430]
7. Liu X, Zhang W, Zhang Q, et al. Development and validation of a machine learning-augmented algorithm for diabetes screening in community and primary care settings: a population-based study. Front Endocrinol (Lausanne) 2022;13:1043919. [doi: 10.3389/fendo.2022.1043919] [Medline: 36518245]
8. Kim HS, Kim DJ, Yoon KH. Medical big data is not yet available: why we need realism rather than exaggeration. Endocrinol Metab (Seoul) 2019 Dec;34(4):349-354. [doi: 10.3803/EnM.2019.34.4.349] [Medline: 31884734]
9. Arencibia MG, Cardero DM. Dilemas éticos en el escenario de la inteligencia artificial. Econom y Socied 2020;25(57):1-18. [doi: 10.15359/eys.25-57.5]
10. Harvey HB, Gowda V. How the FDA regulates AI. Acad Radiol 2020 Jan;27(1):58-61. [doi: 10.1016/j.acra.2019.09.017] [Medline: 31818387]
11. Taloni A, Scorcia V, Giannaccare G. Large language model advanced data analysis abuse to create a fake data set in medical research. JAMA Ophthalmol 2023 Dec 1;141(12):1174-1175. [doi: 10.1001/jamaophthalmol.2023.5162] [Medline: 37943569]
12. Naddaf M. ChatGPT generates fake data set to support scientific hypothesis. Nature New Biol 2023 Nov 30;623(7989):895-896. [doi: 10.1038/d41586-023-03635-w]
13. Lauricella LL, Pêgo-Fernandes PM. Databases, big data and artificial intelligence: what healthcare professionals need to know about them. Sao Paulo Med J 2022;140(6):737-738. [doi: 10.1590/1516-3180.2022.140611082022]
14. Kamble SS, Gunasekaran A, Goswami M, Manda J. A systematic perspective on the applications of big data analytics in healthcare management. Int J Healthc Manag 2019 Jul 3;12(3):226-240. [doi: 10.1080/20479700.2018.1531606]
15. Grunhut J, Wyatt AT, Marques O. Educating future physicians in artificial intelligence (AI): an integrative review and proposed changes. J Med Educ Curric Dev 2021;8:23821205211036836. [doi: 10.1177/23821205211036836] [Medline: 34778562]
16. Nagi F, Salih R, Alzubaidi M, et al. Applications of artificial intelligence (AI) in medical education: a scoping review. Stud Health Technol Inform 2023 Jun 29;305:648-651. [doi: 10.3233/SHTI230581] [Medline: 37387115]
17. Stanfill MH, Marc DT. Health information management: implications of artificial intelligence on healthcare data and information management. Yearb Med Inform 2019 Aug;28(1):56-64. [doi: 10.1055/s-0039-1677913] [Medline: 31419816]
18. Charow R, Jeyakumar T, Younus S, et al. Artificial intelligence education programs for health care professionals: scoping review. JMIR Med Educ 2021 Dec 13;7(4):e31043. [doi: 10.2196/31043] [Medline: 34898458]

XSL•FO
RenderX

19. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097. [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
20. Aguayo-Albasini JL, Flores-Pastor B, Soria-Aledo V. Sistema GRADE: clasificación de la calidad de la evidencia y graduación de la fuerza de la recomendación. Cir Esp 2014 Feb;92(2):82-88. [doi: 10.1016/j.ciresp.2013.08.002]
21. Hong QN, Fàbregues S, Bartlett G, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Educ Inf 2018 Dec 18;34(4):285-291. [doi: 10.3233/EFI-180221]
22. Higgins JPT, Thomas J, Chandler J, et al. Cochrane Handbook for Systematic Reviews of Interventions Version 64: Cochrane; 2023. URL: https://training.cochrane.org/handbook [accessed 2025-01-14]
23. Çalışkan SA, Demir K, Karaca O. Artificial intelligence in medical education curriculum: an e-Delphi study for competencies. PLoS ONE 2022;17(7):e0271872. [doi: 10.1371/journal.pone.0271872] [Medline: 35862401]
24. Singh RP, Hom GL, Abramoff MD, Campbell JP, Chiang MF, AAO Task Force on Artificial Intelligence. Current challenges and barriers to real-world artificial intelligence adoption for the healthcare system, provider, and the patient. Transl Vis Sci Technol 2020 Aug;9(2):45. [doi: 10.1167/tvst.9.2.45] [Medline: 32879755]
25. Liaw W, Kueper JK, Lin S, Bazemore A, Kakadiaris I. Competencies for the use of artificial intelligence in primary care. Ann Fam Med 2022;20(6):559-563. [doi: 10.1370/afm.2887] [Medline: 36443071]
26. McCoy LG, Nagaraj S, Morgado F, Harish V, Das S, Celi LA. What do medical students actually need to know about artificial intelligence? NPJ Digit Med 2020;3(1):86. [doi: 10.1038/s41746-020-0294-7] [Medline: 32577533]
27. Sujan MA, White S, Habli I, Reynolds N. Stakeholder perceptions of the safety and assurance of artificial intelligence in healthcare. Saf Sci 2022 Nov;155(8):105870. [doi: 10.1016/j.ssci.2022.105870]
28. Wiljer D, Hakim Z. Developing an artificial intelligence-enabled health care practice: rewiring health care professions for better care. J Med Imaging Radiat Sci 2019 Dec;50(4 Suppl 2):S8-S14. [doi: 10.1016/j.jmir.2019.09.010] [Medline: 31791914]
29. Sapci AH, Sapci HA. Artificial intelligence education and tools for medical and health informatics students: systematic review. JMIR Med Educ 2020 Jun 30;6(1):e19285. [doi: 10.2196/19285] [Medline: 32602844]
30. Garvey KV, Craig KJT, Russell R, Novak LL, Moore D, Miller BM. Considering clinician competencies for the implementation of artificial intelligence-based tools in health care: findings from a scoping review. JMIR Med Inform 2022 Nov 16;10(11):e37478. [doi: 10.2196/37478] [Medline: 36318697]
31. Abuzaid MM, Elshami W, Fadden SM. Integration of artificial intelligence into nursing practice. Health Technol (Berl) 2022;12(6):1109-1115. [doi: 10.1007/s12553-022-00697-0] [Medline: 36117522]
32. Faes L, Liu X, Wagner SK, et al. A clinician's guide to artificial intelligence: how to critically appraise machine learning studies. Transl Vis Sci Technol 2020 Feb 12;9(2):7. [doi: 10.1167/tvst.9.2.7] [Medline: 32704413]
33. Nagy M, Radakovich N, Nazha A. Why machine learning should be taught in medical schools. Med Sci Educ 2022 Apr;32(2):529-532. [doi: 10.1007/s40670-022-01502-3] [Medline: 35528308]
34. Stöger K, Schneeberger D, Kieseberg P, Holzinger A. Legal aspects of data cleansing in medical AI. Comput Law Secur Rev 2021 Sep;42:105587. [doi: 10.1016/j.clsr.2021.105587]
35. European Commission. Official Journal of the European Union; Regulation (EU) 2023/206 on artificial intelligence. EU Artificial Intelligence Act. 2024. URL: https://artificialintelligenceact.eu/the-act/ [accessed 2025-01-14]
36. Lysaght T, Lim HY, Xafis V, Ngiam KY. AI-assisted decision-making in healthcare. ABR 2019 Sep;11(3):299-314. [doi: 10.1007/s41649-019-00096-0]
37. Savage N. Breaking into the black box of artificial intelligence. Nature New Biol 2022 Mar 29. [doi: 10.1038/d41586-022-00858-1] [Medline: 35352042]
38. Milam ME, Koo CW. The current status and future of FDA-approved artificial intelligence tools in chest radiology in the United States. Clin Radiol 2023 Feb;78(2):115-122. [doi: 10.1016/j.crad.2022.08.135] [Medline: 36180271]
39. Khanra S, Dhir A, Islam A, Mäntymäki M. Big data analytics in healthcare: a systematic literature review. Ent Inf Syst 2020 Aug 8;14(7):878-912. [doi: 10.1080/17517575.2020.1812005]
40. James CA, Wheelock KM, Woolliscroft JO. Machine learning: the next paradigm shift in medical education. Acad Med 2021 Jul 1;96(7):954-957. [doi: 10.1097/ACM.0000000000003943] [Medline: 33496428]
41. AAMC. AAMC new and emerging areas in medicine series; telehealth competencies across the learning continuum. Arizona Telemedine Program.: AAMC; 2021. URL: https://telemedicine.arizona.edu/sites/default/files/training/2022/Feb/5%20-%20AAMC-2021-telehealth-competencies%20across%20the%20learning%20continuum.pdf [accessed 2025-01-14]
42. Zweig M, Desiliva J. Q1 2021 funding report: digital health is all grown up. Rock Health. 2021. URL: https://rockhealth.com/insights/q1-2021-funding-report-digital-health-is-all-grown-up/ [accessed 2025-01-14]
43. Marwaha JS, Landman AB, Brat GA, Dunn T, Gordon WJ. Deploying digital health tools within large, complex health systems: key considerations for adoption and implementation. NPJ Digit Med 2022 Jan 27;5(1):13. [doi: 10.1038/s41746-022-00557-1] [Medline: 35087160]
44. Chan KS, Zary N. Applications and challenges of implementing artificial intelligence in medical education: integrative review. JMIR Med Educ 2019 Jun 15;5(1):e13930. [doi: 10.2196/13930] [Medline: 31199295]

XSL•FO
RenderX

45. Grunhut J, Marques O, Wyatt ATM. Needs, challenges, and applications of artificial intelligence in medical education curriculum. JMIR Med Educ 2022 Jun 7;8(2):e35587. [doi: 10.2196/35587] [Medline: 35671077]
46. Stanford center for professional development. Stanford University. URL: https://online.stanford.edu/ [accessed 2025-01-14]
47. Artificial intelligence in medicine program. Harvard University. URL: https://aim.hms.harvard.edu/ [accessed 2025-01-14]
48. Heckman GA, Hirdes JP, McKelvie RS. The role of physicians in the era of big data. Can J Cardiol 2020 Jan;36(1):19-21. [doi: 10.1016/j.cjca.2019.09.018] [Medline: 31787436]

Abbreviations

AI: artificial intelligence
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by A Hasan Sapci, B Lesselroth; submitted 07.03.24; peer-reviewed by B Matejko, K Varikara, Z Zandesh; revised version received 04.10.24; accepted 02.01.25; published 05.02.25.

Please cite as:
Gazquez-Garcia J, Sánchez-Bocanegra CL, Sevillano JL
AI in the Health Sector: Systematic Review of Key Skills for Future Health Professionals
JMIR Med Educ 2025;11:e58161
URL: https://mededu.jmir.org/2025/1/e58161
doi:10.2196/58161

© Javier Gazquez-Garcia, Carlos Luis Sánchez-Bocanegra, Jose Luis Sevillano. Originally published in JMIR Medical Education (https://mededu.jmir.org), 5.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Motivation Theories and Constructs in Experimental Studies of Online Instruction: Systematic Review and Directed Content Analysis

Adam Gavarkovs[1], PhD; Erin Miller[2], PhD; Jaimie Coleman[3], MPT; Tharsiga Gunasegaran[4], HBSc; Rashmi A Kusurkar[5], MD, PhD; Kulamakan Kulasegaram[6], PhD; Melanie Anderson[7], MLiS; Ryan Brydges[8], PhD

[1]Division of Continuing Professional Development, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada
[2]School of Physical Therapy, Faculty of Health Sciences, Western University, London, ON, Canada
[3]School of Physical Therapy, University of Toronto, Toronto, ON, Canada
[4]University of Toronto, Toronto, ON, Canada
[5]Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[6]Department of Family and Community Medicine, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[7]University Health Network, Toronto, ON, Canada
[8]Department of Medicine, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Adam Gavarkovs, PhD
Division of Continuing Professional Development
Faculty of Medicine
University of British Columbia
555 West 12th Avenue
Suite 200
Vancouver, BC, V5Z 3X7
Canada
Phone: 1 6046753777
Email: adam.g@ubc.ca

Abstract

Background: The motivational design of online instruction is critical in influencing learners' motivation. Given the multifaceted and situated nature of motivation, educators need access to a range of evidence-based motivational design strategies that target different motivational constructs (eg, interest or confidence).

Objective: This systematic review and directed content analysis aimed to catalog the motivational constructs targeted in experimental studies of online motivational design strategies in health professions education. Identifying which motivational constructs have been most frequently targeted by design strategies—and which remain under-studied—can offer valuable insights into potential areas for future research.

Methods: Medline, Embase, Emcare, PsycINFO, ERIC, and Web of Science were searched from 1990 to August 2022. Studies were included if they compared online instructional design strategies intending to support a motivational construct (eg, interest) or motivation in general among learners in licensed health professions. Two team members independently screened and coded the studies, focusing on the motivational theories that researchers used and the motivational constructs targeted by their design strategies. Motivational constructs were coded into the following categories: intrinsic value beliefs, extrinsic value beliefs, competence and control beliefs, social connectedness, autonomy, and goals.

Results: From 10,584 records, 46 studies were included. Half of the studies (n=23) tested strategies aimed at making instruction more interesting, enjoyable, and fun (n=23), while fewer studies tested strategies aimed at influencing extrinsic value beliefs (n=9), competence and control beliefs (n=6), social connectedness (n=4), or autonomy (n=2). A focus on intrinsic value beliefs was particularly evident in studies not informed by a theory of motivation.

Conclusions: Most research in health professions education has focused on motivating learners by making online instruction more interesting, enjoyable, and fun. We recommend that future research expand this focus to include other motivational constructs,

XSL•FO
RenderX

such as relevance, confidence, and autonomy. Investigating design strategies that influence these constructs would help generate a broader toolkit of strategies for educators to support learners' motivation in online settings.

Trial Registration: PROSPERO CRD42022359521; https://www.crd.york.ac.uk/PROSPERO/view/CRD42022359521

(JMIR Med Educ 2025;11:e64179) doi:10.2196/64179

KEYWORDS

motivation; internet; systematic review; experimental studies; online instruction; educator; learner; researcher; health professional; education; tool-kit; autonomy

Introduction

The internet has become a preferred modality for health professions education (HPE) in the postpandemic landscape [1]. A recent global survey found that 60% of health professionals preferred blended learning, while 32% preferred fully online learning [2]. Online instruction can ameliorate barriers due to geography, scheduling, and cost that make in-person learning infeasible for many health professionals and trainees [3]. However, one challenge of online learning is keeping learners motivated. Motivation—the energetic force that instigates and sustains behavior [4]—is key to success when learning online [5,6]. A lack of face-to-face interaction and the metacognitive demands associated with learning online can lead to feelings of isolation, frustration, and diminished motivation [7,8]. To address these challenges and keep learners motivated, educators must build motivational support into online instruction through a process known as motivational design [9].

Motivational design is defined by Keller [9] as "the process of arranging resources and procedures to bring about changes in people's motivation." This process involves selecting, adapting, and applying motivational design strategies, which are resources and procedures that facilitate the motivational processes underpinning learning. For example, Colonnello et al [10] enhanced medical students' motivation by supplementing surgical videos with emotionally salient patient information. Other studies have demonstrated that other motivational design strategies, such as using narration in online modules, can impact learner motivation [11,12].

Motivational design strategies work by influencing various motivational constructs—cognitive factors that shape learners' moment-to-moment motivation [4]. Broad categories of motivational constructs include goals ("What am I aiming to do?"), competence beliefs ("Can I do it?"), value beliefs ("Do I want to do it? Why?"), and attributional beliefs ("Why did it happen this way?") [13]. For example, an educator might use a strategy to make learning seem more relevant, increase learners' interest, or boost their confidence that they can learn the material.

Theories of motivation emphasize that learners' motivation is influenced by several motivational constructs, any one of which may be the cause of poor motivation during online learning [4]. For example, medical students completing an online module on a basic science topic may be confident in their ability to learn but struggle to see the value in the material beyond their next examination. Conversely, students completing a virtual examination with a standardized patient may see the value in what they are learning but not feel confident in their ability to succeed. In the first case, an educator could use a strategy that targets learners' value beliefs (eg, a prompt to reflect on the clinical relevance of the material [14]), while in the second, an educator could use a strategy that targets' learners competence beliefs (eg, providing a demonstration that learners can observe beforehand [9]). Given the multifaceted and situated nature of motivation, educators need access to a range of evidence-based motivational design strategies that target different motivational constructs, such as strategies for enhancing confidence or perceived value [15].

Researchers can support educators by providing evidence on the effectiveness of different motivational design strategies [16]. However, we do not have a good understanding of which motivational constructs are most frequently targeted in research on online motivational design. For example, are researchers disproportionately focused on testing ways to make online instruction more interesting or enjoyable? An expanding literature on serious games and gamification in HPE suggests this may be the case, as games are often framed as a strategy to enhance interest [17-24]. While enhancing interest is important, if researchers focus too narrowly on this construct at the expense of others (eg, confidence), then educators may not receive the full range of design strategies needed to support learner motivation [4]. To inform future research, it is important to identify which motivational constructs have been most emphasized and which remain under-studied.

To address this gap, our review aims to catalog the motivational constructs targeted in studies of online motivational design strategies. This is a novel objective, as no previous reviews have organized the instructional design literature based on the motivational constructs that strategies aim to influence. By identifying which constructs have received the most attention, we aim to guide future literature syntheses on the most effective design strategies for supporting these constructs. Additionally, by identifying under-studied constructs, we aim to guide areas for future primary research. Ultimately, our review is intended as a resource for researchers interested in conducting future studies on motivational design for online instruction. Stimulating ongoing research in this area will ensure that educators have access to evidence-based guidance to design more motivating online instruction.

We hypothesize that there are two reasons why certain motivational constructs may be underrepresented in research on online motivational design strategies: (1) studies are not informed by a theory of motivation or model of motivational design, or (2) studies are informed by such theories but choose

XSL•FO
RenderX

not to focus on specific constructs. To disentangle these explanations, we posed two research questions: (1) Which theories or models of motivation, if any, inform experimental comparison studies of motivational design strategies for online instruction? (2) Within experimental comparison studies of motivational design strategies for online instruction, which motivational constructs, if any, have been targeted?

Methods

We conducted a systematic review and directed content analysis focused on experimental comparison studies in HPE [25]. Experimental comparison studies, which compare 1 version of online instruction to another, are uniquely positioned to generate empirical evidence for the causal effects of motivational design strategies [25-28]. Motivational design is, at its core, a process of making predictions about the causal effects of motivational design strategies ("If I use this strategy, will it cause my learners to be more motivated?"). Since experimental comparison studies are best suited for making causal claims, we consider them a necessary source of evidence for educators and serve as the focus for our review. Bajpai et al [29] adopted a similar position in their recent review of learning theories in randomized trials of digital instruction in HPE.

Given our focus on experimental comparison studies, we identified a systematic review as the most appropriate review methodology [30]. We registered (PROSPERO CRD42022359521) and published a review protocol [31], and report our findings in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 updated guidelines [32], with a few exceptions. We omit items 12 (effect measures), 14 (reporting bias assessment), 15 (certainty assessment), 19 (results of individual studies), 21 (risk of bias due to missing results), and 22 (certainty of evidence), as we did not intend to appraise nor synthesize the outcomes of included studies. Further details on our methods can be found in our published protocol [31]. To increase the clarity and brevity of reporting, this paper omits data related to a few research questions listed in our published protocol. Additional data regarding these questions is available upon request.

Eligibility Criteria

Study Characteristics

We included individual and cluster randomized controlled trials and quasi-experimental studies published in English from 1990 to August 2, 2022 (for databases) and September 15, 2022 (for registries). Our date range aligns with prior reviews of digital education in HPE [33]. We included protocols for planned or ongoing studies but excluded conference abstracts and unpublished studies. Studies were not excluded based on quality or risk of bias as we did not aim to synthesize the results of studies. However, we appraised the risk of bias to provide readers with additional context regarding the quality of studies.

Participants

We included studies focusing on learners in the health professions regardless of training status (see protocol for list of health professions), either exclusively or when mixed with other learners (eg, psychology students).

Interventions

We included studies comparing online instructional designs (or that could have been delivered online, such as CD-ROM instruction), which targeted a motivational construct (eg, interest) or motivation more generally. By "targeting" motivation, we mean that researchers stated that their instructional design aimed to enhance learner motivation to engage with instruction. Several studies demonstrated a cursory treatment of motivation, for example, by discussing the impact of design strategies on constructs (eg, interest) without grounding the construct in a theoretical framework. We decided to include these studies because they contribute to our understanding of the foci among researchers interested in this area of HPE. Studies comparing online instruction against paper-based or face-to-face instruction were excluded.

Outcomes

We included studies that assessed any learner outcome.

Search Strategy and Selection Process

Database Searching

Strategies were developed for Ovid Medline, Embase, Emcare PsycINFO, EBSCO ERIC, and Web of Science Core Collection (Social Sciences Citation Index; Arts & Humanities Citation Index; Book Citation Index-Social Sciences & Humanities; Conference Proceedings Citation Index-Science; Emerging Sources Citation Index; Science Citation Index; Book Citation Index-Social Sciences & Humanities; and Conference Proceedings Citation Index-Social Science & Humanities) by a health sciences librarian (MA) in collaboration with the review team (Multimedia Appendix 1). Appropriate subject headings and keywords for motivation, online instruction, and HPE focused on the licensed professions were used for each database. The results were limited to those published from 1990 to the date of the searches. The searches were run on August 2, 2022, and the 14,736 results were uploaded to Covidence for screening.

Registry Searching

For the Open Science Framework Registries, we developed 12 searches, comprised of different combinations of the highest yielding terms in our database searches (Multimedia Appendix 2). The searches yielded between 7277 and 16,018 hits for each combination of terms. AG manually screened the first 10 pages of results (10 results per page) for each search (1200 studies screened in total) and uploaded 19 potentially relevant studies to Covidence.

Hand and Reference Searching

AG manually screened several published literature reviews on online instruction in HPE [18-23,34-39] and the references of included studies and uploaded 161 potentially relevant studies to Covidence.

XSL•FO
RenderX

Screening

After removing duplicates, we screened 10,584 records. Two team members independently screened abstracts and, as necessary, the paper's full text. Before independent screening, all 6 team members who participated in the screening process practiced screening the same 30 abstracts, and then discussed and refined the inclusion criteria. AG also developed a decision tool to support full-text screening. As screening progressed, AG periodically reviewed conflicts for any systematic issues and further refined the inclusion and exclusion criteria. Two senior team members (EM or RB) not involved in the initial decision resolved all conflicts. We included 61 studies in the data extraction phase. During the extraction phase we excluded an additional 15 studies. In 12 cases, the papers were excluded because they did not discuss the potential motivational effects of a strategy in the introduction or did not state an objective to assess the effects of a strategy on motivation. Therefore, we concluded that these were not motivational design strategies [40-51]. This yielded 46 studies included in our review.

Data Collection and Synthesis Methods

Overview

The data items we extracted can be found in Multimedia Appendix 3. We conducted a directed content analysis during the extraction process [52], coding each study deductively regarding the motivational theories used and the motivational constructs targeted. We piloted and refined the extraction process in Covidence with a few included studies. AG trained team members to extract and code data. Two team members independently extracted data from each study. Conflicts were resolved through discussion, with an experienced team member (ie, currently in, or having completed, a PhD program) not involved in the initial decision leading to resolution.

Theories of Motivation (Aligned With Research Question 1)

We developed an a priori list of 6 prominent theories of motivation and 1 model of motivational design to deductively guide our coding. We defined theories as "prominent" based on meeting one of the following criteria: (1) they were included in a 2020 special issue of *Contemporary Educational Psychology* titled "Prominent Motivation Theories: The Past, Present, and Future" [53-57], or (2) they have been the subject of an AMEE Guide in *Medical Teacher* [58,59]. We also added Keller's ARCS model of motivational design, which we assumed would be cited in HPE studies [24]. Brief descriptions of these theories can be found in Table 1. Beyond this initial list, we considered any theory aiming to explain the energetic basis and direction of learners' engagement to be a theory of motivation [60]. We also coded whether these theories informed 4 key aspects of the research process: the research questions, the design of the experimental conditions, the selection of methods and measures, and the interpretation of results [61].

XSL•FO
RenderX

Table 1. Overview of and reported use of established theories of motivation and models of motivational design.

Theory or model	Description	Frequency used, n (%)	References
SDT[a] [55]	Ryan and Deci's SDT differentiates between types of motivation depending on learners' reasons for engaging in learning, such as feeling pressured to satisfy external demands (external regulation), feeling pressured to quell feelings of guilt or shame (introjected regulation), identifying with the value of an activity (identified regulation), or finding the activity inherently interesting (intrinsic motivation). SDT also emphasizes the influence of the social environment on learners' motivation, as mediated by the satisfaction of feelings of autonomy (ie, being in control of one's actions), competence (ie, feeling efficacious in one's actions), and relatedness (ie, feeling connected to others).	8 (17)	[10,11,62-67]
ARCS[b] model [9]	Keller's ARCS model states that, for learners to become and remain motivated to learn, their attention must be captured via feelings of curiosity, they must perceive instruction to be relevant to their current needs and long-term goals, they must feel confident that they can succeed, and they must feel satisfied with the intrinsic and extrinsic consequences of engaging with instruction.	6 (13)	[5,68-72]
SCT[c] [56]	Bandura's SCT emphasizes the primary role of learners' self-efficacy beliefs (ie, that they can execute courses of action needed to attain particular outcomes) and outcome expectancies (ie, that courses of action will lead to particular outcomes) in motivating their learning goal pursuit.	3 (7)	[64,73,74]
CVT[d] [75]	Pekrun's CVT posits that the achievement emotions that learners experience (as well as their self-regulation and learning) are most proximally a function of the subjective control and value beliefs they ascribe to actions and outcomes for an activity. Subjective control beliefs are based on action-control expectations (ie, expectations that actions can be performed) and action-outcome expectations (ie, expectations that particular actions will lead to certain outcomes). Subjective value beliefs are based on the perceived intrinsic and extrinsic value of engaging in the activity and attaining resultant outcomes.	2 (4)	[10,63]
EVT[e] [53]	Eccles and Wigfield's EVT (now called situated expectancy-value theory) posits that learners' motivation is most proximally a function of their expectations of success and the subjective value they ascribe to an activity. Subjective value is composed of interest value (ie, the interest or enjoyment an activity brings), utility value (ie, an activity's usefulness for attaining other valued goals), attainment value (ie, an activity's importance in confirming a salient aspect of one's identity), and cost (ie, the drawbacks of completing an activity).	1 (2)	[76]
Other theories or models	Theory of narrative engagement [77,78]; 4-phase model of interest development [11]; engagement modes model [73]; information and communication acceptance model [79]; social interdependence theory [80]; Guthrie and Wigfield engagement model [81]	N/A[f]	See description
None mentioned	N/A	24 (52)	[82-105]

[a]SDT: self-determination theory.

[b]ARCS: attention, relevance, confidence, and satisfaction.

[c]SCT: social cognitive theory.

[d]CVT: control-value theory.

[e]EVT: expectancy-value theory

[f]N/A: not applicable.

Motivational Constructs (Aligned With Research Question 2)

We used our list of theories and previous research [13] to create a priori categories of motivational constructs to deductively guide our coding. During the coding process, our categorization scheme changed slightly from that documented in our protocol [31], as we determined that a more parsimonious categorization scheme involved aggregating more constructs into fewer categories (Multimedia Appendix 4). Our list included the following categories of motivational constructs: intrinsic value beliefs (eg, interest), extrinsic value beliefs (eg, instrumentality), competence and control beliefs (eg, self-efficacy), social connectedness (eg, relatedness), autonomy, and goals. Intrinsic value refers to the value derived from the experience of completing an activity (eg, interest or enjoyment), whereas extrinsic value refers to the value derived from attaining outcomes external to an activity (eg, progress toward future goals) [53,55].

Study Risk of Bias Assessment

We rated each study's risk of bias across 9 dimensions contained within the Cochrane Collaboration's Effective Practice and Organization of Care risk of bias tool: random sequence generation, allocation concealment, similar baseline outcome measurements, similar baseline characteristics, incomplete

XSL•FO
RenderX

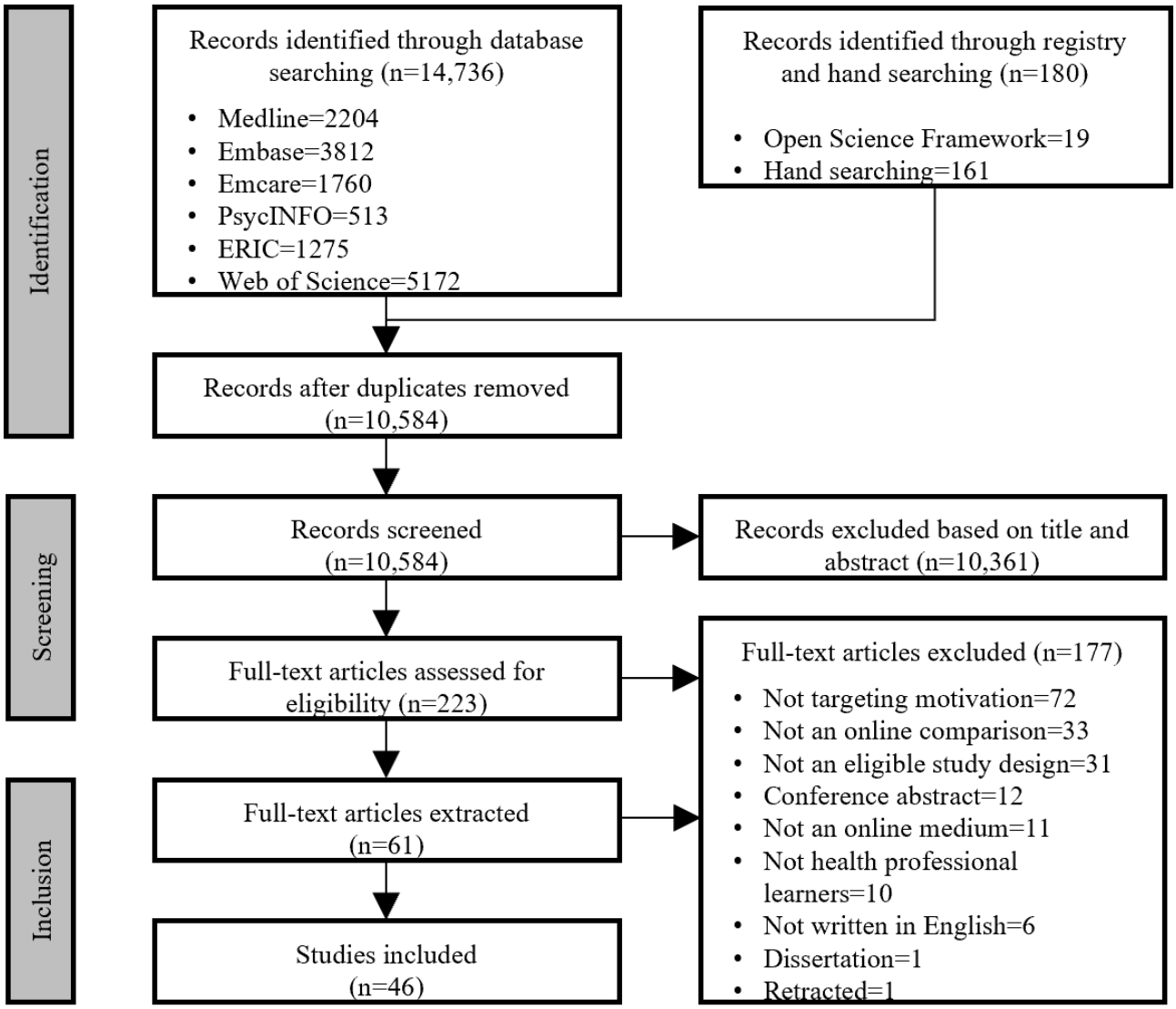
outcome data, blinded outcome measurement, protection against contamination, selective outcome reporting, and other risks of bias [30]. This tool has been used in similar systematic reviews of online instruction in HPE [19,36]. Team members reported particular difficulty in identifying "other risks of bias," and we observed that raters frequently documented different sources of bias (or no bias) within this broad category. Accordingly, we decided to exclude this dimension.

Results

Characteristics of Included Studies

The characteristics of the included studies are presented in Multimedia Appendix 5. Most studies were conducted with trainees (n=40), primarily medical students (n=17) and nursing students (n=11). Study designs were predominantly randomized parallel-group trials (n=27), followed by quasi-experimental trials (n=12), randomized cross-over trials (n=4), and cluster randomized trials (n=3). The risks of bias for each study are presented in Multimedia Appendix 6. Although 74% (34/46) of the included studies were identified as randomized trials, only 30% (14/46) were rated as low risk of bias for random sequence generation, and 33% (15/46) were rated as low risk of bias for allocation concealment. For other dimensions of bias, low risk was observed in 35% (16/46) of studies for baseline outcome measurements, 37% (17/46) for baseline characteristics, 50% (23/46) for blinded outcome measurements, 50% (23/46) for contamination, 57% (26/46) for missing outcome data, and 80% (37/46) for selective outcome reporting. The PRISMA flowchart for our review is presented in Figure 1, and the PRISMA checklist can be found in Multimedia Appendix 7.

Figure 1. PRISMA flowchart. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



Which Theories or Models of Motivation Inform Existing Experimental Studies of Motivational Design Strategies?

Table 1 presents the number of studies that were informed by a theory of motivation or model of motivational design. SDT and the ARCS model were the most commonly used theories, while 24 studies did not cite any theory. Five studies cited more than 1 theory of motivation. Among the 22 studies that used at least 1 theory, we judged the theory as informing the research questions in 20 (91%) studies, informing the experimental conditions in 15 (68%) studies, informing methods and measures

XSL•FO
RenderX

in 17 (77%) studies, and informing the interpretation of results in 17 (77%) studies. Nine studies used theory to inform all 4 aspects of their research process [5,10,11,62,63,68,69,77,80].

Which Motivational Constructs Have Studies Targeted With Their Motivational Design Strategies?

Studies investigated motivational design strategies that targeted intrinsic value beliefs in 23 of the 46 (50%) studies, extrinsic value beliefs in 9 (20%) studies, competence and control beliefs in 6 (13%) studies, social connectedness in 4 (9%) studies, and autonomy in 2 (4%) studies. Ten (22%) studies targeted more than 1 construct; of these, 5 (11%) were informed by the ARCS model. Sixteen (35%) studies did not report targeting any specific motivational construct, instead aiming to enhance motivation in general.

While intrinsic value beliefs were the most commonly targeted construct, researchers drawing on a prominent theory or model (as listed in Table 1) tended to be more pluralistic in their foci. Specifically, studies that used a motivation theory or model targeted intrinsic value beliefs (n=11) at a similar level to extrinsic value beliefs (n=9) and, to a lesser extent, competence and control beliefs (n=6). By contrast, studies that did not use a theory or model focused solely on intrinsic value beliefs (n=10) compared to extrinsic value beliefs (n=0) and competence and control beliefs (n=0).

Discussion

Key Findings and Implications for Future Research

In this systematic review, we analyzed experimental comparison studies of online motivational design strategies in HPE. We aimed to identify which motivational constructs have been most frequently targeted in these studies and which remain understudied, offering insights into potential areas for future research.

A significant finding was that nearly one-third of the studies in our review did not specify which motivational constructs their design strategy was targeting, instead broadly aiming to enhance motivation. We argue that such research is of limited value to educators. Motivational design expertise relies on educators understanding how strategies work, specifically what constructs they influence and under what conditions they are most effective [106,107]. Studies that do not clarify which constructs a design strategy influences, either conceptually or empirically, cannot provide educators with the information needed to build expertise [16]. Therefore, we recommend that researchers explicitly define the motivational constructs their strategies aim to influence and test their impact on those constructs. This recommendation can be supported through the greater use of motivational theories, which were cited in fewer than half of the studies in our review. This lack of theory use is consistent with other reviews, such as those by Maheu-Cadotte et al [19] and Bajpai et al [29], who found similarly low levels of theory use in their reviews of serious games and digital education in HPE. Motivational theory should be used to inform the research questions, the design strategy, the outcome measures, and the interpretation of results. Excellent examples of theory use are present in our sample [5,11,80].

Among the studies that did specify targeted constructs, most focused on intrinsic value beliefs (eg, interest or enjoyment), compared to extrinsic value beliefs, competence and control beliefs, social connectedness, and autonomy. Accordingly, research in this area is disproportionately focused on ways to make online instruction more interesting and enjoyable. Given the volume of studies on design strategies targeting intrinsic value beliefs, we recommend that future research synthesize existing findings to identify the most effective strategies for enhancing interest and enjoyment and outline areas for future research.

A disproportionate focus on enhancing intrinsic value beliefs aligns with an increased uptake of SDT in HPE, as documented in our studies and other reviews [24,108]. SDT emphasizes the role of intrinsic motivation—which is grounded in feelings of interest and enjoyment—in effective learning [55]. However, we found that studies using SDT were often pluralistic in the constructs they targeted, suggesting a more nuanced approach than studies without a theoretical basis. A theoretical perspective, whether based on SDT or another theory, may help researchers avoid equating motivation solely with enjoyment and interest, thus neglecting other facets of motivation, such as confidence and relatedness, despite evidence suggesting that these constructs may be particularly at risk when learning online [7,8]. Supporting this perspective, we found that studies informed by the ARCS model—which explicitly states the importance of supporting learners' attention, relevance, confidence, and satisfaction—were most likely to report targeting multiple motivational constructs. We recommend that studies test design strategies targeting a broader range of motivational constructs to expand the set of design strategies that educators can choose from (eg, confidence-enhancing strategies or relatedness-enhancing strategies). For example, though serious games are often framed as ways to enhance interest and enjoyment, they may also be configured to support feelings of practical relevance or boost confidence [24]. Researchers could build on the serious games literature by investigating ways to design serious games to support feelings of extrinsic value, confidence, social connectedness, and autonomy.

We encourage researchers to study ways of motivating learners in established online modalities (eg, asynchronous modules or webinars) and by using emerging technologies such as virtual reality and artificial intelligence. For example, artificial intelligence chatbots have the potential to provide personalized coaching and feedback during learning [109,110]. Providing such support and scaffolding instruction in a learner's zone of proximal development may foster a sense of autonomy and confidence. As research on the motivational design of emerging online modalities is still in its infancy, future studies could investigate how to design emerging technology-enabled instruction to optimize learner motivation.

The risk of bias was a concern across many of the included studies. To ensure that future research can make more defendable claims regarding the effects of design strategies, researchers should clearly specify procedures for random sequence generation and allocation concealment, which are often missing from published papers. They should also capture

XSL•FO
RenderX

relevant variables at baseline, blind assessors to condition, and attempt to limit attrition and contamination [27].

Limitations

Several limitations are worth noting. We did not include any synonyms for the word "motivation" (eg, "engagement" or "satisfaction") or motivational constructs (eg, "value," "relevance," or "confidence") in our search terms because we believed these terms would greatly increase the number of nonrelevant studies in our search results. We assumed that studies using synonyms for "motivation" or referencing motivational constructs would also use the word "motivation" and thus would be retrieved in our searches. Consequently, we may have missed some otherwise eligible studies that exclusively referenced concepts that are related to, or treated as synonymous to, motivation (eg, engagement) or motivational constructs (eg, confidence). We also chose to exclude studies written in a language other than English, which may have resulted in missed studies.

We decided to focus our review on experimental studies because they provide a critical source of evidence regarding the effectiveness of design strategies. We acknowledge that many different kinds of studies can generate evidence to support educators' motivational design efforts when producing online learning [31,111]. For example, qualitative studies can help us understand how learners make meaning of instructional designs in context [112], and single-group studies can investigate the factors influencing engagement with motivational design strategies [113]. It may be that studies leveraging nonexperimental designs demonstrate a different distribution of foci regarding motivational constructs. We recommend that a breadth of methodologies, including but not limited to experimental comparison studies, be used to investigate novel motivational design strategies in the future.

Finally, our review focused on online instruction in HPE, and it is unclear whether the trends we observed apply to other types of HPE, such as in-person simulation. While the trend toward enhancing interest and enjoyment may also be present in other HPE contexts—such as through the gamification of in-person instruction [114-116]—we cannot make definitive claims about the generalizability of our results to other types of HPE. Conducting similar reviews in other areas of HPE may be a focus of future research.

Conclusions

A key challenge for educators when teaching online involves keeping learners motivated. To address this challenge, educators need access to motivational design strategies that target a range of motivational constructs. The existing research provides an important starting point, but there is much work to be done. Researchers can use our findings to guide future primary and secondary research that generates a more robust evidence base for educators wishing to motivate their learners.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Database search strategies.
[DOCX File , 53 KB - mededu_v11i1e64179_app1.docx]

Multimedia Appendix 2
Registry search strategy.
[DOCX File , 18 KB - mededu_v11i1e64179_app2.docx]

Multimedia Appendix 3
Data items.
[DOCX File , 13 KB - mededu_v11i1e64179_app3.docx]

Multimedia Appendix 4
Categories of motivational constructs.
[DOCX File , 14 KB - mededu_v11i1e64179_app4.docx]

Multimedia Appendix 5
Characteristics of included studies.
[DOCX File , 41 KB - mededu_v11i1e64179_app5.docx]

Multimedia Appendix 6
Risk of bias ratings for included studies.
[DOCX File , 23 KB - mededu_v11i1e64179_app6.docx]

XSL•FO
RenderX

Multimedia Appendix 7
PRISMA 2020 checklist. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
[PDF File (Adobe PDF File), 138 KB - mededu_v11i1e64179_app7.pdf]

References

1. Heldt JP, Agrawal A, Loeb R, Richards MC, Castillo EG, DeBonis K. We're not sure we like it but we still want more: trainee and faculty perceptions of remote learning during the COVID-19 pandemic. Acad Psychiatry 2021;45(5):598-602 [FREE Full text] [doi: 10.1007/s40596-021-01403-4] [Medline: 33594628]
2. Cassidy D, Edwards G, Bruen C, Kelly H, Arnett R, Illing J. Are we ever going back? Exploring the views of health professionals on postpandemic continuing professional development modalities. J Contin Educ Health Prof 2023;43(3):172-180 [FREE Full text] [doi: 10.1097/CEH.0000000000000482] [Medline: 36877815]
3. Cook DA. The value of online learning and MRI: finding a niche for expensive technologies. Med Teach 2014;36(11):965-972. [doi: 10.3109/0142159X.2014.917284] [Medline: 25072533]
4. Cook DA, Artino AR. Motivation to learn: an overview of contemporary theories. Med Educ 2016;50(10):997-1014 [FREE Full text] [doi: 10.1111/medu.13074] [Medline: 27628718]
5. Cook DA, Beckman TJ, Thomas KG, Thompson WG. Measuring motivational characteristics of courses: applying Keller's instructional materials motivation survey to a web-based course. Acad Med 2009;84(11):1505-1509. [doi: 10.1097/ACM.0b013e3181baf56d] [Medline: 19858805]
6. Song HS, Kalet AL, Plass JL. Interplay of prior knowledge, self‐regulation and motivation in complex multimedia learning environments. J Comput Assist Learn 2016;32(1):31-50. [doi: 10.1111/jcal.12117]
7. Butz NT, Stupnisky RH. A mixed methods study of graduate students' self-determined motivation in synchronous hybrid learning environments. Internet High Educ 2016;28:85-95. [doi: 10.1016/j.iheduc.2015.10.003]
8. Scheiter K. The learner control principle in multimedia learning. In: Mayer RE, Fiorella L, editors. The Cambridge Handbook of Multimedia Learning. Cambridge, England: Cambridge University Press; 2021:418-429.
9. Keller JM. Motivational design for learning and performance. In: The ARCS Model Approach. Boston, MA: Springer; 2010:1-345.
10. Colonnello V, Mattarozzi K, Agostini A, Russo PM. Emotionally salient patient information enhances the educational value of surgical videos. Adv Health Sci Educ Theory Pract 2020;25(4):799-808. [doi: 10.1007/s10459-020-09957-y] [Medline: 31960188]
11. Dousay TA. Effects of redundancy and modality on the situational interest of adult learners in multimedia learning. Educ Technol Research Dev 2016;64(6):1251-1271. [doi: 10.1007/s11423-016-9456-3]
12. Bland T, Guo M, Dousay TA. Multimedia design for learner interest and achievement: a visual guide to pharmacology. BMC Med Educ 2024;24(1):113 [FREE Full text] [doi: 10.1186/s12909-024-05077-y] [Medline: 38317141]
13. Pintrich PR. A motivational science perspective on the role of student motivation in learning and teaching contexts. J Educ Psychol 2003;95(4):667-686. [doi: 10.1037/0022-0663.95.4.667]
14. Gavarkovs A, Crukley J, Miller E, Kusurkar R, Kulasegaram K, Brydges R. Effectiveness of life goal framing to motivate medical students during online learning: a randomized controlled trial. Perspect Med Educ 2023;12(1):444-454 [FREE Full text] [doi: 10.5334/pme.1017] [Medline: 37901885]
15. Gavarkovs AG, Glista D, O'Hagan R, Moodie S. Applying the purpose, autonomy, confidence, engrossment model of motivational design to support motivation for continuing professional development. J Contin Educ Health Prof 2025. [doi: 10.1097/CEH.0000000000000595] [Medline: 39907433]
16. Gavarkovs AG, Kusurkar RA, Kulasegaram K, Brydges R. Going beyond the comparison: toward experimental instructional design research with impact. Adv Health Sci Educ Theory Pract 2024. [doi: 10.1007/s10459-024-10365-9] [Medline: 39196469]
17. Nagengast B, Marsh HW, Scalas LF, Xu MK, Hau K, Trautwein U. Who took the "x" out of expectancy-value theory? A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. Psychol Sci 2011;22(8):1058-1066. [doi: 10.1177/0956797611415540] [Medline: 21750248]
18. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, Wortley D, Lilienthal A, Tudor Car L, et al. Serious gaming and gamification education in health professions: systematic review. J Med Internet Res 2019;21(3):e12994 [FREE Full text] [doi: 10.2196/12994] [Medline: 30920375]
19. Maheu-Cadotte M, Cossette S, Dubé V, Fontaine G, Lavallée A, Lavoie P, et al. Efficacy of serious games in healthcare professions education: a systematic review and meta-analysis. Simul Healthc 2021;16(3):199-212. [doi: 10.1097/SIH.0000000000000512] [Medline: 33196609]
20. Min A, Min H, Kim S. Effectiveness of serious games in nurse education: a systematic review. Nurse Educ Today 2022;108:105178. [doi: 10.1016/j.nedt.2021.105178] [Medline: 34717098]
21. Silva RDOS, Pereira AM, Araújo DCSAD, Rocha KSS, Serafini MR, de Lyra Jr DP. Effect of digital serious games related to patient care in pharmacy education: a systematic review. Simul Gaming 2021;52(5):554-584. [doi: 10.1177/1046878120988895]

XSL•FO
RenderX

22. Sipiyaruk K, Hatzipanagos S, Reynolds PA, Gallagher JE. Serious games and the COVID-19 pandemic in dental education: an integrative review of the literature. Computers 2021;10(4):42. [doi: 10.3390/computers10040042]
23. Wang R, DeMaria S, Goldberg A, Katz D. A systematic review of serious games in training health care professionals. Simul Healthc 2016;11(1):41-51. [doi: 10.1097/SIH.0000000000000118] [Medline: 26536340]
24. Krath J, Schürmann L, von Korflesch HF. Revealing the theoretical basis of gamification: a systematic review and analysis of theory in research on gamification, serious games and game-based learning. Comput Human Behav 2021;125:106963. [doi: 10.1016/j.chb.2021.106963]
25. Cook DA. The research we still are not doing: an agenda for the study of computer-based learning. Acad Med 2005;80(6):541-548. [doi: 10.1097/00001888-200506000-00005] [Medline: 15917356]
26. Cook DA, Beckman TJ. Reflections on experimental research in medical education. Adv Health Sci Educ 2010;15(3):455-464. [doi: 10.1007/s10459-008-9117-3] [Medline: 18427941]
27. Shadish WR, Cook TD, Campbell TD. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin; 2001.
28. Mayer RE. How to assess whether an instructional intervention has an effect on learning. Educ Psychol Rev 2023;35(2):64. [doi: 10.1007/s10648-023-09783-9]
29. Bajpai S, Semwal M, Bajpai R, Car J, Ho AHY. Health professions' digital education: review of learning theories in randomized controlled trials by the digital health education collaboration. J Med Internet Res 2019;21(3):e12912 [FREE Full text] [doi: 10.2196/12912] [Medline: 30860483]
30. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane handbook for systematic reviews of interventions version 6.4. Cochrane. The Cochrane Collaboration. 2023. URL: https://training.cochrane.org/handbook/current [accessed 2025-03-21]
31. Gavarkovs A, Kusurkar RA, Kulasegaram K, Crukley J, Miller E, Anderson M, et al. Motivational design for web-based instruction in health professions education: protocol for a systematic review and directed content analysis. JMIR Res Protoc 2022;11(11):e42681 [FREE Full text] [doi: 10.2196/42681] [Medline: 36350706]
32. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev 2021;10(1):89 [FREE Full text] [doi: 10.1186/s13643-021-01626-4] [Medline: 33781348]
33. Car J, Carlstedt-Duke J, Tudor Car L, Posadzki P, Whiting P, Zary N, Digital Health Education Collaboration. Digital education in health professions: the need for overarching evidence synthesis. J Med Internet Res 2019;21(2):e12913 [FREE Full text] [doi: 10.2196/12913] [Medline: 30762583]
34. Arruzza E, Chau M. A scoping review of randomised controlled trials to assess the value of gamification in the higher education of health science students. J Med Imaging Radiat Sci 2021;52(1):137-146. [doi: 10.1016/j.jmir.2020.10.003] [Medline: 33153931]
35. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Instructional design variations in internet-based learning for health professions education: a systematic review and meta-analysis. Acad Med 2010;85(5):909-922. [doi: 10.1097/ACM.0b013e3181d6c319] [Medline: 20520049]
36. Fontaine G, Cossette S, Maheu-Cadotte M, Mailhot T, Deschênes MF, Mathieu-Dupuis G, et al. Efficacy of adaptive e-learning for health professionals and students: a systematic review and meta-analysis. BMJ Open 2019;9(8):e025252 [FREE Full text] [doi: 10.1136/bmjopen-2018-025252] [Medline: 31467045]
37. Szeto MD, Strock D, Anderson J, Sivesind TE, Vorwald VM, Rietcheck HR, et al. Gamification and game-based strategies for dermatology education: narrative review. JMIR Dermatol 2021;4(2):e30325 [FREE Full text] [doi: 10.2196/30325] [Medline: 37632819]
38. Tudor Car L, Kyaw BM, Teo A, Fox TE, Vimalesvaran S, Apfelbacher C, et al. Outcomes, measurement instruments, and their validity evidence in randomized controlled trials on virtual, augmented, and mixed reality in undergraduate medical education: systematic mapping review. JMIR Serious Games 2022;10(2):e29594 [FREE Full text] [doi: 10.2196/29594] [Medline: 35416789]
39. Xu Y, Lau Y, Cheng LJ, Lau ST. Learning experiences of game-based educational intervention in nursing students: a systematic mixed-studies review. Nurse Educ Today 2021;107:105139. [doi: 10.1016/j.nedt.2021.105139] [Medline: 34563963]
40. Becker EA, Godwin EM. Methods to improve teaching interdisciplinary teamwork through computer conferencing. J Allied Health 2005;34(3):169-176. [Medline: 16252680]
41. Crowley RS, Legowski E, Medvedeva O, Tseytlin E, Roh E, Jukic D. Evaluation of an intelligent tutoring system in pathology: effects of external representation on performance gains, metacognition, and acceptance. J Am Med Inform Assoc 2007;14(2):182-190 [FREE Full text] [doi: 10.1197/jamia.M2241] [Medline: 17213494]
42. DeBate RD, Severson HH, Cragun D, Bleck J, Gau J, Merrell L, et al. Randomized trial of two e-learning programs for oral health students on secondary prevention of eating disorders. J Dent Educ 2014;78(1):5-15. [Medline: 24385519]
43. Gauthier A, Corrin M, Jenkinson J. Exploring the influence of game design on learning and voluntary use in an online vascular anatomy study aid. Comput Educ 2015;87:24-34. [doi: 10.1016/j.compedu.2015.03.017]

XSL•FO
RenderX

44. Harned MS, Dimeff LA, Woodcock EA, Kelly T, Zavertnik J, Contreras I, et al. Exposing clinicians to exposure: a randomized controlled dissemination trial of exposure therapy for anxiety disorders. Behav Ther 2014;45(6):731-744 [FREE Full text] [doi: 10.1016/j.beth.2014.04.005] [Medline: 25311284]
45. Hege I, Dietl A, Kiesewetter J, Schelling J, Kiesewetter I. How to tell a patient's story? Influence of the case narrative design on the clinical reasoning process in virtual patients. Med Teach 2018;40(7):736-742 [FREE Full text] [doi: 10.1080/0142159X.2018.1441985] [Medline: 29490538]
46. Hege I, Ropp V, Adler M, Radon K, Mäsch G, Lyon H, et al. Experiences with different integration strategies of case-based e-learning. Med Teach 2007;29(8):791-797. [doi: 10.1080/01421590701589193] [Medline: 18236274]
47. Hendriks WJAJ, Bakker N, Pluk H, de Brouwer A, Wieringa B, Cambi A, et al. Certainty-based marking in a formative assessment improves student course appreciation but not summative examination scores. BMC Med Educ 2019;19(1):178 [FREE Full text] [doi: 10.1186/s12909-019-1610-2] [Medline: 31151456]
48. Kalet AL, Song HS, Sarpel U, Schwartz R, Brenner J, Ark TK, et al. Just enough, but not too much interactivity leads to better clinical skills performance after a computer assisted learning module. Med Teach 2012;34(10):833-839. [doi: 10.3109/0142159X.2012.706727] [Medline: 22917265]
49. Noll C, von Jan U, Raap U, Albrecht U. Mobile augmented reality as a feature for self-oriented, blended learning in medicine: randomized controlled trial. JMIR mHealth uHealth 2017;5(9):e139 [FREE Full text] [doi: 10.2196/mhealth.7943] [Medline: 28912113]
50. Janda MS, Botticelli AT, Mattheos N, Nebel D, Wagner A, Nattestad A, et al. Computer-mediated instructional video: a randomised controlled trial comparing a sequential and a segmented instructional video in surgical hand wash. Eur J Dent Educ 2005;9(2):53-58. [doi: 10.1111/j.1600-0579.2004.00366.x] [Medline: 15811151]
51. Van Es SL, Kumar RK, Pryor WM, Salisbury EL, Velan GM. Cytopathology whole slide images and adaptive tutorials for senior medical students: a randomized crossover trial. Diagn Pathol 2016;11:1 [FREE Full text] [doi: 10.1186/s13000-016-0452-z] [Medline: 26746436]
52. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
53. Eccles JS, Wigfield A. From expectancy-value theory to situated expectancy-value theory: a developmental, social cognitive, and sociocultural perspective on motivation. Contemp Educ Psychol 2020;61(4):101859. [doi: 10.1016/j.cedpsych.2020.101859]
54. Graham S. An attributional theory of motivation. Contemp Educ Psychol 2020;61:101861. [doi: 10.1016/j.cedpsych.2020.101861]
55. Ryan RM, Deci EL. Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions. Contemp Educ Psychol 2020;61:101860. [doi: 10.1016/j.cedpsych.2020.101860]
56. Schunk DH, DiBenedetto MK. Motivation and social cognitive theory. Contemp Educ Psychol 2020;60:101832. [doi: 10.1016/j.cedpsych.2019.101832]
57. Urdan T, Kaplan A. The origins, evolution, and future directions of achievement goal theory. Contemp Educ Psychol 2020;61:101862. [doi: 10.1016/j.cedpsych.2020.101862]
58. Artino AR, Holmboe ES, Durning SJ. Control-value theory: using achievement emotions to improve understanding of motivation, learning, and performance in medical education: AMEE Guide No. 64. Med Teach 2012;34(3):e148-e160. [doi: 10.3109/0142159X.2012.651515] [Medline: 22364472]
59. Ten Cate TJ, Kusurkar RA, Williams GC. How self-determination theory can assist our understanding of the teaching and learning processes in medical education. AMEE guide No. 59. Med Teach 2011;33(12):961-973. [doi: 10.3109/0142159X.2011.595435] [Medline: 22225433]
60. Schunk DH, Pintrich PR, Meece JL. Motivation in Education: Theory, Research, and Applications. 3rd Edition. Upper Saddle River, NJ: Pearson Education Internat; 2010.
61. Cheung JJH, Apramian T, Brydges R. Starting your research project: from problem to theory to question. In: Nestel D, Hui J, Kunkler K, Scerbo M, Calhoun A, editors. Healthcare Simulation Research. Cham, Switzerland: Springer; 2019.
62. Liu C, Lim R, Taylor S, Calvo RA. Students' behavioural engagement in reviewing their tele-consultation feedback within an online clinical communication skills platform. Comput Hum Behav 2019;94:35-44. [doi: 10.1016/j.chb.2019.01.002]
63. Wang M, Wu B, Kirschner PA, Michael Spector J. Using cognitive mapping to foster deeper learning with complex problems in a computer-based environment. Comput Hum Behav 2018;87:450-458. [doi: 10.1016/j.chb.2018.01.024]
64. Zwart DP, Goei SL, Van Luit JEH, Noroozi O. Nursing students' satisfaction with the instructional design of a computer-based virtual learning environment for mathematical medication learning. Interact Learn Environ 2022;31(10):7392-7407. [doi: 10.1080/10494820.2022.2071946]
65. Haftador AM, Shirazi F, Mohebbi Z. Online class or flipped-jigsaw learning? Which one promotes academic motivation during the COVID-19 pandemic? BMC Med Educ 2021;21(1):499 [FREE Full text] [doi: 10.1186/s12909-021-02929-9] [Medline: 34548075]
66. Mahnken AH, Baumann M, Meister M, Schmitt V, Fischer MR. Blended learning in radiology: is self-determined learning really more effective? Eur J Radiol 2011;78(3):384-387. [doi: 10.1016/j.ejrad.2010.12.059] [Medline: 21288674]

XSL•FO
RenderX

67. Rudolphi-Solero T, Lorenzo-Alvarez R, Ruiz-Gomez MJ, Sendra-Portero F. Impact of compulsory participation of medical students in a multiuser online game to learn radiological anatomy and radiological signs within the virtual world Second Life. Anat Sci Educ 2022;15(5):863-876. [doi: 10.1002/ase.2134] [Medline: 34449983]
68. Drees C, Ghebremedhin E, Hansen M. Development of an interactive e-learning software "Histologie für Mediziner" for medical histology courses and its overall impact on learning outcomes and motivation. GMS J Med Educ 2020;37(3):Doc35 [FREE Full text] [doi: 10.3205/zma001328] [Medline: 32566737]
69. Pittenger A, Doering A. Influence of motivational design on completion rates in online self‐study pharmacy‐content courses. Distance Educ 2010;31(3):275-293. [doi: 10.1080/01587919.2010.513953]
70. EL Machtani EL Idrissi W, Chemsi G, EL Kababi K, Radid M. The impact of serious game on the nursing students' learning, behavioral engagement, and motivation. Int J Emerg Technol Learn 2022;17(01):18-35. [doi: 10.3991/ijet.v17i01.26857]
71. Rondon-Melo S, Andrade CRFD. Educação mediada por tecnologia em Fonoaudiologia: impacto na motivação para aprendizagem sobre o Sistema Miofuncional Orofacial. Codas 2016;28(3):269-277 [FREE Full text] [doi: 10.1590/2317-1782/20162015143] [Medline: 27305632]
72. Su C. The effects of students' learning anxiety and motivation on the learning achievement in the activity theory based gamified learning environment. Eurasia J Math Sci Technol Educ 2016;13(5):1229-1258. [doi: 10.12973/eurasia.2017.00669a]
73. Hedman M, Schlickum M, Felländer-Tsai L. Surgical novices randomized to train in two video games become more motivated during training in MIST-VR and GI Mentor II than students with no video game training. Stud Health Technol Inform 2013;184:189-194. [Medline: 23400154]
74. Maag M. The effectiveness of an interactive multimedia learning tool on nursing students' math knowledge and self-efficacy. Comput Inform Nurs 2004;22(1):26-33. [doi: 10.1097/00024665-200401000-00007] [Medline: 15069846]
75. Pekrun R. The control-value theory of achievement emotions: assumptions, corollaries, and implications for educational research and practice. Educ Psychol Rev 2006;18(4):315-341. [doi: 10.1007/s10648-006-9029-9]
76. Wingo MT, Thomas KG, Thompson WG, Cook DA. Enhancing motivation with the "virtual" supervisory role: a randomized trial. BMC Med Educ 2015;15:76 [FREE Full text] [doi: 10.1186/s12909-015-0348-8] [Medline: 25889758]
77. Mohan D, Fischhoff B, Angus DC, Rosengart MR, Wallace DJ, Yealy DM, et al. Serious games may improve physician heuristics in trauma triage. Proc Natl Acad Sci U S A 2018;115(37):9204-9209 [FREE Full text] [doi: 10.1073/pnas.1805450115] [Medline: 30150397]
78. Mohan D, Farris C, Fischhoff B, Rosengart MR, Angus DC, Yealy DM, et al. Efficacy of educational video game versus traditional educational apps at improving physician decision making in trauma triage: randomized controlled trial. BMJ 2017;359:j5416 [FREE Full text] [doi: 10.1136/bmj.j5416] [Medline: 29233854]
79. Lee MK. Effects of mobile phone-based app learning compared to computer-based web learning on nursing students: pilot randomized controlled trial. Healthc Inform Res 2015;21(2):125-133 [FREE Full text] [doi: 10.4258/hir.2015.21.2.125] [Medline: 25995965]
80. Peterson AT, Roseth CJ. Effects of four CSCL strategies for enhancing online discussion forums: Social interdependence, summarizing, scripts, and synchronicity. Int J Educ Res 2016;76:147-161. [doi: 10.1016/j.ijer.2015.04.009]
81. Seibert D, Guthrie J, Adamo G. Improving learning outcomes: Integration of standardized patients & telemedicine technology. Nurs Educ Perspect 2004;25(5):232-237. [Medline: 15508562]
82. Buijs-Spanjers KR, Hegge HH, Jansen CJ, Hoogendoorn E, de Rooij SE. A web-based serious game on delirium as an educational intervention for medical students: randomized controlled trial. JMIR Serious Games 2018;6(4):e17 [FREE Full text] [doi: 10.2196/games.9886] [Medline: 30368436]
83. Allen EB, Walls RT, Reilly FD. Effects of interactive instructional techniques in a web-based peripheral nervous system component for human anatomy. Med Teach 2008;30(1):40-47. [doi: 10.1080/01421590701753518] [Medline: 18278650]
84. Berndt M, Thomas F, Bauer D, Härtl A, Hege I, Kääb S, et al. The influence of prompts on final year medical students' learning process and achievement in ECG interpretation. GMS J Med Educ 2020;37(1):Doc11 [FREE Full text] [doi: 10.3205/zma001304] [Medline: 32270025]
85. Blackmore C, Tantam D, Van DE. The role of the eTutor - evaluating tutor input in a virtual learning community for psychotherapists and psychologists across Europe. Int J Psychother 2006;10(2):35-46 [FREE Full text]
86. Bock A, Thomas C, Heitzer M, Winnand P, Peters F, Lemos M, et al. Transferring the sandwich principle to instructional videos: is it worth the effort? BMC Med Educ 2021;21(1):525 [FREE Full text] [doi: 10.1186/s12909-021-02967-3] [Medline: 34627213]
87. Booth R, Sinclair B, McMurray J, Strudwick G, Watson G, Ladak H, et al. Evaluating a serious gaming electronic medication administration record system among nursing students: protocol for a pragmatic randomized controlled trial. JMIR Res Protoc 2018;7(5):e138 [FREE Full text] [doi: 10.2196/resprot.9601] [Medline: 29807885]
88. Brull S, Finlayson S, Kostelec T, MacDonald R, Krenzischeck D. Using gamification to improve productivity and increase knowledge retention during orientation. J Nurs Adm 2017;47(9):448-453. [doi: 10.1097/NNA.0000000000000512] [Medline: 28834805]
89. Buijs-Spanjers KR, Hegge HHM, Cnossen F, Hoogendoorn E, Jaarsma DADC, de Rooij SE. Dark play of serious games: effectiveness and features (G4HE2018). Games Health J 2019;8(4):301-306. [doi: 10.1089/g4h.2018.0126] [Medline: 30964340]

XSL•FO
RenderX

90. Cao R, Sunaga M, Miyoshi T, Kinoshita A. Development and evaluation of a study level announcement system in e-learning. J Med Dent Sci 2018;65(3):113-122. [doi: 10.11480/jmds.650301]
91. Dankbaar MEW, Alsma J, Jansen EEH, van Merrienboer JJG, van Saase JLCM, Schuit SCE. An experimental study on the effects of a simulation game on students' clinical cognitive skills and motivation. Adv Health Sci Educ 2016;21(3):505-521 [FREE Full text] [doi: 10.1007/s10459-015-9641-x] [Medline: 26433730]
92. Dankbaar MEW, Richters O, Kalkman CJ, Prins G, Ten Cate OTJ, van Merrienboer JJG, et al. Comparative effectiveness of a serious game and an e-module to support patient safety knowledge and awareness. BMC Med Educ 2017;17(1):30. [doi: 10.1186/s12909-016-0836-5] [Medline: 28148296]
93. Frith KH, Kee CC. The effect of communication on nursing student outcomes in a web-based course. J Nurs Educ 2003;42(8):350-358. [doi: 10.3928/0148-4834-20030801-06] [Medline: 12938897]
94. Goldingay S, Land C. Emotion: the 'e' in engagement in online distance education in social work. JOFDL 2014;18(1):58-72. [doi: 10.61468/jofdl.v18i1.226]
95. Hwang GJ, Chang CY. Facilitating decision-making performances in nursing treatments: a contextual digital game-based flipped learning approach. Interact Learn Environ 2023;31(1):156-171. [doi: 10.1080/10494820.2020.1765391]
96. Inangil D, Dincer B, Kabuk A. Effectiveness of the use of animation and gamification in online distance education during pandemic. Comput Inform Nurs 2022;40(5):335-340 [FREE Full text] [doi: 10.1097/CIN.0000000000000902] [Medline: 35266898]
97. Jones EP, Wahlquist AE, Hortman M, Wisniewski CS. Motivating students to engage in preparation for flipped classrooms by using embedded quizzes in pre-class videos. Innov Pharm 2021;12(1):6 [FREE Full text] [doi: 10.24926/iip.v12i1.3353] [Medline: 34007679]
98. Karaksha A, Grant G, Anoopkumar-Dukie S, Nirthanan SN, Davey AK. Student engagement in pharmacology courses using online learning tools. Am J Pharm Educ 2013;77(6):125 [FREE Full text] [doi: 10.5688/ajpe776125] [Medline: 23966728]
99. Koop CFA, Marschollek M, Schmiedl A, Proskynitopoulos PJ, Behrends M. Does an audiovisual dissection manual improve medical students' learning in the gross anatomy dissection course? Anat Sci Educ 2021;14(5):615-628. [doi: 10.1002/ase.2012] [Medline: 33460300]
100. Metz CJ, Metz MJ. The benefits of incorporating active learning into online, asynchronous coursework in dental physiology. Adv Physiol Educ 2022;46(1):11-20 [FREE Full text] [doi: 10.1152/advan.00110.2021] [Medline: 34709946]
101. Pereira AC, Dias da Silva MA, Patel US, Tanday A, Hill KB, Walmsley AD. Using quizzes to provide an effective and more enjoyable dental education: a pilot study. Eur J Dent Educ 2022;26(2):404-408. [doi: 10.1111/eje.12716] [Medline: 34510674]
102. Rajan KK, Pandit AS. Comparing computer-assisted learning activities for learning clinical neuroscience: a randomized control trial. BMC Med Educ 2022;22(1):522 [FREE Full text] [doi: 10.1186/s12909-022-03578-2] [Medline: 35780115]
103. Scales CD, Moin T, Fink A, Berry SH, Afsar-Manesh N, Mangione CM, et al. A randomized, controlled trial of team-based competition to increase learner participation in quality-improvement education. Int J Qual Health Care 2016;28(2):227-232 [FREE Full text] [doi: 10.1093/intqhc/mzw008] [Medline: 26857941]
104. Sward KA, Richardson S, Kendrick J, Maloney C. Use of a web-based game to teach pediatric content to medical students. Ambul Pediatr 2008;8(6):354-359. [doi: 10.1016/j.ambp.2008.07.007] [Medline: 19084784]
105. Woelber JP, Hilbert TS, Ratka-Krüger P. Can easy-to-use software deliver effective e-learning in dental education? A randomised controlled study. Eur J Dent Educ 2012;16(3):187-192. [doi: 10.1111/j.1600-0579.2012.00741.x] [Medline: 22783845]
106. Hardré P, Ge X, Thomas M. Toward a model of development for instructional design expertise. Educ Technol 2005;45(1):53-57.
107. Ertmer PA, Stepich DA, York CS, Stickman A, Wu X, Zurek S, et al. How instructional design experts use knowledge and experience to solve ill-structured problems. Perform Improv Q 2008;21(1):17-42. [doi: 10.1002/piq.20013]
108. Kusurkar RA. Self-determination theory in health professions education research and practice. In: Ryan RM, editor. The Handbook of Self-Determination Theory. New York, NY: Oxford University Press; 2023:665-683.
109. Eysenbach G. The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. JMIR Med Educ 2023;9:e46885 [FREE Full text] [doi: 10.2196/46885] [Medline: 36863937]
110. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023;9:e45312 [FREE Full text] [doi: 10.2196/45312] [Medline: 36753318]
111. Cook DA. The failure of e-learning research to inform educational practice, and what we can do about it. Med Teach 2009;31(2):158-162. [doi: 10.1080/01421590802691393] [Medline: 19330674]
112. Ellaway RH, Pusic M, Yavner S, Kalet AL. Context matters: emergent variability in an effectiveness trial of online teaching modules. Med Educ 2014;48(4):386-396. [doi: 10.1111/medu.12389] [Medline: 24606622]

XSL•FO
RenderX

113. Brisson BM, Hulleman CS, Häfner I, Gaspard H, Flunger B, Dicke A, et al. Who sticks to the instructions—and does it matter? Antecedents and effects of students' responsiveness to a classroom-based motivation intervention. Z Erziehungswissenschaft 2020;23(1):121-144. [doi: 10.1007/s11618-019-00922-z]
114. Rutledge C, Walsh CM, Swinger N, Auerbach M, Castro D, Dewan M, et al. Gamification in action: theoretical and practical considerations for medical educators. Acad Med 2018;93(7):1014-1020. [doi: 10.1097/ACM.0000000000002183] [Medline: 29465450]
115. Singhal S, Hough J, Cripps D. Twelve tips for incorporating gamification into medical education. MedEdPublish (2016) 2019;8:216 [FREE Full text] [doi: 10.15694/mep.2019.000216.1] [Medline: 38089323]
116. Davis K, Lo H, Lichliter R, Wallin K, Elegores G, Jacobson S, et al. Twelve tips for creating an escape room activity for medical education. Med Teach 2022;44(4):366-371. [doi: 10.1080/0142159X.2021.1909715] [Medline: 33872114]

Abbreviations

HPE: health professions education
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by B Lesselroth; submitted 10.07.24; peer-reviewed by S Papadakis, C Lo; comments to author 24.11.24; revised version received 28.02.25; accepted 06.03.25; published 11.04.25.

Please cite as:
Gavarkovs A, Miller E, Coleman J, Gunasegaran T, Kusurkar RA, Kulasegaram K, Anderson M, Brydges R
Motivation Theories and Constructs in Experimental Studies of Online Instruction: Systematic Review and Directed Content Analysis
JMIR Med Educ 2025;11:e64179
URL: https://mededu.jmir.org/2025/1/e64179
doi:10.2196/64179
PMID:

©Adam Gavarkovs, Erin Miller, Jaimie Coleman, Tharsiga Gunasegaran, Rashmi A Kusurkar, Kulamakan Kulasegaram, Melanie Anderson, Ryan Brydges. Originally published in JMIR Medical Education (https://mededu.jmir.org), 11.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Online-Based and Technology-Assisted Psychiatric Education for Trainees: Scoping Review

Mohd Amiruddin Mohd Kassim[1,2,3*], MD; Sidi Muhammad Yusoff Azli Shah[2*], MB Bch BAO; Jane Tze Yn Lim[1,2*], MD, DrPsych; Tuti Iryani Mohd Daud[1,2*], MBBS, DrPsych

[1]Department of Psychiatry, Faculty of Medicine, Universiti Kebangsaan Malaysia, Bandar Tun Razak, Kuala Lumpur, Malaysia

[2]Department of Psychiatry, Hospital Canselor Tuanku Muhriz, Bandar Tun Razak, Kuala Lumpur, Malaysia

[3]Department of Psychiatry and Psychological Health, Faculty of Medicine and Health Sciences, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia

[*]all authors contributed equally

Corresponding Author:
Tuti Iryani Mohd Daud, MBBS, DrPsych
Department of Psychiatry
Faculty of Medicine
Universiti Kebangsaan Malaysia
Jalan Yaacob Latif
Bandar Tun Razak, Kuala Lumpur, 56000
Malaysia
Phone: 60 3 9145 6143
Email: tutimd@hctm.ukm.edu.my

Abstract

Background: The concept of online learning in medical education has been gaining traction, but whether it can accommodate the complexity of higher-level psychiatric training remains uncertain.

Objective: This review aims to identify the various online-based and technology-assisted educational methods used in psychiatric training and to examine the outcomes in terms of trainees' knowledge, skills, and levels of confidence or preference in using such technologies.

Methods: A comprehensive search was conducted in PubMed, Cochrane, PsycINFO, Scopus, and ERIC to identify relevant literature from 1991 until 2024. Studies in English and those that had English translations were identified. Studies that incorporated or explored the use of online-based or technology-assisted learning as part of psychiatric training in trainees and had outcomes of interest related to changes in the level of knowledge or skills, changes in the level of preference or confidence in using online-based or technology-assisted learning, and feedback of participants were included. Studies were excluded if they were conducted on populations excluding psychiatric trainees or residents, were mainly descriptive of the concept of the intervention without any relevant study outcome, were not in English or did not have English translations, or were review articles.

Results: A total of 82 articles were included in the review. The articles were divided into 3 phases: prior to 2015, 2015 to 2019 (prepandemic), and 2020 onward (postpandemic). Articles mainly originated from Western countries, and there was a significant increase in relevant studies after the pandemic. There were 5 methods identified, namely videoconference, online modules/e-learning, virtual patients, software/applications, and social media. These were applied in various aspects of psychiatric education, such as theory knowledge, skills training, psychotherapy supervision, and information retrieval.

Conclusions: Videoconference-based learning was the most widely implemented approach, followed by online modules and virtual patients. Despite the outcome heterogeneity and small sample sizes in the included studies, the application of such approaches may have utility in terms of knowledge and skills attainment and could be beneficial for the training of future psychiatrists, especially those in underserved low- and middle-income countries.

(JMIR Med Educ 2025;11:e64773) doi:10.2196/64773

KEYWORDS

online learning; telepsychiatry; remote learning; virtual; training; education; psychiatry; trainees; residents

XSL•FO
RenderX

Introduction

The incorporation of online-based or technology-assisted methods in medical education is not new. Virtual grand rounds, web-based learning, online journal clubs, and virtual clinical cases and labs are among the many examples of their ubiquitous implementation [1]. The mass adoption of technology-based education is attributed to its numerous perceived advantages, including the ability to transcend geographical boundaries, the presence of learner-centered approaches, the development of students' self-directed learning skills, and the asynchronous interaction between teachers and students. As education is getting more globalized due to increasing connectivity, these benefits are being increasingly valued [2].

However, questions remain about whether the many advantages of online education are as intuitively apparent and relatable in the field of psychiatry. Traditionally considered as a face-to-face medical discipline, concerns arise regarding the unique interpersonal nature of psychiatry, with its emphasis on empathetic responsiveness toward patients. These concerns are particularly relevant when considering a virtual or simulated patient, and this represents one of the frontier aspects of online education [3]. It is an undeniable fact that appreciating cues from patients is an experiential aspect of knowledge, which is often deemed irreplicable in online sessions.

As technology-based education is increasingly recognized as being noninferior to physical education in undergraduate studies [4], it is imperative to investigate its application in training postgraduate students. The outcome is significant, as it pertains to the production of future psychiatric specialists. This inquiry is especially relevant today, given the radical and drastic transition to technology-based education at all levels due to the recent COVID-19 pandemic [5]. Most medical fraternities are able to integrate online-based and technology-assisted components in their syllabi to enhance the training of trainees or residents without much difficulty. However, acknowledging the distinct nature of psychiatry, which often can be rather ambiguous and subject to nuance, it is important to evaluate the suitability of such an approach to augment the training of future psychiatrists.

Given these considerations, the overarching goal of this study is to systematically map and summarize the existing literature on online-based and technology-assisted psychiatric education for trainees. Specifically, this review aims to identify the various online-based and technology-assisted educational methods used in psychiatric training and to examine the outcome of the aforementioned technologies in terms of trainees' knowledge, skills, or levels of confidence or preference in using such technologies. We hypothesized that online-based and technology-assisted education can be integrated into psychiatric training to improve trainees' knowledge, skills, and competency levels.

This review is aimed at psychiatric educators and training directors looking for ways to incorporate technology into their programs, psychiatric trainees who want to understand how online learning fits into their training, and policymakers or accreditation bodies shaping the future of psychiatric education. It is also relevant for researchers and academics interested in digital learning and medical education trends.

Methods

Search Strategy

The scoping review was conducted according to the recent methodological framework by Westphaln et al [6], which was derived from the earlier work of Arksey and O'Malley [7]. Five databases (PubMed, PsycINFO, Cochrane, Scopus, and ERIC) were searched from March until June 2024. The keywords applied in PubMed were as follows: ((“resident*”[Title/Abstract] OR “trainee*”[Title/Abstract] OR “postgrad*”[Title/Abstract] OR “graduate*”[Title/Abstract]) AND (“psychiatr*”[Title/Abstract] OR “psychologic* medicine”[Title/Abstract]) AND (“education”[Title/Abstract] OR “training”[Title/Abstract] OR “development”[Title/Abstract] OR “learning”[Title/Abstract] OR “teaching”[Title/Abstract] OR “internship”[Title/Abstract] OR “traineeship”[Title/Abstract] OR “residency”[Title/Abstract] OR “course”[Title/Abstract] OR “lesson”[Title/Abstract] OR “program”[Title/Abstract] OR “programme”[Title/Abstract] OR “class”[Title/Abstract] OR “workshop”[Title/Abstract] OR “module”[Title/Abstract] OR “mooc”[Title/Abstract] OR “academic”[Title/Abstract] OR “clerkship”[Title/Abstract] OR “curriculum”[Title/Abstract]) AND (“on-line”[Title/Abstract] OR “online”[Title/Abstract] OR “digital”[Title/Abstract] OR “virtual”[Title/Abstract] OR “internet-based”[Title/Abstract] OR “internet based”[Title/Abstract] OR “web-based”[Title/Abstract] OR “web based”[Title/Abstract] OR “telepsychiatry”[Title/Abstract] OR “tele-psychiatry”[Title/Abstract] OR “cyber”[Title/Abstract] OR “electronic”[Title/Abstract] OR “e-learning”[Title/Abstract] OR “tele-education”[Title/Abstract] OR “videoconferencing”[Title/Abstract] OR “elearning”[Title/Abstract] OR “distance”[Title/Abstract])).

Different search configurations were used for the databases, and the search strategies are presented in Multimedia Appendix 1.

To ensure completeness, the authors also conducted backward citation searches from key articles and performed searches in Google Scholar to look for grey literature, such as conference proceedings and theses, relevant to the topic. Google Scholar was adopted as it has extensive coverage of academic work and is one of the commonly used search engines for grey literature [8].

Inclusion and Exclusion Criteria

Acknowledging the emergence of the field and the relatively limited number of studies, the authors made a conscious decision to include a variety of publication types in this review, including original articles, empirical and brief reports, case reports, and short communications. Studies were included in the review if they met the following criteria: (1) incorporated or explored the use of online-based or technology-assisted learning as part of psychiatric training or education; (2) were conducted in populations that included psychiatric trainees or residents; (3) had an outcome of interest related to changes in the level of knowledge or skills or the level of preference or confidence in

XSL•FO
RenderX

using online-based or technology-assisted learning, or included feedback of participants on the aforementioned approach (regardless of qualitative or quantitative results); and (4) were written in English or had English translations.

Studies were excluded if they (1) were conducted on populations excluding psychiatric trainees or residents; (2) were mainly descriptive of the concept of the intervention without an evaluation or any relevant study outcome related to the application of online-based or technology-assisted psychiatric education; or (3) were review articles. In addition, studies that only used online questionnaires to conduct pre-post assessments for psychiatric education, which were otherwise not delivered through an online or technology-assisted platform, and studies that primarily assessed the learning needs in online-based or technology-assisted psychiatric education without an evaluation of the intervention itself were also excluded. Studies conducted in languages other than English and those without an English translation were omitted due to limitations in language proficiency and to prevent inaccuracies or misinterpretation of the study findings.

Data Screening and Extraction

The search results from the 5 databases were exported to Rayyan online reference manager. Duplicates of similar articles detected by Rayyan were screened manually by MAMK and SMYAS to minimize errors in excluding articles. Prior to the title and abstract screening process, both MAMK and SMYAS underwent screening training to promote standardization and to identify possible conflicts. Then, according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) approach, the titles and abstracts were screened by MAMK and SMYAS to assess suitability for further examination based on the following predetermined criteria:

- Did the study use online-based or technology-assisted instruments as part of the education approach?
- Was the study focused on psychiatric education or related necessary skills?
- Was the study conducted in samples that included psychiatric trainees or residents?

In this phase, articles were divided into the following categories: accept, maybe, and exclude. Subsequently, the full texts of eligible articles (accept and maybe categories) were retrieved. This was followed by a blinded screening phase during which MAMK and SMYAS independently examined the articles in accordance with the inclusion and exclusion criteria. Any disputes regarding the acceptability of the articles in the title and abstract screening phase and in the full-text screening phase were resolved by an impartial third referee (JTYL or TIMD). Data extraction from all included studies was then conducted, gathering parameters such as author names, study year and country, aims and objectives, interventions applied, and key outcomes of interest. The data were extracted by MAMK and SMYAS to Excel (Microsoft Corp) sheets with predefined data fields.

The quality of the studies was assessed according to the 10-item Medical Education Research Study Quality Instrument (MERSQI) for quantitative studies [9] and the Standards for Reporting Qualitative Research (SRQR) for qualitative studies [10]. Previously, a scoping review adopted the SRQR as a 21-score checklist to assess the quality of included qualitative studies within the review [11].

Ethical Considerations

This review received ethical approval from the Research Ethics Committee of the National University of Malaysia (JEP-2023-789).

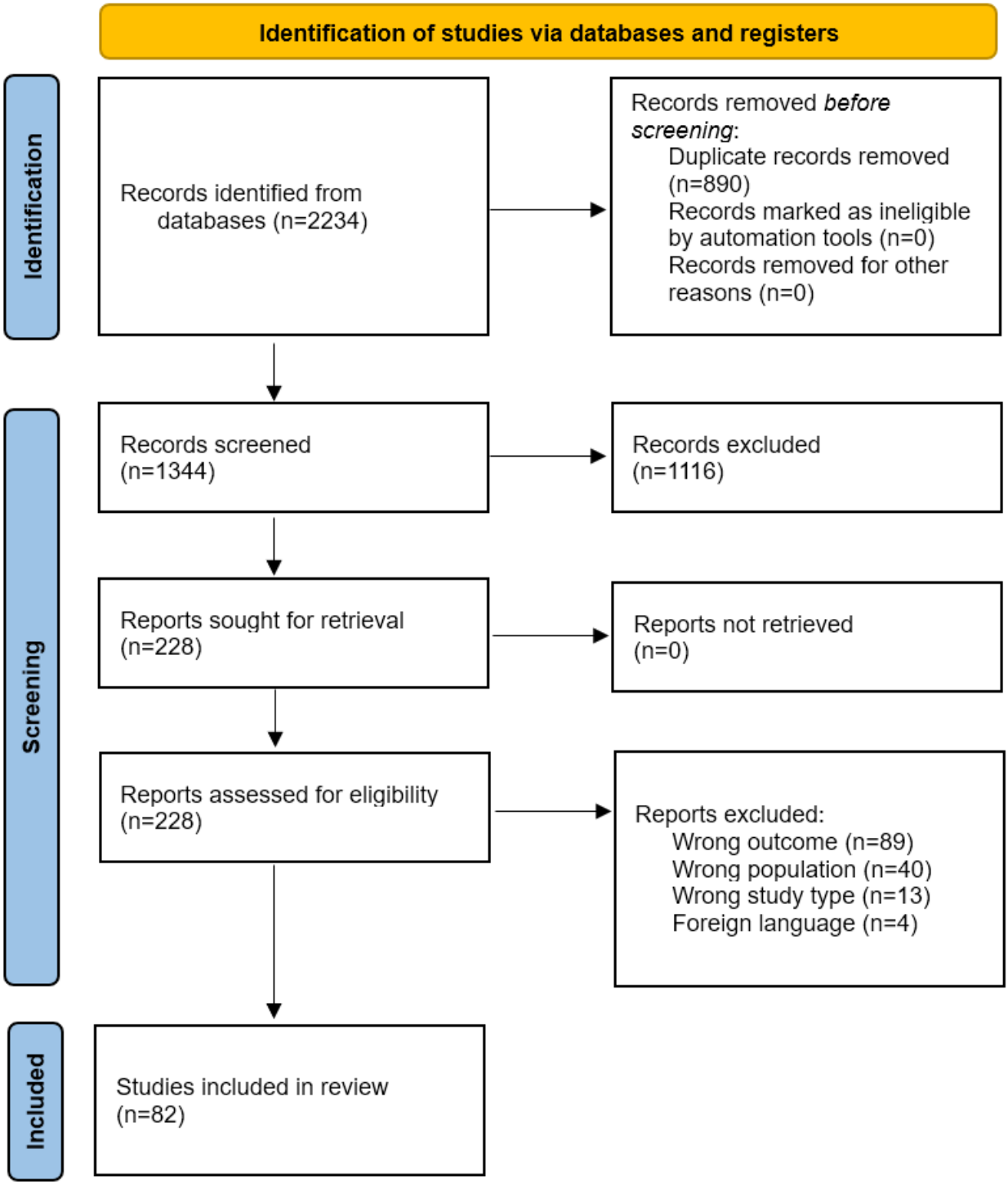
Results

Overview of the Included Studies

The initial search across the 5 databases yielded a total of 2234 articles, from which 890 duplicates were subsequently removed. After screening the titles and abstracts, 1344 articles were excluded and 228 articles proceeded to full-text screening. Of these, 89 articles were excluded for wrong outcomes (not evaluating the outcome of interest), 40 articles for wrong population (samples excluded psychiatric trainees or residents), 13 articles for wrong study type (review study design), and 4 articles for being in a foreign language. Thus, 82 articles were included in this review (Figure 1).

XSL•FO
RenderX

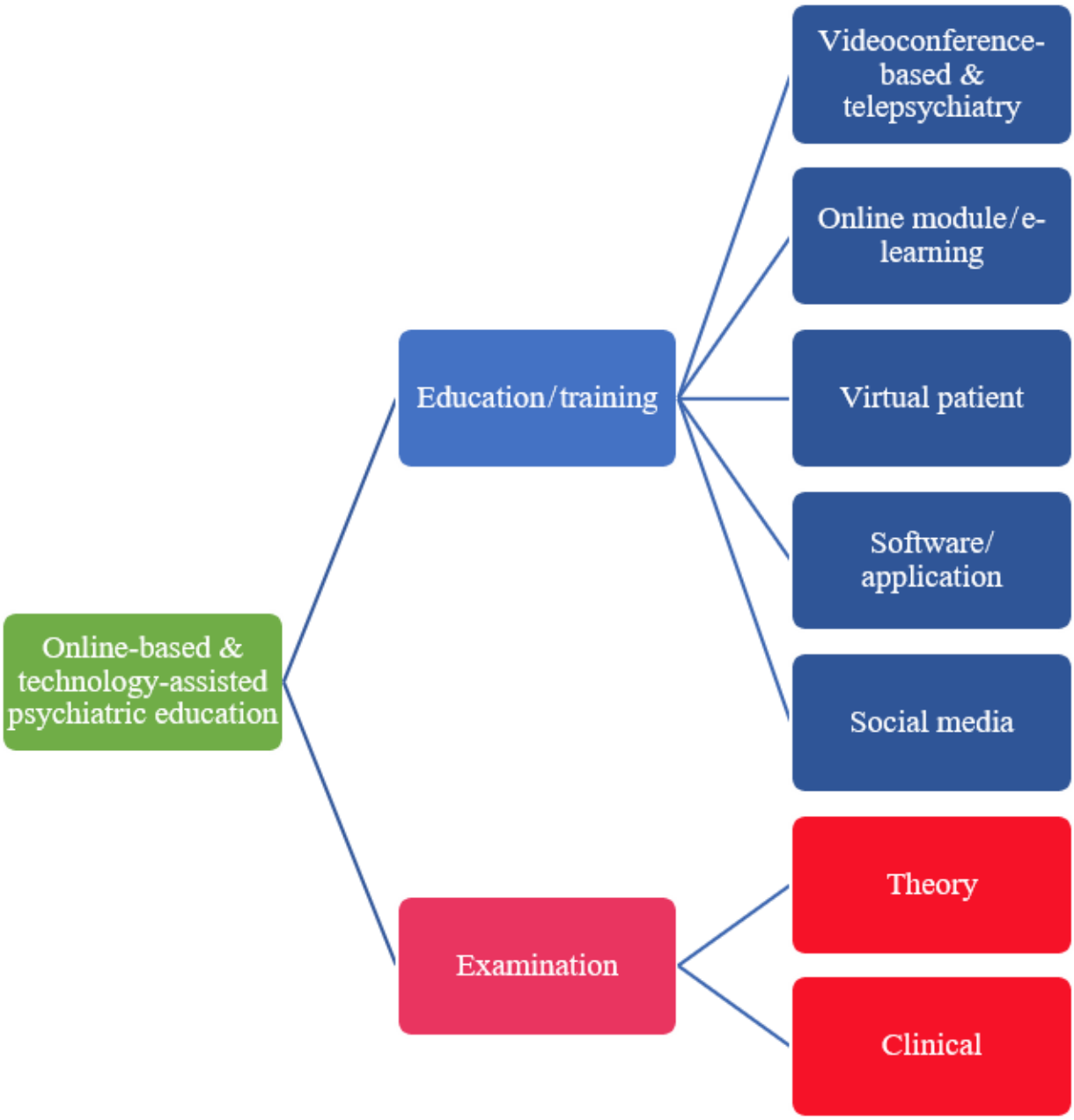
Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.



Among the 82 articles included, 2 themes were identified: education and assessment (Figure 2). Under education, the trend could be divided into 5 subthemes, namely online software (or e-learning platform or massive open online course [MOOC]), videoconference (or telepsychiatry), virtual patient (or simulation), software/application, and social media. Under assessment, there were 2 subthemes, namely theory and clinical examination.

XSL•FO
RenderX

Figure 2. Trend of the implementation of online-based and technology-assisted psychiatric education.

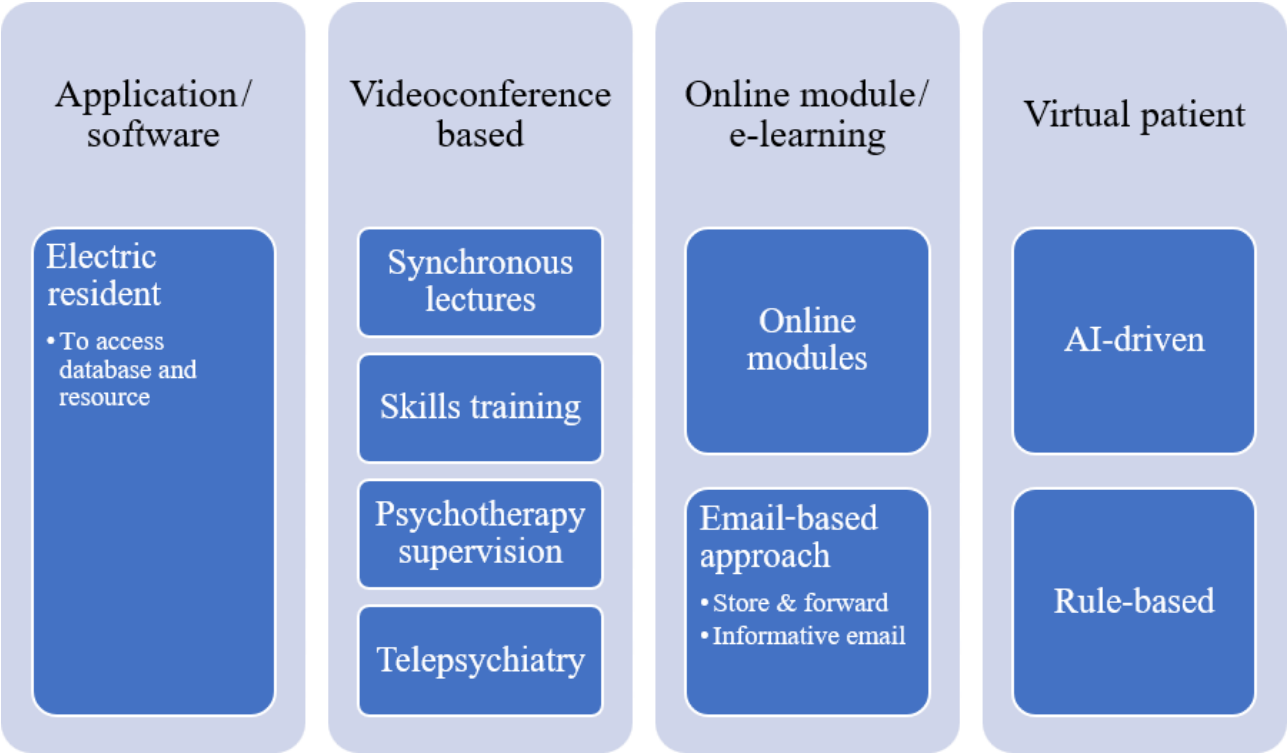


Studies Prior to 2015

A total of 20 articles published before 2015 were included (Table S1 in Multimedia Appendix 2). Specifically, 2 articles were from the 1990s, 9 articles were from the 2000s, and 9 articles were from 2011 to 2014. These articles were mainly from Western countries, such as the United States and Canada. The articles primarily involved telepsychiatry and videoconferencing as tools for education, training, and supervision (n= 10/20, 50%), followed by web-based approaches or e-learning (n= 5/20, 25%), virtual patients (n= 3/20, 15%), and software (n=1/20, 5%). The taxonomy of the methods applied in psychiatric education during this period is presented in Figure 3. Studies mainly had a 1-group, posttest-only study design; 1-group, pretest-posttest study design; and randomized controlled trial design. Out of the 20 articles, 8 had objective measures, such as the change in the level of knowledge before and after the intervention, while the other 12 had subjective measures, such as the level of satisfaction and participant feedback.

XSL•FO
RenderX

Figure 3. Taxonomy of online-based and technology-assisted learning prior to 2015. AI: artificial intelligence.



Despite the limitation of the internet in earlier years, there have been some attempts to integrate IT as part of training. Within this period, depending on financial capabilities and regions, few types of internet connections were available, for example, dial-up connection (especially in the 1990s), Integrated Service Digital Network (ISDN), and broadband connection. The type of internet connection has influenced the strategy and the experience in the education process. With slower internet, an indirect instructional model or independent study was used, with trainees searching for relevant information to enhance their knowledge in particular topics or cases that they had been consulted on. One of the earliest studies highlighted the use of computers and the internet for accessing MEDLINE to assist residents with their consultations, checking drug interactions, and reviewing literature pertinent to cases that they were consulted on [12]. In a survey, printed materials were preferred when learning something new, but digital media or online resources were preferred when revising or searching for resources during patient care [13].

As seen in other fields of medicine, the advancement of relevant infrastructure enabled faster internet connectivity, which allowed widespread and accessible knowledge sharing through videoconference-based seminars, and this is also applicable in the field of psychiatry. This has allowed direct instructional models through approaches such as online lectures and seminars. However, these approaches were met with mixed feedback. One study in the United States in 2004 highlighted that satisfaction with videoconference-based lecturers was contingent on the internet speed, with trainees or residents in centers having a slower internet speed reporting less satisfaction with the lecture experience and the overall sound quality [14]. Another study in Australia in 2008 reported higher preference among participants to attend seminars from remote sites, and most participants felt that the videoconference-based seminars were beneficial for their practice [15]. Meanwhile, in a study in South Africa by Chipps et al [16], while videoconferencing was perceived as an excellent education tool by half of the psychiatric registrars, only 39% of them felt that it was as effective as face-to-face teaching. This led to decreased interest in further videoconference-based training. Additionally, another randomized controlled trial in Iran in 2014, which aimed to compare the effectiveness of face-to-face communication skills training sessions against distant learning in improving empathy, found that the level of empathy was significantly increased in the attending group but not in the distant learning group [17].

One study in Norway in 1998 explored the use of videoconferencing technology in terms of psychotherapy supervision [18]. The psychiatric trainees conducted face-to-face psychotherapy sessions with their patients and later had alternating face-to-face and online psychotherapy supervisions with their supervisors. Through semistructured interviews after the completion of the psychotherapy session, it was noted that while the reduced nonverbal cues were an issue, the limitations of the videoconferencing supervision paradoxically had some positive effects among the trainees in terms of the supervision process, such as verbalization and structure. The positive effects were also contributed by the ease of logistics and by having a neutral space separate from the supervisor's office.

As an extension to videoconference use, telepsychiatry serves as a valuable tool to expand the reach of psychiatric services. As such, it has been incorporated as part of training for psychiatric trainees or residents. In terms of supervision, most of the studies included direct, side-by-side supervision by attendings for assessing patients [19-22]. On the other hand, 1 study adopted a different approach, with attending psychiatrists sitting in with residents during their first session to help

XSL•FO
RenderX

familiarize them with conducting treatment via telepsychiatry, and in later sessions, the involvement of supervising attendings was on an "as needed basis" [23]. Most feedback by trainees on telepsychiatry programs indicated that telepsychiatry enhanced their skills and knowledge, with majority of trainees stating that it was interesting and enhanced their training. However, some trainees mentioned technical issues with this approach and the difficulty in assessing the influence on patients.

On the other hand, improved access to the internet has expanded the utility of asynchronous learning methods. In earlier years, a web-based email approach was applied to promote exposure or learning about stigma education [24] and child and psychiatry cases [25] through approaches such as the "store and forward" concept. However, in Western countries where technology was more advanced compared to the rest of the world at that point of time, e-learning materials were typically in the form of slides of didactic content with recorded audios and videos. This delivery method was used in learning evidence-based medicine [26] and to improve electrocardiogram reading skills [27]. A study by Garside et al [28] in 2009 managed to introduce direct and interactive instruction strategies to learn about how to fill Form 1 of the Mental Health Act. This was achieved by integrating slides of relevant materials regarding Form 1 and the laws related to it, together with interactive Flash animations and practice cases, using questions, and there was immediate expert feedback for each question. Throughout these studies, there were statistically significant improvements in the levels of knowledge and skills of the trainees, suggesting the potential of such an approach to augment the training of psychiatric trainees.

Interest in a virtual patient as an education tool for psychiatric training emerged in the 2000s and 2010s, and facilitated more immersive learning. Within this period, 2 types of virtual patients were studied: artificial intelligence (AI)-driven virtual patients and rule-based virtual patients. Kenny et al [29] and Pataki et al [30] described the use of an AI-driven virtual patient to simulate an adolescent patient with posttraumatic stress disorder (PTSD) ("Justina"). The virtual patient was developed according to the criteria of PTSD based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) [31] and involved technologies such as voice recognition, response selection, behavior generation, and a visual graphics engine. "Justina" received good feedback from residents who mentioned that the experience they had from assessing the virtual patient closely matched their actual experience, but there were times when the virtual patient was not able to understand the questions from the residents. In another study, the rule-based virtual patient concept was applied to assess the doctor's competence in obtaining informed consent before prescribing antipsychotics in a simulated patient with psychosis [32]. A Flash-based video was shown to introduce the clinical scenario, followed by a series of menu options from which they could choose their next action. After completion of the scenario, the program provided a feedback screen regarding the appropriateness of their actions, as well as a link to relevant resources to improve their knowledge. There were statistically significant improvements in all items of the Confidence Scale (pretest to posttest).

Studies From 2015 to 2019 (Before the COVID-19 Pandemic)

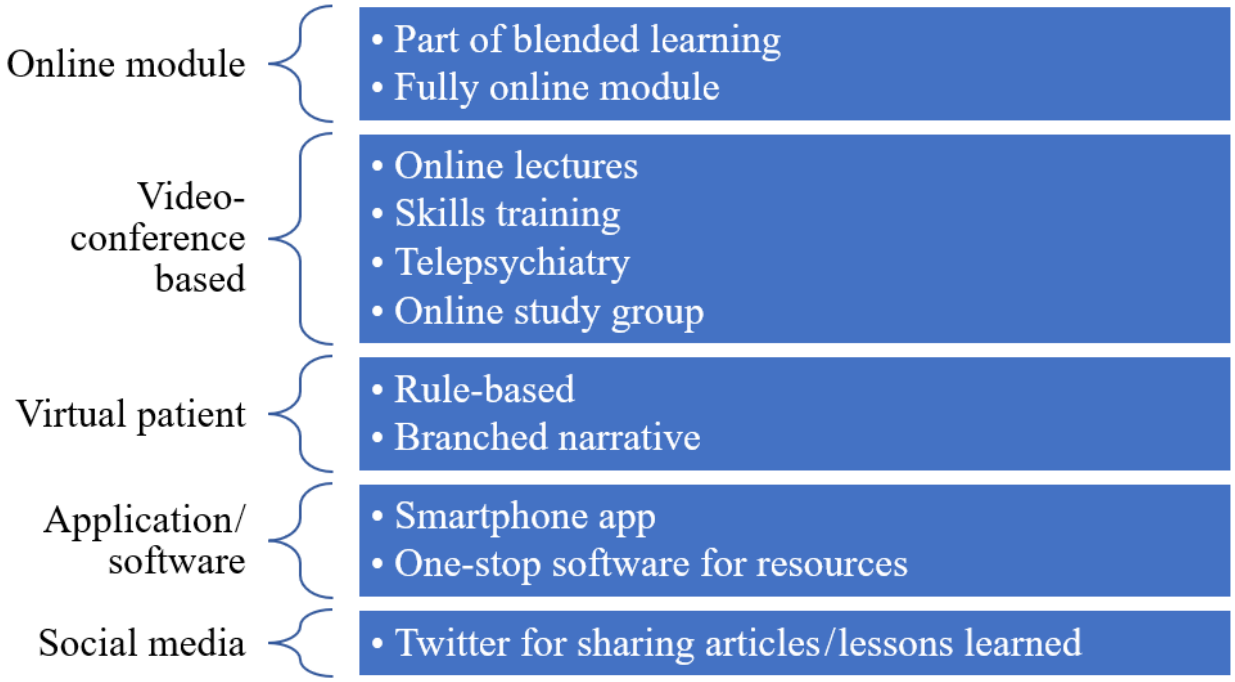
During the period from 2015 to 2019, there was an acceleration of internet-based or technology-assisted education in psychiatry, with 21 studies identified over 5 years (Table S2 in Multimedia Appendix 2). Western countries, particularly the United States and Canada, were at the forefront of these initiatives. Majority of the studies involved videoconference-based learning (n= 7/21, 33.3%), followed by use of online modules or e-learning platforms (n= 5/21, 23.8%), development of software or applications (n= 3/21, 14.3%), adoption of virtual patient approaches (n= 3/21, 14.3%), and use of social media as a learning tool (n= 1/21, 4.8%) (Figure 4). The most common study design was a 1-group, pretest-posttest design, followed by a 1-group, posttest-only design. There was only 1 randomized controlled trial and 1 crossover intervention study. Studies mainly applied subjective measures to assess the outcome of interest, and there was an increasing trend of objective measures in 9 studies, mainly to assess changes in knowledge.

There was more organized and comprehensive use of online resources, with UpToDate, PubMed, Wikipedia, and e-journals being popular among trainees [33]. However, some residents cited insufficient time, insufficient faculty guidance, and lack of resources specific to psychiatry as barriers to using online resources in their practice. Moreover, 86% of respondents felt that there is a need for more psychiatry-specific online resources, and 79% believed that online resources should be more visual and interactive. In another survey, some trainees preferred to read or take notes on paper for academic purposes [34]. However, they still preferred on-screen reading for checking medication dosing and information.

As accessibility to the internet improved with wider coverage and greater stability, there was an increased effort to conduct online modules or e-learning to complement the traditional methods of psychiatric learning, at least for delivering theoretical aspects. For example, Hickey et al [35] conducted a blended course of traditional lectures, online modules, and videotape reviews in psychotherapy education. The authors developed 2 modules on Davanloo Intensive Short-term Dynamic Psychotherapy (IS-DTP), which consisted of PowerPoint presentations, videos, and pre- and posttests. In a crossover intervention design, residents were divided into 2 groups, with each group receiving an online module and a face-to-face lecture but in a different order for 2 topics. It was found that there was a statistically significant improvement in knowledge acquisition in both the online module group and face-to-face lecture group, and there was no significant difference in comparison between both groups regardless of how the topics were delivered.

XSL•FO
RenderX

Figure 4. Taxonomy of online-based and technology-assisted learning in 2015-2019.



Three studies described blended learning but with a different model compared to the earlier study [36-38]. For example, a bookend blended learning model was adopted, in which there was self-paced virtual learning via online modules in the beginning, followed by synchronous in-person interactive discussion, and the approach ended with case supervision or reflection. This approach was used in learning PTSD [36] and in an integrative psychiatry curriculum [37], and residents appreciated the convenience and suitability to augment their training owing to 2 main factors. The first factor was that the self-paced online modules allowed for more time for personalized training and to process materials, and the second factor was the ability to access quality information linked directly to published sources. In another variation of blended learning, another study applied the flipped classroom model, in which the required reading was provided earlier, and then, participants engaged in interactive asynchronous discussion [38]. Interestingly, less than half of the participants indicated improvement in their knowledge, while the rest felt no difference, and a small number of participants felt that the approach had a negative impact on their knowledge, indicating the need to re-evaluate the suitability of the asynchronous flipped classroom approach for psychiatric training.

Fully online courses or modules have been continuously adapted to different aspects of psychiatric training. Brief online courses or modules were employed in rather specific topics under psychiatric-related theoretical knowledge, such as catatonia [39], tobacco use disorder [40], and substance use disorder [41]. Certain modules can be completed in 10 minutes, and brief modules often include slide presentations and relevant videos (recorded webinars, patient interviews, or demonstrations of symptoms and signs), which trainees can complete at their own pace. Despite the brevity of the approach, there were significant increases in knowledge [39,40] or improved levels of stigma [41] between pre- and posttests, with some of the improvements being retained at the 3-month or 6-month follow-up, suggesting the utility of such an approach.

Videoconference-based discussions or webinars remain useful tools in psychiatric education. Few studies used videoconference platforms to deliver online lectures and online case discussions tailored for aspects of psychiatric training, such as career development activities [42] and continuous education in conflict zones [43]. Videoconference platforms have also been used for skills training. Puspitasari et al [44] conducted a study on online behavioral activation training over 4 weeks with the aim to compare trainer-led online training and self-paced online training. There was a significant increase in behavioral activation skill assessment total scores in both groups, and there were significant between-group differences favoring the trainer-led online group at both the posttraining assessment and 3-month follow-up. Although direct instructional strategies were applied in both groups and both approaches were conducted online, the interactive component of the session in the trainer-led group caused a significant difference in the improvement of the skills of the trainees.

Building on the earlier work of Pignatiello et al [21] and Volpe et al [22], a study by Teshima et al [45] focused on feedback from over 300 residents in telelink telepsychiatry training in Canada. The trainees appreciated the opportunity to learn about different approaches to interviewing clients. Although the residents found the technology a “bit unnatural” at the beginning of the session and realized that it was “challenging to interview [patients] at a distance,” they were still able to obtain nonverbal cues, which were important to understand their patients. Almost all participants agreed that the telepsychiatry experience was interesting, while 97% agreed with the statement “the experience helped me understand more about providing psychiatric services to underserved areas.” As the telelink program was an established and mandatory program for psychiatric trainees in University of Toronto, this study had one of the largest sample

XSL•FO
RenderX

sizes in comparison to other studies, enhancing its reliability despite its single study site.

Davidson and Evans [46] illustrated how a videoconference platform was employed for an online study group as an augmentation tool in preparing for the Royal Australian and New Zealand College of Psychiatrist (RANZCP) objective structured clinical examination (OSCE). Four New Zealand trainees used Google Hangout (now known as Google Meet) for their online OSCE practice, with the exam questions based on online past papers on the RANZCP website [46]. The members rotated their role to be a candidate, examiner, or role-player, adding to the experiential learning. The 4 trainees passed their OSCE and acknowledged the benefit of a virtual study group in enhancing their preparation.

As demonstrated by earlier studies, the concept of a rule-based virtual patient was proven useful in learning PTSD and significantly improved the knowledge and confidence level of residents in managing PTSD [47,48]. In another variation, Wilkening et al [49] employed a branched-narrative virtual patient for advanced psychopharmacology sessions, in which residents were presented with a challenge and given choices, and the consequences would depend on the choice selected. The integration of the virtual patient concept led to statistically significant improvements in knowledge levels in advanced psychopharmacology, supporting the efficacy of the virtual patient concept as part of psychiatric education.

Abundance of reliable resources is a boon to evidence-based medicine; however, due to the hectic nature of clinicians, a quick decision must be made quite often. Thus, few software programs were developed to act as a 1-stop center for reference to help expedite and guide their practice. Adeponle et al [50] developed the Psychiatry Toolkit, which allowed direct, immediate, and full access (institutional login) to desired journals, articles, and relevant databases, including PubMed, PsycINFO, and UpToDate [50]. Another study by Dirlam et al [51] described Mental Health EMR Tools, which is a large database that allows residents to access prevetted, curated, and continuously updated information to help with their clinical practice. In addition, acknowledging the potential of a smartphone as an educational tool, Zhang et al [52] developed the Delirium University Health Network Application as a tool for delirium education. It was initially developed as an online application and was later piloted as a smartphone app via the Android Play Store. The app included many important contents related to delirium, such as the DSM-5 diagnostic criteria for delirium, common causes of delirium, pharmacological and nonpharmacological interventions for delirium, and objective assessment questionnaires (eg, Confusion Assessment Method [CAM]). Overall, Mental Health EMR Tools and the delirium smartphone app received good feedback from users, who appreciated the convenience of getting information from a consolidated source. On the other hand, while the Psychiatry Toolkit helped residents to look for answers to their clinical questions, the adoption rate of the toolkit among respondents was relatively low at 47% [50].

An interesting study described an innovative approach to adjunct psychiatric education using social media. Walsh et al [53] described the use of Twitter to disseminate education resources considered helpful in training. Under Twitter account @PhippsPsych, residents took turns to post tweets or retweet contents, such as take-home points from psychiatry grand rounds, links to journal articles, and references to psychiatry in current events. While the study had a rather small sample size with 49 residents, there was a significant increase in the proportion of participants using Twitter for medical education from 8.2% to 28.6%. However, residents' ratings regarding the usefulness of social media in medical education did not change from pre- to postsurvey, and corroborated by the fact that 60% of residents reported that the knowledge gained from following the account had no impact on their clinical practice, 37.2% reported a minimal or average impact and only 2.8% reported a great impact.

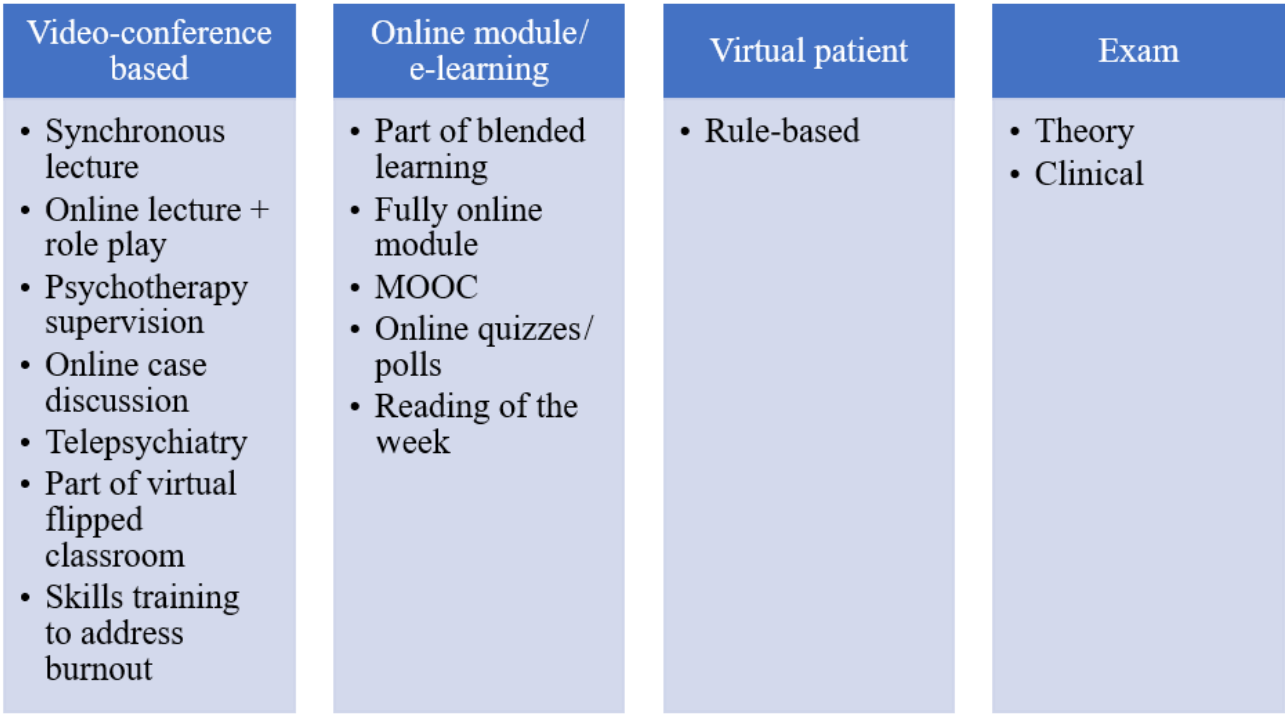
Studies From 2020 Onward (After the COVID-19 Pandemic)

The number of articles on online-based or technology-assisted learning in psychiatry education for psychiatric residents or trainees saw a significant spike during and after the COVID-19 pandemic. There were 41 relevant articles from 2020 until June 2024 (Table S3 in Multimedia Appendix 2). While the United States, Canada, and the United Kingdom had the highest number of studies, there was also notable involvement from Global South countries, such as Malaysia, Thailand, Pakistan, India, and Tunisia. Majority of the articles involved the application of videoconference-based learning or webinar concepts (n= 24/41, 58.5%) and online modules (n= 8/41, 19.5%) (Figure 5). Studies mainly had subjective measures as outcomes and had a 1-group posttest-only study design. There were 11 studies with a 1-group pretest-posttest design and 1 study with a nonequivalent group posttest-only design.

Transition to videoconference-based learning and webinars was necessary for the continuation of training and education during the COVID-19 pandemic, including for psychiatric residents. There were few variations in how videoconference-based learning was applied. In some articles, it was rather straightforward with synchronous online lectures through the videoconference platform, and some of the lectures were then followed by virtual group discussions or brainstorming sessions in the breakout rooms of the platform to make it more interactive. This approach was commonly used in the theoretical aspects of psychiatric training, such as in alcohol use disorder [54], fundamentals of remote psychotherapy [55], research in psychiatry [56], digital psychiatry [57], complex child and adolescent cases [58], biostatistics and methodology courses [59], and journal clubs [60].

XSL•FO
RenderX

Figure 5. Taxonomy of online-based and technology-assisted learning in 2020-2024. MOOC: massive open online course.



In another variation, the online lecture was paired with virtual role-play or simulation sessions. With the added experiential learning component, residents were able to apply their knowledge accordingly in case scenarios. In a study by Blamey et al [61], psychiatric trainees attended a 2-hour virtual lecture on the necessary skills for their on-call work, and then, the trainees participated in a series of 2-hour simulated on-call shifts once a week for 10 weeks, covering 10 common scenarios for psychiatric on-call work. Acknowledging the importance of understanding the complexities of health systems in delivering effective and safe patient care, Li et al [62] developed an online curriculum for core competencies in health systems science. The residents underwent 10-minute virtual didactics prior to the virtual simulation of case scenarios using the Zoom platform.

Looking at the outcomes, trainees or residents perceived a high level of satisfaction with the program and its online delivery, as well as an increase in confidence in skills and perceived learning gains [56,57,60]. Among the studies, 1 study had an objective outcome, in which significant improvements in the knowledge level regarding alcohol use were noted among residents after the videoconference-based lecture and recorded training video session, indicating the potential efficacy of such a program [54]. The virtual role-play concept was especially credited to be a useful technique to enhance the interactive learning of residents [61]; however, the virtual format can be awkward owing to the need for turn-taking, which in turn affects the interactivity, especially when there are few residents involved concurrently in a single case scenario [62].

As demonstrated by Gammon et al [18], a videoconference platform may be used for psychotherapy supervision. Due to the COVID-19 pandemic, this was of value, and there were 2 studies involving psychiatric trainees receiving psychotherapy supervision virtually. There were concerns regarding the loss of nonverbal cues or subtleties of communication during remote supervision, in addition to the “Zoom fatigue” phenomenon, all of which influenced residents to favor face-to-face supervision more than remote supervision [63]. However, the flexibility of remote supervision and the option to allow residents to attend the supervision even when they were busy with their ward work or were on-call were certainly advantageous, and the quality of psychotherapy skills attainment based on subjective assessments by supervisors was not significantly different between remote supervision and traditional face-to-face supervision [63,64]. In fact, as reported by Famina et al [65], supervising attendings noted that the quality of psychiatric care was not different between remote sessions and in-person sessions, and there was not much difference in terms of the ability to empathize and to interpret nonverbal cues.

With regard to telepsychiatry, the authors of an article mentioned their experience of a sudden unprecedented change to their service, which involved a transition to telepsychiatry due to the COVID-19 pandemic [66]. The service initially involved phone consultation, and residents and attendings subsequently switched to video consultation after obtaining approval to use a video platform [66]. While both trainees and attendings strongly agreed that the change to virtual care was necessary, the attendings felt that trainee supervision and training worsened during the pandemic. The trainees also felt less comfortable conducting virtual care and less confident in their assessments to the extent that they found video consultations “frustrating,” especially when attempting to interview patients who had difficulty engaging in virtual interactions (eg, those with delirium, neurocognitive disorders, or mania) [66]. This

XSL•FO
RenderX

experience was echoed in a survey by Cruz et al [67], which showed that the top 5 concerns shared by residents and the faculty about telepsychiatry were the inability to perform a physical exam, poor internet connection, unknown liability risks related to telepsychiatry, certain cultures being less accepting, and nonverbal cues being missed.

Contrary to that experience, the authors of another study described their telepsychiatry experience in a rather positive note. Because of a rapid shift, telepsychiatry sessions were still following the prior model of in-person direct supervision involving attendings and residents, and both had to don a mask while conducting the telepsychiatry sessions, which affected the voice projection and the ability of the patient to hear the treatment plan [68]. Subsequently, with further understanding of the videoconference platform, they were able to continue direct supervision with slight modification as attendings joined the session from their private offices, allowing residents and attendings to remain mask-free and improving the audio for patients. Majority of the residents felt that telepsychiatry had positively impacted their clinical education experience, and it was significantly associated with comfort with practicing telepsychiatry in the future [68].

The flipped classroom model, which is a type of blended learning, reversed the settings, with direct instructions to be provided at home and learning activities involving higher order thinking to be done at school. COVID-19 restrictions necessitated the change to a fully virtual flipped classroom, and psychiatric training was not an exception. A study from Pakistan described a program using the flipped classroom model for an online trauma curriculum [69]. Under this program, reading materials and videos were provided to trainees earlier, and then, the trainees participated in virtual brainstorming, role-play, and case-based discussions. On a larger scale, the Metis didactic courses for psychiatric residents in Sweden (the pedagogical model has been in line with the flipped classroom concept from its inception in 2007) were switched to a digital format to ensure continuation of learning [70]. Each course consists of 3 phases: distance-based self-study, classroom-based meeting days for lectures and supervision, and distance-based examination. The second phase was subsequently transitioned to an online classroom for the same activities. While the fully online flipped classroom concept improved the level of knowledge and skills of psychiatric trainees, some residents preferred to return to face-to-face learning [69,70]. Interestingly, female participants and those aged younger than 50 years were more inclined to continue with online-based course meetings [70].

There was a growing issue of burnout among psychiatric trainees due to the pandemic and its consequences. As such, training programs included skills training as a necessary curriculum component to address burnout, and these were delivered virtually using videoconference platforms, such as the virtual Balint group [71,72], Mind-Body skills program [73], brief mindfulness-based cognitive behavioral therapy (CBT)-informed virtual well-being program [74], and virtual medical improvement program [75]. An interesting example was the virtual Balint group, an initiative that provides a cathartic space and helps to improve morale. With the idea to improve the understanding of patients' problems rather than finding solutions, residents were encouraged to participate with the camera on and the mic on mute when someone was presenting, and the presenter was free to express their experience of doctor-patient interactions, with the guarantee of nonjudgment and confidentiality [72]. During the discussion, the hands-up function of the Zoom platform was used when someone wanted to speak in order to control flow and avoid interruption. Participants were very positive of the virtual Balint group, with trainees feeling well supported by this initiative. However, many participants preferred face-to-face sessions but nevertheless would choose an online session over no session at all. In another study, the virtual Balint group was credited for promoting a sense of connectedness among peers and providing freedom to speak without needing to censor themselves. However, the virtual nature of the Balint group itself led some participants to feel an abrupt ending to the session, which does not occur in face-to-face sessions [71].

In terms of online modules, the most common design was problem-focused case vignettes, alongside interactive presentation and audiovisual content. Some of the modules also had tests at the end. This design was adopted in few of the studies to cover various aspects of psychiatry training, such as forensic psychiatry [76], catatonia [77], neuropsychiatry [78], tobacco use disorder [79], cultural sensitivity [80], and antiracism intervention [81]. A study by Owais et al [82] used the same online module concept in a blended learning approach for an electroconvulsive therapy curriculum, together with didactic seminars and hands-on clinical management. By combining indirect and direct instruction strategies, there was significant improvement in terms of knowledge attainment after the modules (smaller sample sizes) [76-78,82], and generally, the modules had high satisfaction levels reported by residents [80,81].

While most modules were confined to certain training institutions or regions, 1 article described a larger-scale MOOC to augment psychiatric training. Gargot et al [83] described the First European Psychiatric Association MOOC on CBT, which lasted for a month. With a focus on the theoretical aspects of CBT through recorded lectures, presentations, online forums, and online examinations, the self-paced MOOC had large participation, with 7116 participants enrolling from at least 49 countries. Although the eventual completion rate was 26%, a large number of participants (n=1828) completed the MOOC and the average score for the tests increased steadily from 21.4 out of 25 in the first week to 23.13 out of 25 in the final week, indicating the potential of the MOOC to fill the training gap.

In another study, a website was developed as an innovative, free psychiatry Continuing Professional Development (CPD) resource for Canadian psychiatrists and residents. Referred to as Reading of the Week, this website summarizes the latest psychiatric literature, provides expert commentaries, and promotes discussions on social media platforms [84]. The innovations in psychiatric education as described in these articles received good feedback from trainees. For the Reading of the Week initiative, in which the survey evaluation was based on the 6-level evaluation framework by Moore et al [85], positive feedback and satisfaction were reported by participants across

XSL•FO
RenderX

the 6 levels, including knowledge outcomes (level 3), behavior outcomes (levels 4 and 5), and practice outcomes (level 6) [84].

The pandemic also forced the examination process to be performed in a digital format. Generally, examinations in psychiatric training can be divided into theory examinations and clinical examinations (including OSCE). The Royal College of Psychiatrists conducted Member of the Royal College of Psychiatrists theory examinations via a digital platform using a combination of AI and in-person online proctoring [86]. Multiple choice questions were directly assessed, but for questions involving very short answers, smart algorithms were developed to recognize versions of correct answers, and answers that were nonexact matches were reviewed by a designated examiner. On the other hand, for the assessment of clinical psychiatry skills, there was mixed feedback from both examiners and trainees. Depending on the format of the clinical examination, examiners generally manned the stations or the breakout rooms. Integrating videoconference technology for the purpose of clinical examinations had inherent issues, such as connectivity problems, a sense of disconnect, lack of a framework to mentally reset, difficulty in building rapport, and an inadequate capacity to assess clinical skills [87]. However, some residents believed that online assessments were convenient for both participants and patients, reducing anxiety by being in a familiar environment and improving patient access [88]. Some of the candidates even stated that virtual communication was nearly as good as face-to-face communication and online examination was “better than expected” [89].

There is a paucity of studies comparing online training or learning with face-to-face or in-person training. In a quasiexperimental study in Germany to explore whether the satisfaction of online CBT training is noninferior to that of in-person CBT training, the 2 study groups had the same theoretical CBT content, a similar duration of training, comparable audiences, and an identical trainer [90]. The online training was conducted according to the inverted-classroom concept, with participants being required to watch recorded video lectures on the Moodle platform and then have a Zoom discussion at a fixed time for 6 to 7 sessions. It was found that evaluations of the online training group were noninferior to those of the in-person group in terms of information content, didactic presentation, assessment of the trainer as a suitable role-model, working atmosphere, own commitment, and practical relevance, suggesting that the delivery of CBT knowledge through an online platform may be sufficient.

In another study, Hewson et al [91] described a rather indirect comparison between face-to-face basic psychiatry skills simulation training and synchronous online training. The transition to online training via Zoom was due to evolving COVID-19 restrictions at that time and was not in the initial plan. In subgroup analyses, the face-to-face group showed statistically significant improvements in confidence across all psychiatry skills tested, whereas the online group showed significant improvements in confidence in all but 2 skills, namely psychiatric risk assessment and assessment of physical health problems in elderly patients with cognitive impairment. However, the face-to-face group included foundation doctors (junior doctors) and the online group included psychiatry and general practitioner trainees, suggesting that the lack of a significant improvement in confidence in those 2 skills could be related to a higher baseline self-confidence level prior to the simulation training.

A rule-based virtual patient appears to be a mainstay model of a virtual patient in psychiatric education. Rakofsky et al [92] developed a virtual patient-based assessment simulator as a tool to assess the proficiency of residents regarding psychopharmacological knowledge and practice. Combining virtual human avatars, AI, and an advanced pedagogical design, it allows for a realistic interaction, including live voice communication. According to the rule-based virtual patient concept, residents had choices of questions and answers to choose from, and they were given immediate feedback on all their choices alongside the rationale. Looking at the performance of the residents, the mean total score of the simulator by class correlated significantly with the mean score of the somatic therapies subscale of the Psychiatry Residency in Training Exam (PRITE), suggesting construct validity of the virtual patient simulator.

In a survey assessing residents’ perceptions of the pandemic’s impact on their didactic experience and training preferences, it was found that trainees appreciated several positive aspects of virtual didactics, such as being easy to attend and being engaging, and they were able to invite guest speakers from other institutions easily [93]. However, some negative experiences were also reported, including the “Zoom fatigue” phenomenon and frequent distractions, and some topics did not translate well to a virtual environment. Residents from Thailand, which was hit hard by the pandemic and had a significant shift in psychiatric training to online sessions, also reported mixed experiences. Although all residents had good results and passed their examinations, they felt that studying online and the uncertainty with virtual psychotherapy were major inconveniences in their training [94]. In another study, residents were ambivalent. They perceived face-to-face teaching to be superior, but majority of them did not think a complete return to in-person learning would be the most effective option when this becomes possible, implying a preference to continue with some online components in the training [95].

A summary of the key benefits and limitations of the 5 different online-based and technology-assisted educational methods (videoconference, online module/e-learning, virtual patient, software/applications, and social media) is provided in Table 1.

XSL•FO
RenderX

Table 1. Summary of the implementation of online-based and technology-assisted psychiatric education.

Method	Key benefits	Limitations
Videoconference	• Flexible and applicable for various objectives (lectures, skills training, psychotherapy supervision, etc) • Accommodates different instructional strategies (direct, indirect, interactive, and experiential) • Convenience of attending sessions regardless of location or schedule	• Relies on the internet speed • Struggles with nonverbal cues • Frequent distractions and Zoom fatigue
Online module	• Allows self-paced learning • Numerous relevant materials can be designed and included (animations, prerecorded videos, and quizzes) • Possibility of reaching a wide range of audiences via a MOOC[a]	• Lacks direct supervision • More suitable for theoretical aspects of training than clinical aspects • Often requires collaboration and resources to design the modules
Virtual patient	• Valuable for learning uncommon cases • Engaging learning experience	• Requires high levels of resources for development • Can be frustrating to interact in case of speech recognition issues
Software/applications	• Serves as a convenient point of reference • Integration with a smartphone • Highly favored in checking medication dosing and information	• Requires certain levels of resources for development • Software that primarily acts as a gateway for institutional logins to certain websites is not often used
Social media	• Promotes a continuous learning opportunity • Possibility of greater dissemination of knowledge to a larger audience • Provides information on current evidence-based studies	• Reported knowledge gain that translates into clinical practice is still less significant • Relies on the effort of the individual to follow the account and review the shared resource

[a]MOOC: massive open online course.

Discussion

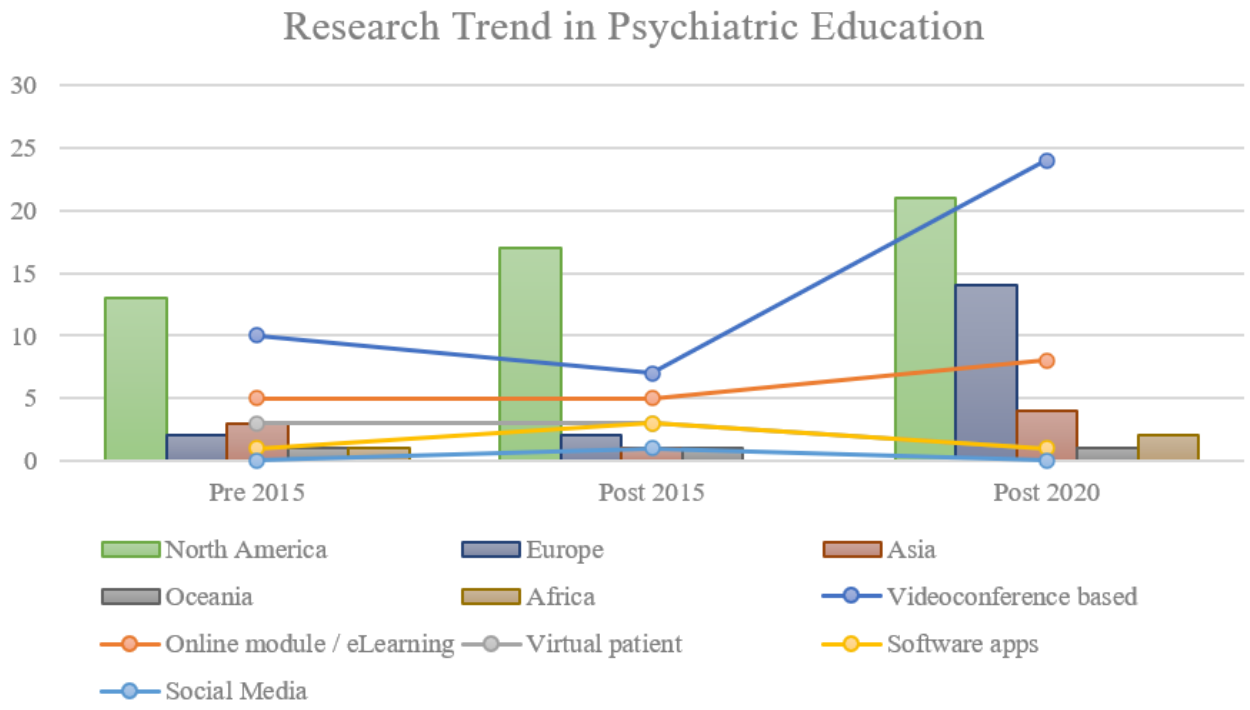
Summary of the Results

The findings across the 3 phases (prior to 2015, 2015-2019 [prepandemic], and 2020 onward [postpandemic]) illustrated the creative integration of online-based and technology-assisted learning in psychiatric education for trainees (Figure 6). Five approaches were identified: videoconference, online module/e-learning, virtual patient, software/applications, and social media. These methods were used for various objectives, including but not limited to teaching theory knowledge, skills training, psychotherapy supervision, and information retrieval. North American countries were leading in research output, followed by European countries. There was consistent research presence from countries in Asia, Africa, and Oceania throughout the 3 phases, with a slight increase after 2020.

The trend of online-based and technology-assisted psychiatric education showed changes from one phase to another. Videoconference-based learning was consistently the preferred way of integrating technology into the learning or training of trainees. Meanwhile, there was a growing trend of online module/eLearning platform use, virtual patient use, and development of applications or software from the "prior to 2015" phase until the prepandemic phase (2015-2019). The trend however stagnated in the postpandemic phase. Videoconference-based learning was more dominantly used due to 3 factors. First, as seen in other fields, the sudden shift to online was not accompanied by the readiness of other technology modalities to assume the responsibility to continue the learning process [96]. Second, the improved accessibility and availability of videoconference platforms, such as Zoom, Google Meet, and Microsoft Teams, served as a plus point [97], and less resources were needed to shift learning to those platforms rather than developing online modules in an expedited manner. The third factor corresponds to the nature of psychiatric learning, which is preferably face-to-face, but among the choices that were presented, videoconference-based learning was a method that may offer some compromise in terms of allowing direct and synchronous instructional methods despite physical distance [98].

XSL•FO
RenderX

Figure 6. Trend of online-based and technology-assisted psychiatric education in the 3 study phases (prior to 2015, 2015-2019 [prepandemic], and 2020 onward [postpandemic]).



In terms of didactic teaching, evidence showed that online-based or technology-assisted learning is beneficial and well-tolerated by residents. Barring minor issues pertaining to connectivity, a recurring problem reported by several studies [59,65,78], theoretical learning of core knowledge in psychiatry through online platforms has been helpful in improving knowledge, and the impact cannot be understated. Some trainees or residents even expressed their preference for receiving academic or continuous professional development activities virtually, citing convenience and ease of access as important factors that encourage participation despite their busy schedules [72,93]. Nevertheless, psychiatry is not purely a theoretical field in medicine. Psychiatry knowledge must be paired with competency in necessary skills, such as communication skills and psychotherapy skills, to ensure robust and quality training. Learning these skills through an online platform is possible, and a study showed that learning CBT online was not inferior to learning it through in-person training [90]. Nonetheless, proper hands-on guidance remains necessary for mastering these skills, as it might not be possible to fully replicate or demonstrate those subtle, nuanced techniques through a screen. Hence, harmonizing virtual theoretical learning and practical hands-on learning to develop comprehensive blended learning may be a more interesting proposition [82]. Nevertheless, it must be highlighted that the approach of providing knowledge about certain skills and evaluating knowledge levels in trainees is vastly different from teaching skill competencies and then evaluating the levels of competencies of trainees. Because of this issue, some practitioners preferred in-person learning to ensure an optimum level of skill attainment, as compared to through online learning.

As demonstrated by many studies successfully integrating simulation in their medical training, the virtual patient concept (essentially simulating the experience of seeing a patient with the assistance of technology) has been quite helpful in psychiatry education. This concept has a unique strength: the ability to simulate cases that may be uncommon in clinical practice [99]. For example, a study used this concept to portray a refugee with PTSD symptoms, enriching the training of psychiatric trainees and increasing their confidence in managing PTSD cases [48]. Adjacent to the virtual patient concept is the virtual reality (VR) concept. VR allows for an immersive experience, frequently described as “being there,” which involves more senses beyond just sight [100]. It has been used as part of the training curricula in medical fields, including orthopedics [101], surgery [102], and ophthalmology [103], with varying degrees of success. In psychiatric services, VR has been implemented as part of therapy or treatment, for example, exposure therapy for phobic disorders [104] and social skills training in patients with autism spectrum disorder [105]. Unfortunately, VR has not been extensively used in psychiatric education yet, perhaps due to the limitation of the current technology in grasping the complexity of psychiatric cases. As technology rapidly evolves, it remains an exciting avenue to explore in the future.

Competency is often assessed through an examination process. During the pandemic, the transition to online examinations became common worldwide, and psychiatry was no different. Online assessments were applied to various aspects of the psychiatry curriculum. While there were few issues with online theory examinations [86], the same cannot be said for other aspects. For example, an article described the challenges in sitting for the virtual Clinical Assessment of Skills and Competencies (CASC) under the Royal College of Psychiatrists, United Kingdom, where constant worry about internet issues, a sense of disconnect, and an inability to mentally reset between stations affected performance and the overall experience [87].

XSL•FO
RenderX

Another article described the experience of the online Basic Specialist Training examination under the College of Psychiatrists of Ireland [89]. Despite acknowledging the superiority of face-to-face examinations, the online examination was described as nearly as good and more favorable in view of respondents being in a familiar environment as well as saving cost and time to travel [89]. Although it is not possible to compare these examinations directly, the Basic Specialist Training examination highlighted that such examinations could be conducted virtually, but thorough preparation and strong technical support are warranted.

Another interesting aspect of how technology can be valuable in psychiatric training is through the development of cultural competence. Cultural competence refers to the capacity to respond to the unique needs of the population. In the context of psychiatry, it refers to the development of knowledge, skills, and attitude, which can enable the formulation of an intervention that considers the sociocultural backgrounds and sensitivities of psychiatric patients. In turn, this allows for a comprehensive and tailored treatment for patients, especially those from racial and minority ethnic groups. Previous measures to promote cultural competence included learning trips and student exchange programs. However, technology can also offer interesting and possibly cheaper options to achieve the same goals. Trinh et al [80] described an online module program to promote culture sensitivity in a psychiatry department. Three modules were developed, including presentation slides, case vignettes, and recorded videos, covering important topics, such as DSM-5 Outline for Cultural Formulation, Cultural Formulation Interview, and cultural identity as a multidimensional construct. This program was initially met with surprising feedback, with most clinicians indicating that they were not familiar with what questions to ask to elicit a cultural history; however, after completion, the respondents endorsed the module as useful and reported that they would change their practice, suggesting that a brief online module may have potential in this area.

Exploring the application of online-based or technology-assisted learning in psychiatry education for trainees holds significance in the training of future psychiatrists, especially in low- and middle-income countries (LMICs). To put this in context, most African countries have a massive shortage of psychiatrists, with an average of 0.1 per 100,000 people in the 47 countries across the World Health Organization African region [106]. Moreover, both Liberia [107] and Timor-Leste [108] had only 2 trained psychiatrists, indicating the pressing need to support the mental health systems in these countries. While there is no easy fix for this situation, efforts to increase the number of psychiatrists is of utmost importance. In this context, a concerted and collaborative effort among universities or training programs across regions or continents for the training of future psychiatrists in LMICs is needed to alleviate this issue, and the use of online platforms could be the key to bridging the gap.

Limitations

Our review has some limitations. First, we did not limit the types of publications, resulting in variations in the quality and rigor of the studies. Additionally, majority of the articles were from Western countries, with very few from LMICs, which might reduce generalizability. It is important to note that while a healthy number of studies were included in this scoping review, ultimately obvious heterogeneity was present in terms of the outcomes measured. Majority of the studies had a 1-group intervention study design and had a rather small sample size. Therefore, while some of the included studies might have shown positive responses or outcomes, the findings need to be interpreted carefully in the context of these factors, which might affect generalizability. Moreover, the lack of well-designed comparative intervention studies limits the understanding of the effectiveness of online learning in comparison to traditional face-to-face learning. Another key limitation is the profound lack of an objective assessment as part of the outcome measure within the included studies. Most of the studies assessed satisfaction and attitudes toward the interventions, rather than the actual impact of the interventions.

Moving forward, more well-designed studies in psychiatric education for trainees are needed, especially with objective assessments, to truly evaluate the suitability of online-based and technology-assisted learning. There is a clear paucity of studies evaluating the efficacy of psychotherapy skills training delivered virtually, which may be of significance to LMICs that need a higher number of competent mental health professionals. Lastly, the emergence of AI systems, such as ChatGPT and DeepSeek, can be a game changer for the psychiatric education of trainees, and further exploration is required on how to maximize the benefits of these systems while developing safe and competent psychiatrists in the future.

Conclusion

Videoconference-based learning was the most widely implemented approach, followed by online modules and virtual patients. Despite the outcome heterogeneity and small sample sizes in the included studies, the application of such approaches may have utility in terms of knowledge and skills attainment. With further fine-tuning, these approaches could become effective solutions to address the significant deficiency of psychiatrists, especially in LMICs.

Acknowledgments

The authors would like to thank Professor Dr Samy Azer for his input in preparing the manuscript. No funding was received to assist with the preparation of this manuscript.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

XSL•FO
RenderX

Authors' Contributions

MAMK and SMYAS: conceptualization, data collection, data analysis, and writing – original draft preparation. TIMD and JTYL: conceptualization, data analysis, writing – review and editing, and supervision. All authors contributed to the article and approved the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[DOCX File , 14 KB - mededu_v11i1e64773_app1.docx]

Multimedia Appendix 2
Data of the 3 study phases.
[DOCX File , 90 KB - mededu_v11i1e64773_app2.docx]

Multimedia Appendix 3
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 101 KB - mededu_v11i1e64773_app3.pdf]

References

1. Ahmady S, Kallestrup P, Sadoughi MM, Katibeh M, Kalantarion M, Amini M, et al. Distance learning strategies in medical education during COVID-19: A systematic review. J Educ Health Promot 2021;10:421 [FREE Full text] [doi: 10.4103/jehp.jehp_318_21] [Medline: 35071627]
2. Saiyad S, Virk A, Mahajan R, Singh T. Online Teaching in Medical Training: Establishing Good Online Teaching Practices from Cumulative Experience. Int J Appl Basic Med Res 2020;10(3):149-155 [FREE Full text] [doi: 10.4103/ijabmr.IJABMR_358_20] [Medline: 33088735]
3. Brenner AM. Uses and limitations of simulated patients in psychiatric education. Acad Psychiatry 2009 Apr 27;33(2):112-119. [doi: 10.1176/appi.ap.33.2.112] [Medline: 19398623]
4. Rauch C, Utz J, Rauch M, Kornhuber J, Spitzer P. E-learning is not inferior to on-site teaching in a psychiatric examination course. Front Psychiatry 2021 Apr 13;12:624005 [FREE Full text] [doi: 10.3389/fpsyt.2021.624005] [Medline: 33927651]
5. Beketov V, Menshikova I, Khudarova A. Fast track to full online education in the medical field: evaluating effectiveness and identifying problems from the COVID-19 experience. Int J Web-Based Learn Teach Technol 2022;17(1):1-24 [FREE Full text] [doi: 10.4018/IJWLTT.315824]
6. Westphaln KK, Regoeczi W, Masotya M, Vazquez-Westphaln B, Lounsbury K, McDavid L, et al. From Arksey and O'Malley and Beyond: Customizations to enhance a team-based, mixed approach to scoping review methodology. MethodsX 2021;8:101375 [FREE Full text] [doi: 10.1016/j.mex.2021.101375] [Medline: 34430271]
7. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol Theory Pract 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
8. Haddaway NR, Collins AM, Coughlin D, Kirk S. The role of google scholar in evidence reviews and its applicability to grey literature searching. PLoS One 2015 Sep 17;10(9):e0138237 [FREE Full text] [doi: 10.1371/journal.pone.0138237] [Medline: 26379270]
9. Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. JAMA 2007 Sep 05;298(9):1002-1009. [doi: 10.1001/jama.298.9.1002] [Medline: 17785645]
10. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
11. McCloskey K, Neuzil K, Basak R, Chan KH. Quality of reporting for qualitative studies in pediatric urology-A scoping review. J Pediatr Urol 2023 Oct;19(5):643-651. [doi: 10.1016/j.jpurol.2023.04.027] [Medline: 37481426]
12. Powsner S, Byck R. Implementing a computer system for psychiatric training : the electric resident. Acad Psychiatry 1991 Jun;15(2):100-105. [doi: 10.1007/BF03341304] [Medline: 24430518]
13. Briscoe GW, Fore Arcand LG, Lin T, Johnson J, Rai A, Kollins K. Students' and residents' perceptions regarding technology in medical training. Acad Psychiatry 2006 Dec 01;30(6):470-479. [doi: 10.1176/appi.ap.30.6.470] [Medline: 17139018]
14. Walter DA, Rosenquist PB, Bawtinhimer G. Distance learning technologies in the training of psychiatry residents: a critical assessment. Acad Psychiatry 2004 Mar 01;28(1):60-65. [doi: 10.1176/appi.ap.28.1.60] [Medline: 15140810]
15. Greenwood J, Williams R. Continuing professional development for Australian rural psychiatrists by videoconference. Australas Psychiatry 2008 Aug 01;16(4):273-276. [doi: 10.1080/10398560801982994] [Medline: 18608161]

XSL•FO
RenderX

16. Chipps J, Ramlall S, Mars M. Videoconference-based education for psychiatry registrars at the University of KwaZulu-Natal, South Africa. Afr J Psychiatry (Johannesbg) 2012 Jul 19;15(4):248-254. [doi: 10.4314/ajpsy.v15i4.32] [Medline: 22829227]
17. Nasr Esfahani M, Behzadipour M, Jalali Nadoushan A, Shariat SV. A pilot randomized controlled trial on the effectiveness of inclusion of a distant learning component into empathy training. Med J Islam Repub Iran 2014;28:65 [FREE Full text] [Medline: 25405130]
18. Gammon D, Sørlie T, Bergvik S, Sørensen Høifødt T. Psychotherapy supervision conducted via videoconferencing: A qualitative study of users' experiences. Nordic Journal of Psychiatry 2009 Jul 12;52(5):411-421. [doi: 10.1080/08039489850139445]
19. Szeftel R, Hakak R, Meyer S, Naqvi S, Sulman-Smith H, Delrahim K, et al. Training psychiatric residents and fellows in a telepsychiatry clinic: a supervision model. Acad Psychiatry 2008;32(5):393-399. [doi: 10.1176/appi.ap.32.5.393] [Medline: 18945978]
20. Dzara K, Sarver J, Bennett JI, Basnet P. Resident and medical student viewpoints on their participation in a telepsychiatry rotation. Acad Psychiatry 2013 May 01;37(3):214-216. [doi: 10.1176/appi.ap.12050101] [Medline: 23632939]
21. Pignatiello A, Teshima J, Boydell KM, Minden D, Volpe T, Braunberger PG. Child and youth telepsychiatry in rural and remote primary care. Child Adolesc Psychiatr Clin N Am 2011 Jan;20(1):13-28. [doi: 10.1016/j.chc.2010.08.008] [Medline: 21092909]
22. Volpe T, Boydell KM, Pignatiello A. Attracting Child Psychiatrists to a Televideo Consultation Service: The TeleLink Experience. Int J Telemed Appl 2013;2013:146858 [FREE Full text] [doi: 10.1155/2013/146858] [Medline: 23864854]
23. DeGaetano N, Greene C, Dearaujo N, Lindley S. A pilot program in telepsychiatry for residents: initial outcomes and program development. Acad Psychiatry 2015 Feb;39(1):114-118. [doi: 10.1007/s40596-014-0122-y] [Medline: 24777712]
24. Bayar MR, Poyraz BC, Aksoy-Poyraz C, Arikan MK. Reducing mental illness stigma in mental health professionals using a web-based approach. Isr J Psychiatry Relat Sci 2009;46(3):226-230 [FREE Full text] [Medline: 20039525]
25. Rahman A, Nizami A, Minhas A, Niazi R, Slatch M, Minhas F. E-Mental health in Pakistan: a pilot study of training and supervision in child psychiatry using the internet. Psychiatr Bull 2006;30(4):149-152. [doi: 10.1192/pb.30.4.149]
26. Kulier R, Hadley J, Weinbrenner S, Meyerrose B, Decsi T, Horvath AR, et al. Harmonising evidence-based medicine teaching: a study of the outcomes of e-learning in five European countries. BMC Med Educ 2008 Apr 29;8:27 [FREE Full text] [doi: 10.1186/1472-6920-8-27] [Medline: 18442424]
27. DeBonis K, Blair T, Payne S, Wigan K, Kim S. Viability of a web-based module for teaching electrocardiogram reading skills to psychiatry residents: learning outcomes and trainee interest. Acad Psychiatry 2015 Dec;39(6):645-648. [doi: 10.1007/s40596-014-0249-x] [Medline: 25391493]
28. Garside S, Levinson A, Kuziora S, Bay M, Norman G. Efficacy of teaching clinical clerks and residents how to fill out the Form 1 of the Mental Health Act using an e-learning module. Electron J e-Learning 2009;7(3):239-246 [FREE Full text]
29. Kenny P, Parsons T, Pataki C, Pato M, St George C, Sugar J, et al. Virtual Justina: A PTSD virtual patient for clinical classroom training. Annu Rev CyberTherapy Telemed 2008;6(1):113-118 [FREE Full text]
30. Pataki C, Pato MT, Sugar J, Rizzo AS, Parsons TD, St George C, et al. Virtual patients as novel teaching tools in psychiatry. Acad Psychiatry 2012 Sep 01;36(5):398-400. [doi: 10.1176/appi.ap.10080118] [Medline: 22983473]
31. Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR). Washington, DC: American Psychiatric Association; 2000.
32. Gorrindo T, Baer L, Sanders K, Birnbaum RJ, Fromson JA, Sutton-Skinner KM, et al. Web-based simulation in psychiatry residency training: a pilot study. Acad Psychiatry 2011;35(4):232-237. [doi: 10.1176/appi.ap.35.4.232] [Medline: 21804041]
33. Torous J, Franzan J, O'Connor R, Mathew I, Keshavan M, Kitts R, et al. Psychiatry residents' use of educational websites: a pilot survey study. Acad Psychiatry 2015 Dec;39(6):630-633. [doi: 10.1007/s40596-015-0335-8] [Medline: 26077007]
34. Reddy A, Duenas H, Luker A, Thootkur M, Kim JW. Preference of electronic versus paper reading resources among trainees in psychiatry. Acad Psychiatry 2018 Oct 3;42(5):743-745. [doi: 10.1007/s40596-018-0930-6] [Medline: 29725992]
35. Hickey C, McAleer S, Khalili D. E-learning: is it any better than traditional approaches in psychotherapy education? Arch Psychiatry Psychother 2015;17(4):48-52. [doi: 10.12740/APP/60644]
36. Kuhn E, Hugo E. Technology-based blended learning to facilitate psychiatry resident training in prolonged exposure (PE) therapy for PTSD. Acad Psychiatry 2017 Feb 5;41(1):121-124. [doi: 10.1007/s40596-016-0648-2] [Medline: 27921266]
37. Ranjbar N, Ricker M, Villagomez A. The integrative psychiatry curriculum: development of an innovative model. Glob Adv Health Med 2019;8:2164956119847118 [FREE Full text] [doi: 10.1177/2164956119847118] [Medline: 31080697]
38. Law M, Rapoport MJ, Seitz D, Davidson M, Madan R, Wiens A. Evaluation of a national online educational program in geriatric psychiatry. Acad Psychiatry 2016 Dec 25;40(6):923-927 [FREE Full text] [doi: 10.1007/s40596-015-0377-y] [Medline: 26108395]
39. Cooper J, Roig Llesuy J. Catatonia education: needs assessment and brief online intervention. Acad Psychiatry 2017 Jun;41(3):360-363. [doi: 10.1007/s40596-016-0632-x] [Medline: 27837452]
40. Williams JM, Poulsen R, Chaguturu V, Tobia A, Palmeri B. Evaluation of an online residency training in tobacco use disorder. Am J Addict 2019 Jul 16;28(4):277-284. [doi: 10.1111/ajad.12885] [Medline: 30993797]

XSL•FO
RenderX

41. Avery J, Knoepflmacher D, Mauer E, Kast KA, Greiner M, Avery J, et al. Improvement in residents' attitudes toward individuals with substance use disorders following an online training module on stigma. HSS J 2019 Feb 01;15(1):31-36 [FREE Full text] [doi: 10.1007/s11420-018-9643-3] [Medline: 30863230]
42. Kupfer D, Schatzberg A, Dunn L, Schneider A, Moore T, DeRosier M. Career development institute with enhanced mentoring: a revisit. Acad Psychiatry 2016 Jun;40(3):424-428 [FREE Full text] [doi: 10.1007/s40596-015-0362-5] [Medline: 26048460]
43. Hameed Y, Al Taiar H, O'Leary D, Kaynge L. Can Online Distance Learning improve access to learning in conflict zones? The Oxford Psychiatry in Iraq (OxPIQ) Experience. Br J Med Pract 2018;11(2):a1114 [FREE Full text]
44. Puspitasari AJ, Kanter JW, Busch AM, Leonard R, Dunsiger S, Cahill S, et al. A randomized controlled trial of an online, modular, active learning training program for behavioral activation for depression. J Consult Clin Psychol 2017 Aug;85(8):814-825. [doi: 10.1037/ccp0000223] [Medline: 28726481]
45. Teshima J, Hodgins M, Boydell KM, Pignatiello A. Resident evaluation of a required telepsychiatry clinical experience. Acad Psychiatry 2016 Apr;40(2):348-352. [doi: 10.1007/s40596-015-0373-2] [Medline: 26122350]
46. Davidson D, Evans L. Virtual study groups and online Observed Structured Clinical Examinations practices - enabling trainees to enable themselves. Australas Psychiatry 2018 Aug 03;26(4):429-431. [doi: 10.1177/1039856218765886] [Medline: 29609477]
47. Pantziaras I, Fors U, Ekblad S. Innovative training with virtual patients in transcultural psychiatry: the impact on resident psychiatrists' confidence. PLoS One 2015;10(3):e0119754 [FREE Full text] [doi: 10.1371/journal.pone.0119754] [Medline: 25794169]
48. Pantziaras I, Fors U, Ekblad S. Training with virtual patients in transcultural psychiatry: do the learners actually learn? J Med Internet Res 2015 Feb 16;17(2):e46 [FREE Full text] [doi: 10.2196/jmir.3497] [Medline: 25689716]
49. Wilkening GL, Gannon JM, Ross C, Brennan JL, Fabian TJ, Marcsisin MJ, et al. Evaluation of branched-narrative virtual patients for interprofessional education of psychiatry residents. Acad Psychiatry 2017 Feb 14;41(1):71-75. [doi: 10.1007/s40596-016-0531-1] [Medline: 26976401]
50. Adeponle A, Skakum K, Cooke C, Fleisher W. The University of Manitoba Psychiatry Toolkit: Development and evaluation. Acad Psychiatry 2016 Aug;40(4):608-611. [doi: 10.1007/s40596-015-0419-5] [Medline: 26443030]
51. Dirlam C, Vallera V, Nelson KJ, Bass D. A psychiatry resident's perspective of the "virtual preceptor" as an electronic medical record clinical education support tool. Acad Psychiatry 2018 Oct 10;42(5):741-742. [doi: 10.1007/s40596-017-0880-4] [Medline: 29322307]
52. Zhang MW, Ho RC, Sockalingam S. Methodology of development of a Delirium clinical application and initial feasibility results. Technol Health Care 2015;23(4):411-417. [doi: 10.3233/THC-150904] [Medline: 25735309]
53. Walsh A, Peters M, Saralkar R, Chisolm M. Psychiatry Residents Integrating Social Media (PRISM): Using Twitter in graduate medical education. Acad Psychiatry 2019 Jun;43(3):319-323. [doi: 10.1007/s40596-018-1017-0] [Medline: 30635806]
54. Kumar PR, Yee A, Francis B. Impact of alcohol withdrawal training program on knowledge, attitude, and perception among healthcare providers in a hospital setting. J Subst Use 2022;27(1):80-85. [doi: 10.1080/14659891.2021.1897696]
55. Shekunov J, Swintak C, Somers K, Kolla BB, Ruble A, Bhatt-Mackin S, et al. The Virtual Couch: a curriculum on the question of the fundamentals of remote psychotherapy-pilot study. Acad Psychiatry 2024 Feb;48(1):52-56 [FREE Full text] [doi: 10.1007/s40596-023-01805-6] [Medline: 37365485]
56. Tapoi C, de Filippis R, Di Lodovico L, Filip M, Fusar-Poli L, Salas DG, et al. 10th EPA Summer School on Research 2021: sharing experience of the first online edition. Inf Psychiatr 2022;98(6):469-474. [doi: 10.1684/ipe.2022.2443]
57. Noori S, Khasnavis S, DeCroce-Movson E, Blay-Tofey M, Vitiello E. A curriculum on digital psychiatry for a US-based psychiatry residency training program: pilot implementation study. JMIR Form Res 2024 May 13;8:e41573 [FREE Full text] [doi: 10.2196/41573] [Medline: 38739423]
58. Kiing J, Feldman H, Ladish C, Srinivasan R, Donnelly CL, Chong SC, et al. International Interprofessional Collaborative Office Rounds (iiCOR): Addressing children's developmental, behavioral, and emotional health using distance technology. Front Public Health 2021;9:657780 [FREE Full text] [doi: 10.3389/fpubh.2021.657780] [Medline: 34055722]
59. Ouanes S, Larnaout A, Jouini L. Use of modern technology in psychiatry training in a middle-income country. Asia Pac Psychiatry 2021 Dec 07;13(4):e12496. [doi: 10.1111/appy.12496] [Medline: 34873857]
60. de Cates AN, Mullin D, Stirland L, Pinto da Costa M, Tracy D. Breaking down barriers: promoting journals beyond the page with open access journal clubs. BJPsych Bull 2024 Apr 01;49(1):1-4 [FREE Full text] [doi: 10.1192/bjb.2024.3] [Medline: 38557559]
61. Blamey H, Harrison C, Roddick A, Malhotra T, Saunders K. Simulated virtual on-call training programme for improving non-specialised junior doctors' confidence in out-of-hours psychiatry: quantitative assessment. BJPsych Bull 2023 Oct;47(5):287-295 [FREE Full text] [doi: 10.1192/bjb.2022.40] [Medline: 36073524]
62. Li L, Ray JM, Bathgate M, Kulp W, Cron J, Huot SJ, et al. Implementation of simulation-based health systems science modules for resident physicians. BMC Med Educ 2022 Jul 30;22(1):584 [FREE Full text] [doi: 10.1186/s12909-022-03627-w] [Medline: 35906583]

XSL•FO
RenderX

63. Collin G, Turner O, Luthra V. Evaluation of doctors' experience of psychodynamic psychotherapy training at Leeds and York Partnership NHS Foundation Trust during the COVID-19 pandemic. Br J Psychother 2023 Jun 05;39(3):425-447. [doi: 10.1111/bjp.12833]
64. Shanley I, Jones C, Reddi N. Medical psychotherapy training and the COVID-19 pandemic. Br J Psychother 2022 May;38(2):338-352 [FREE Full text] [doi: 10.1111/bjp.12719] [Medline: 35601048]
65. Famina S, Farooqui AA, Caudill RL. Early use of telepsychotherapy in resident continuity clinics-our experience and a review of literature. Mhealth 2020 Jan;6:1-1 [FREE Full text] [doi: 10.21037/mhealth.2019.09.11] [Medline: 32190612]
66. Beran C, Sowa NA. Adaptation of an academic inpatient consultation-liaison psychiatry service during the SARS-CoV-2 pandemic: effects on clinical practice and trainee supervision. J Acad Consult Liaison Psychiatry 2021 Mar;62(2):186-192 [FREE Full text] [doi: 10.1016/j.psym.2020.11.002] [Medline: 33288272]
67. Cruz C, Orchard K, Shoemaker EZ, Hilty DM. A survey of residents/fellows, program directors, and faculty about telepsychiatry: clinical experience, interest, and views/concerns. J Technol Behav Sci 2021 Feb 09;6(2):327-337 [FREE Full text] [doi: 10.1007/s41347-020-00164-5] [Medline: 33585672]
68. McCutcheon S. Putting the cart before the horse: outcomes following rapid implementation of telepsychiatry in an outpatient resident clinic. Acad Psychiatry 2020 Dec;44(6):655-658 [FREE Full text] [doi: 10.1007/s40596-020-01316-8] [Medline: 32944873]
69. Nadeem T, Asad N, Hamid SN, Gul B, Aftab R. Experiences from implementing an Lessons from teaching psychiatry trainees at a tertiary care hospital in Karachi, Pakistan. Asian J Psychiatr 2021 Nov;65:102865. [doi: 10.1016/j.ajp.2021.102865] [Medline: 34560566]
70. Knez R, El Alaoui S, Ivarson J, Risö Bergerlind LL, Stasinakis S, Ahlgren AM, et al. Medical residents' and teachers' perceptions of the digital format of nation-wide didactic courses for psychiatry residents in Sweden: a survey-based observational study. BMC Med Educ 2023 Jan 05;23(1):9 [FREE Full text] [doi: 10.1186/s12909-022-03989-1] [Medline: 36604728]
71. Nalan P, Manning A. The juice is worth the squeeze: psychiatry residents' experience of Balint group. Int J Psychiatry Med 2022 Nov;57(6):508-520 [FREE Full text] [doi: 10.1177/00912174221127084] [Medline: 36112941]
72. Elzain M, Murthy S, Omer S, McCarthy G. Reflective practice in psychiatric training: Balint groups during COVID-19. Ir J Psychol Med 2023 Sep 15;40(3):326-329. [doi: 10.1017/ipm.2022.51] [Medline: 36519310]
73. Ranjbar N, Erb M, Tomkins J, Taneja K, Villagomez A. Implementing a mind-body skills group in psychiatric residency training. Acad Psychiatry 2022 Aug 02;46(4):460-465 [FREE Full text] [doi: 10.1007/s40596-021-01507-x] [Medline: 34341965]
74. Chacko E, Vara A, Cheung G, Naskar C, Ramalho R, Bell R. A mindfulness-based cognitive therapy informed virtual psychiatry trainee wellbeing programme: Development and preliminary feedback. Australas Psychiatry 2022 Oct;30(5):663-667. [doi: 10.1177/10398562221119090] [Medline: 35973679]
75. Westcott S, Simms K, van Kampen K, Jafine H, Chan T. Off-script, online: virtual medical improv pilot program for enhancing well-being and clinical skills among psychiatry residents. Acad Psychiatry 2023 Aug;47(4):374-379 [FREE Full text] [doi: 10.1007/s40596-023-01778-6] [Medline: 37101105]
76. Wasser T, Hu J, Danzig A, Yarnell-MacGrory S, Guzman J, Michaelsen K. Teaching forensic concepts to residents using interactive online modules. J Am Acad Psychiatry Law 2020 Mar;48(1):77-83. [doi: 10.29158/JAAPL.003890-20] [Medline: 31753964]
77. Wortzel JR, Maeng DD, Francis A, Oldham MA. Evaluating the effectiveness of an educational module for the Bush-Francis Catatonia Rating Scale. Acad Psychiatry 2022 Apr 07;46(2):185-193. [doi: 10.1007/s40596-021-01582-0] [Medline: 34997564]
78. Jacoby N, Gullick M, Sullivan N, Shalev D. Development and evaluation of an innovative neurology e-learning didactic curriculum for psychiatry residents. Acad Psychiatry 2023 Jun 14;47(3):237-244 [FREE Full text] [doi: 10.1007/s40596-023-01769-7] [Medline: 36918470]
79. Williams JM, Steinberg ML, Wang H, Chaguturu V, Poulsen R, Tobia A, et al. Practice change after training psychiatry residents in tobacco use disorder. Psychiatr Serv 2020 Feb 01;71(2):209-212 [FREE Full text] [doi: 10.1176/appi.ps.201900272] [Medline: 31690223]
80. Trinh N, O'Hair C, Agrawal S, Dean T, Emmerich A, Rubin D, et al. Lessons learned: developing an online training program for cultural sensitivity in an academic psychiatry department. Psychiatr Serv 2021 Oct 01;72(10):1233-1236. [doi: 10.1176/appi.ps.202000015] [Medline: 34106742]
81. Brown TR, Amir H, Hirsch D, Jansen MO. Designing a novel digitally delivered antiracism intervention for mental health clinicians: exploratory analysis of acceptability. JMIR Hum Factors 2024 Apr 03;11:e52561 [FREE Full text] [doi: 10.2196/52561] [Medline: 38568730]
82. Owais S, Saperson K, Levinson AJ, Payne S, Lamont R, Brown MV, et al. Evaluation of the online component of a blended learning electroconvulsive therapy curriculum for psychiatry residents to treat depression in older adults. Acad Psychiatry 2024 Feb 26;48(1):36-40. [doi: 10.1007/s40596-023-01825-2] [Medline: 37493958]

XSL•FO
RenderX

83. Gargot T, Arnaoutoglou NA, Costa T, Sidorova O, Liu-Thwaites N, Moorey S, et al. Can we really teach cognitive behavioral therapy with a massive open online course? Eur Psychiatry 2020 Mar 10;63(1):e38 [FREE Full text] [doi: 10.1192/j.eurpsy.2020.29] [Medline: 32151289]
84. Gratzer D, Islam F, Sockalingam S, Beckett R. Reading of the Week: a continuing professional development program for psychiatrists and residents that Osler would have liked. Can Med Educ J 2022 Mar;13(1):81-85 [FREE Full text] [doi: 10.36834/cmej.72089] [Medline: 35291453]
85. Moore D, Green J, Gallis HA. Achieving desired results and improved outcomes: integrating planning and assessment throughout learning activities. J Contin Educ Health Prof 2009;29(1):1-15. [doi: 10.1002/chp.20001] [Medline: 19288562]
86. Scheeres K, Agrawal N, Ewen S, Hall I. Transforming MRCPsych theory examinations: digitisation and very short answer questions (VSAQs). BJPsych Bull 2022 Feb;46(1):52-56 [FREE Full text] [doi: 10.1192/bjb.2021.23] [Medline: 33752773]
87. Chu K, Sathanandan S. The virtual Clinical Assessment of Skills and Competence: the impact and challenges of a digitised final examination. BJPsych Bull 2023 Apr;47(2):110-115 [FREE Full text] [doi: 10.1192/bjb.2021.112] [Medline: 34937596]
88. Gentry MT, Murray AP, Altchuler SI, McKean AJ, Joyce JB, Hilty DM. Development and implementation of a virtual clinical skills examination in general psychiatry. Acad Psychiatry 2023 Feb 02;47(1):48-52 [FREE Full text] [doi: 10.1007/s40596-022-01691-4] [Medline: 35918600]
89. Usman M, Adamis D, McCarthy G. Perspectives of psychiatric trainees and examiners on the assessment of communication skills during an online clinical examination: a qualitative study. Ir J Psychol Med 2023 Jun 15;41(4):1-7. [doi: 10.1017/ipm.2023.19] [Medline: 37318020]
90. Soll D, Fuchs R, Mehl S. Teaching cognitive behavior therapy to postgraduate health care professionals in times of COVID 19 - An asynchronous blended learning environment proved to be non-inferior to in-person training. Front Psychol 2021 Sep 27;12:657234 [FREE Full text] [doi: 10.3389/fpsyg.2021.657234] [Medline: 34646190]
91. Hewson T, Foster H, Sanderson R. Using socially distanced and online simulation training to improve the confidence of junior doctors in psychiatry. BJPsych Bull 2023 Aug 18;47(4):235-241 [FREE Full text] [doi: 10.1192/bjb.2022.18] [Medline: 35300744]
92. Rakofsky JJ, Talbot TB, Dunlop BW. A virtual standardized patient-based assessment tool to evaluate psychiatric residents' psychopharmacology proficiency. Acad Psychiatry 2020 Dec;44(6):693-700 [FREE Full text] [doi: 10.1007/s40596-020-01286-x] [Medline: 32681418]
93. Ben Ammer A, Bryan J, Asghar-Ali A. The impact of COVID-19 in reshaping graduate medical education: harnessing hybrid learning and virtual training. Cureus 2024 Mar;16(3):e56790 [FREE Full text] [doi: 10.7759/cureus.56790] [Medline: 38650783]
94. Kalayasiri R, Wainipitapong S. Training of psychiatry and mental health in a low- and middle-income country: Experience from Thailand before and after COVID-19 outbreak. Asia Pac Psychiatry 2021 Dec 07;13(4):e12493 [FREE Full text] [doi: 10.1111/appy.12493] [Medline: 34873871]
95. Heldt JP, Agrawal A, Loeb R, Richards MC, Castillo EG, DeBonis K. We're not sure we like it but we still want more: trainee and faculty perceptions of remote learning during the COVID-19 pandemic. Acad Psychiatry 2021 Oct 17;45(5):598-602 [FREE Full text] [doi: 10.1007/s40596-021-01403-4] [Medline: 33594628]
96. Hodges C, Moore S, Lockee B, Trust T, Bond M. The Difference Between Emergency Remote Teaching and Online Learning. EDUCAUSE. 2020 Mar 27. URL: https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning [accessed 2025-01-03]
97. Adipat S. Why web-conferencing matters: rescuing education in the time of COVID-19 pandemic crisis. Front Educ 2021 Sep 22;6:752522. [doi: 10.3389/feduc.2021.752522]
98. Sterling J. 10 Ways Video Conferencing Changed the Education Industry for Good. Usherwood Office Technology. 2024 Mar 18. URL: https://www.usherwood.com/blog/10-ways-video-conferencing-changed-the-education-industry-for-good [accessed 2025-01-07]
99. Ayaz O, Ismail F. Healthcare Simulation: A Key to the Future of Medical Education - A Review. Adv Med Educ Pract 2022;13:301-308 [FREE Full text] [doi: 10.2147/AMEP.S353777] [Medline: 35411198]
100. Slater M, Banakou D, Beacco A, Gallego J, Macia-Varela F, Oliva R. A separate reality: an update on place illusion and plausibility in virtual reality. Front Virtual Real 2022 Jun 27;3:914392. [doi: 10.3389/frvir.2022.914392]
101. Combalia A, Sanchez-Vives MV, Donegan T. Immersive virtual reality in orthopaedics-a narrative review. Int Orthop 2024 Jan 11;48(1):21-30 [FREE Full text] [doi: 10.1007/s00264-023-05911-w] [Medline: 37566225]
102. Desselle MR, Brown RA, James AR, Midwinter MJ, Powell SK, Woodruff MA. Augmented and virtual reality in surgery. Comput Sci Eng 2020 Feb 11;22(3):18-26. [doi: 10.1109/mcse.2020.2972822]
103. Ma M, Saha C, Poon S, Yiu R, Shih K, Chan Y. Virtual reality and augmented reality- emerging screening and diagnostic techniques in ophthalmology: A systematic review. Surv Ophthalmol 2022;67(5):1516-1530. [doi: 10.1016/j.survophthal.2022.02.001] [Medline: 35181279]
104. van Loenen I, Scholten W, Muntingh A, Smit J, Batelaan N. The effectiveness of virtual reality exposure-based cognitive behavioral therapy for severe anxiety disorders, obsessive-compulsive disorder, and posttraumatic stress disorder: meta-analysis. J Med Internet Res 2022 Feb 10;24(2):e26736. [doi: 10.2196/26736] [Medline: 35142632]

XSL•FO
RenderX

105. Ke F, Moon J, Sokolikj Z. Virtual reality–based social skills training for children with autism spectrum disorder. J Spec Educ Technol 2020 Sep 15;37(1):49-62. [doi: 10.1177/0162643420945603]
106. Mental health a human right, but only 1 psychiatrist per 1,000,000 people in sub-Saharan Africa – UNICEF/WHO. UNICEF. 2023 Oct 10. URL: https://www.unicef.org/esa/press-releases/mental-health-a-human-right [accessed 2024-07-01]
107. Ghebrehiwet S, Ogundare T, Owusu M, Harris BL, Ojediran B, Touma M, et al. Building a postgraduate psychiatry training program in Liberia through cross-country collaborations: initiation stages, challenges, and opportunities. Front Public Health 2023;11:1020723 [FREE Full text] [doi: 10.3389/fpubh.2023.1020723] [Medline: 37727607]
108. Mathur A. Let's work to recognize mental health as a universal human right. World Health Organization. 2023 Oct 10. URL: https://www.who.int/timorleste/news/detail/10-10-2023-let-s-work-to-recognize-mental-health-as-a-universal-human-right [accessed 2024-07-01]

Abbreviations

AI: artificial intelligence
CBT: cognitive behavioral therapy
DSM: Diagnostic and Statistical Manual of Mental Disorders
LMIC: low- and middle-income country
MOOC: massive open online course
OSCE: objective structured clinical examination
PTSD: posttraumatic stress disorder
RANZCP: Royal Australian and New Zealand College of Psychiatrist
SRQR: Standards for Reporting Qualitative Research
VR: virtual reality

Edited by B Lesselroth; submitted 26.07.24; peer-reviewed by D Hilty, K Drude; comments to author 30.09.24; revised version received 31.01.25; accepted 25.02.25; published 15.04.25.

Please cite as:
Mohd Kassim MA, Azli Shah SMY, Lim JTY, Mohd Daud TI
Online-Based and Technology-Assisted Psychiatric Education for Trainees: Scoping Review
JMIR Med Educ 2025;11:e64773
URL: https://mededu.jmir.org/2025/1/e64773
doi:10.2196/64773
PMID:

©Mohd Amiruddin Mohd Kassim, Sidi Muhammad Yusoff Azli Shah, Jane Tze Yn Lim, Tuti Iryani Mohd Daud. Originally published in JMIR Medical Education (https://mededu.jmir.org), 15.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Leveraging Generative Artificial Intelligence to Improve Motivation and Retrieval in Higher Education Learners

Noahlana Monzon, BS; Franklin Alan Hays, BS, PhD

Department of Nutritional Sciences, University of Oklahoma Health Sciences, 1200 N Stonewall Ave, 3064 Allied Health Building, Oklahoma City, OK, United States

Corresponding Author:
Franklin Alan Hays, BS, PhD
Department of Nutritional Sciences, University of Oklahoma Health Sciences, 1200 N Stonewall Ave, 3064 Allied Health Building, Oklahoma City, OK, United States

Abstract

Generative artificial intelligence (GenAI) presents novel approaches to enhance motivation, curriculum structure and development, and learning and retrieval processes for both learners and instructors. Though a focus for this emerging technology is academic misconduct, we sought to leverage GenAI in curriculum structure to facilitate educational outcomes. For instructors, GenAI offers new opportunities in course design and management while reducing time requirements to evaluate outcomes and personalizing learner feedback. These include innovative instructional designs such as flipped classrooms and gamification, enriching teaching methodologies with focused and interactive approaches, and team-based exercise development among others. For learners, GenAI offers unprecedented self-directed learning opportunities, improved cognitive engagement, and effective retrieval practices, leading to enhanced autonomy, motivation, and knowledge retention. Though empowering, this evolving landscape has integration challenges and ethical considerations, including accuracy, technological evolution, loss of learner's voice, and socioeconomic disparities. Our experience demonstrates that the responsible application of GenAI's in educational settings will revolutionize learning practices, making education more accessible and tailored, producing positive motivational outcomes for both learners and instructors. Thus, we argue that leveraging GenAI in educational settings will improve outcomes with implications extending from primary through higher and continuing education paradigms.

(JMIR Med Educ 2025;11:e59210) doi:10.2196/59210

KEYWORDS

educational technology; retrieval practice; flipped classroom; cognitive engagement; personalized learning; generative artificial intelligence; higher education; university education; learners; instructors; curriculum structure; learning; technologies; innovation; academic misconduct; gamification; self-directed; socio-economic disparities; interactive approach; medical education; chatGPT; machine learning; AI; large language models

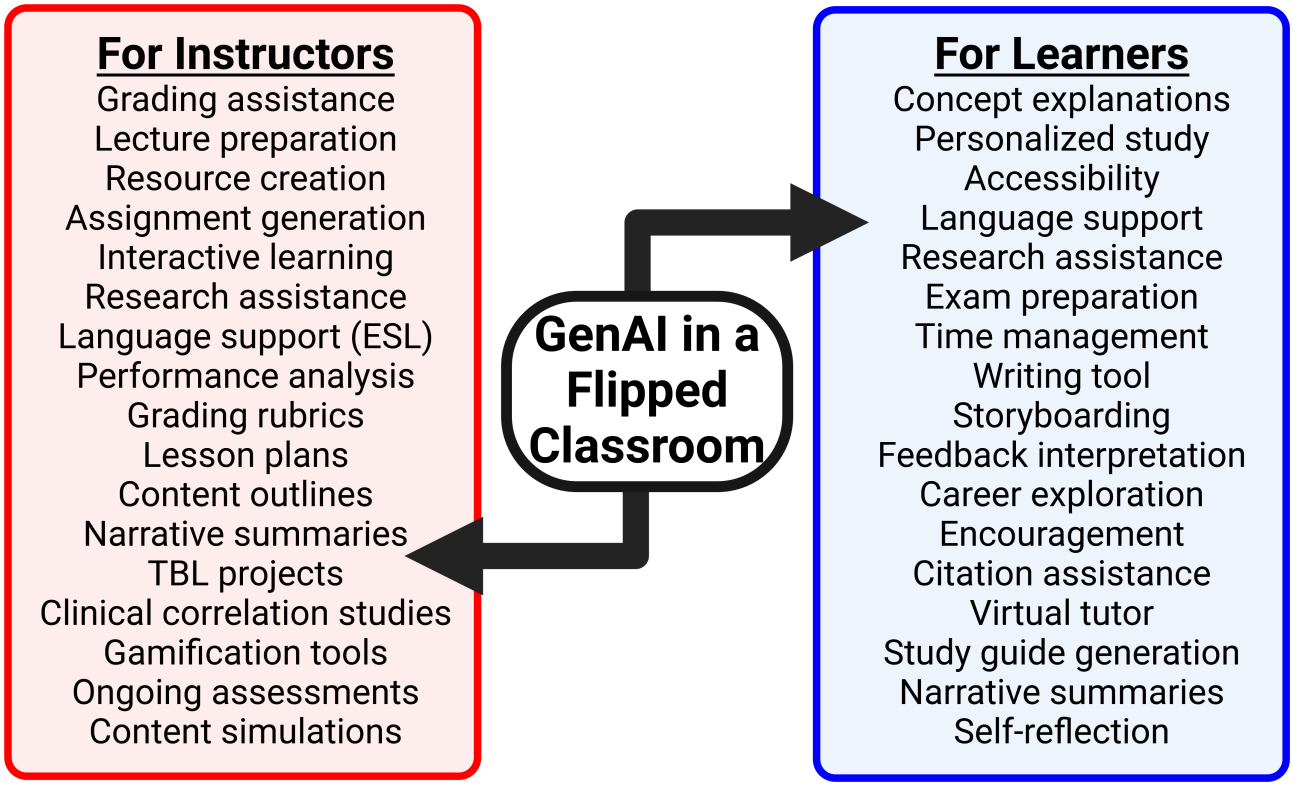
Introduction

Generative artificial intelligence (GenAI) is impacting educational spaces in ways that few technologies have since the personal computer and calculator [1-3]. Though GenAI is not a new concept, its inroads into education and pedagogy started in earnest following the release of "ChatGPT" (November 30, 2022). We observed learners using ChatGPT within weeks of its release. GenAI continues to rapidly evolve with new "GPTs," models, websites, application programming interfaces, and GenAI-enabled hardware [3-6]. GenAI is now "mainstream" with low activation barriers for use. This new reality is sending shockwaves through educational institutions and districts, including higher and clinical education. Learners, instructors, and administration work to understand and define implications while either leveraging or obstructing GenAI implementation. Indeed, banning and blocking GenAI use in some educational settings remains with no clear consensus on what role GenAI can, or should, play going forward. With this rapidly evolving space, it's challenging to differentiate inflated expectations and hype from productivity and enlightenment, to borrow from the Gartner Hype Cycle. One could argue that the GenAI "peak of inflated expectations" has yet to be reached. However, the new reality in clinical and higher education is one where GenAI will play a role going forward, both within classrooms and clinical practice [3,6-8]. What that looks like remains variable and will change depending on the knowledge of those involved, personal perspectives on GenAI, learner and programmatic needs, and accreditation standards and expectations. Our immediate approach was to embrace GenAI, like ChatGPT and other tools as they have come online (eg, Claude, Bard, and CoPilot), as a new tool to facilitate learning, retrieval, and motivation in a higher education, clinically focused, instructional environment. The rationale is that modern GenAI can generate diverse outputs (eg, text, images, videos, and language) derived from data-centric training sets [9] using narrative style prompts or data inputs (eg, content outlines, documents, PDFs, and images). Thus, there is a pragmatic reality that new GenAI tools have a

XSL•FO
RenderX

low activation barrier for use while being capable of generating high-quality output focused on user needs. It's evident that modern GenAIs ability to generate extensive, coherent, responses is fundamental to increasing engagement, communication, and motivation in educational settings. Though not seamless, especially when considering potential for malfeasance or hallucinations [10], GenAI can integrate throughout curricula to reduce overhead and improve outcomes (Figure 1). This integration fosters environments that inspire and empower learners, promote motivation and collaboration, and facilitates the creation of dynamic and individualized curricula.

Figure 1. Generative artificial intelligence application in a flipped classroom model enhances both learner (blue) and instructor (red) experiences. Shown are examples of generative artificial intelligence benefit and impact within respective domains. Bidirectional arrows indicate reciprocal enhancement of generative artificial intelligence applications, demonstrating improvements in instructor-driven activities inherently enrich learner experience, thereby reinforcing a flipped classroom learning environment. GenAI: generative artificial intelligence.



The objective of this viewpoint is to advance a positive perspective on leveraging GenAI tools in modern medical education environments while presenting examples and methods tested in our hands since ChatGPT's initial release. This viewpoint is presented from both learner (Mrs Monzon) and instructor (Dr Hays) perspectives as, in our experience, both offer unique opportunities on GenAI use. Learners are focused on knowledge acquisition and retrieval from an individualized perspective. Not all learners have the same motivation, hidden curriculum, previous knowledge, and experience, or ability to learn and retain learning objectives defined by instructors. Likewise, instructors are unable to individualize curricula across multiple learners or sections while ensuring productive exposure to core learning objectives defined by accreditation and program standards. It's a conundrum of modern higher education, learners seeking individualized instruction amidst information overload while instructors are bandwidth-limited and hamstrung by program and accreditation demands. GenAI tools will directly impact this reality in a positive manner and empower both learners and instructors. The current challenge is what does that look like? How can GenAI be integrated into learning environments to facilitate learning and retrieval, drive motivation, and improve outcomes while avoiding pitfalls such as loss of voice, data ownership and use, academic misconduct or malfeasance, and incorrect information? This future must balance innovation and GenAI integration with established guidelines, integrity and safety guardrails, and equity. By presenting a nuanced perspective of the interplay between GenAI and learning theories from both learner and instructor perspectives, this viewpoint intends to inform GenAI integration that is inclusive, forward-thinking, and collaborative while not ignoring tangible GenAI benefits for all stakeholders in the learning ecosystem. Integration should not overshadow essential human elements of teaching and learning but rather complement and enhance both, thereby creating a dynamic and inclusive educational environment that is responsive. Finally, we argue that GenAI should not be ignored but embraced. It's imperative that learners are exposed to new technologies that will increasingly impact workforce dynamics going forward. Instructors are innovators and our learners are digital natives surrounded by AI technologies. We implemented and evolved methods described in this viewpoint within graduate (PhD and MS), clinical (dietetics and RD), and undergraduate curricula. Leveraging GenAI in courses does require initial effort, yet subsequent improvements in effort needed, instructional quality, and learner feedback justify the initial cost. GenAI has proven, in our hands, to positively impact every pedagogical niche. It should be noted that we acknowledge significant ethical

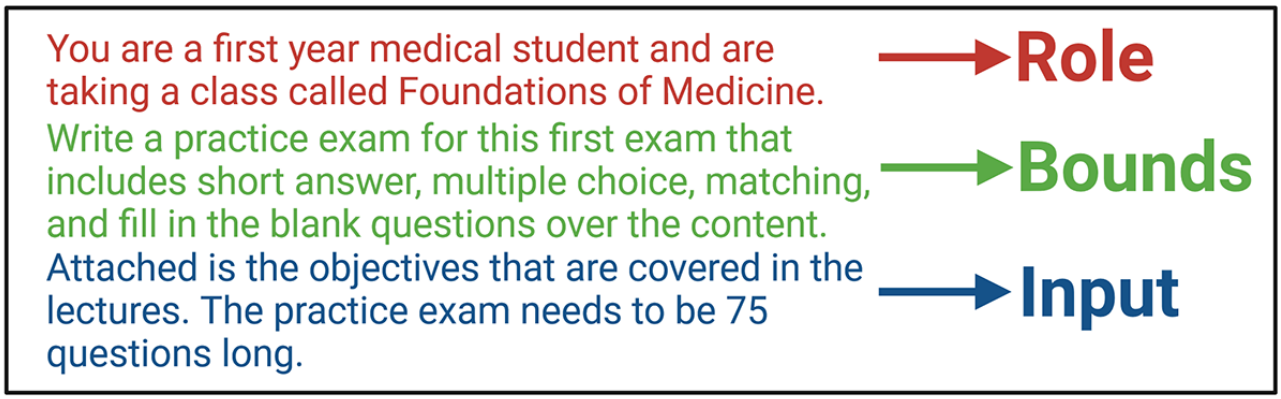
XSL•FO
RenderX

concerns regarding GenAI use in educational settings. This has been covered extensively elsewhere [10,11] and the current viewpoint starts with the perspective that GenAI can, and should, play a role in educational settings.

Learner Perspective

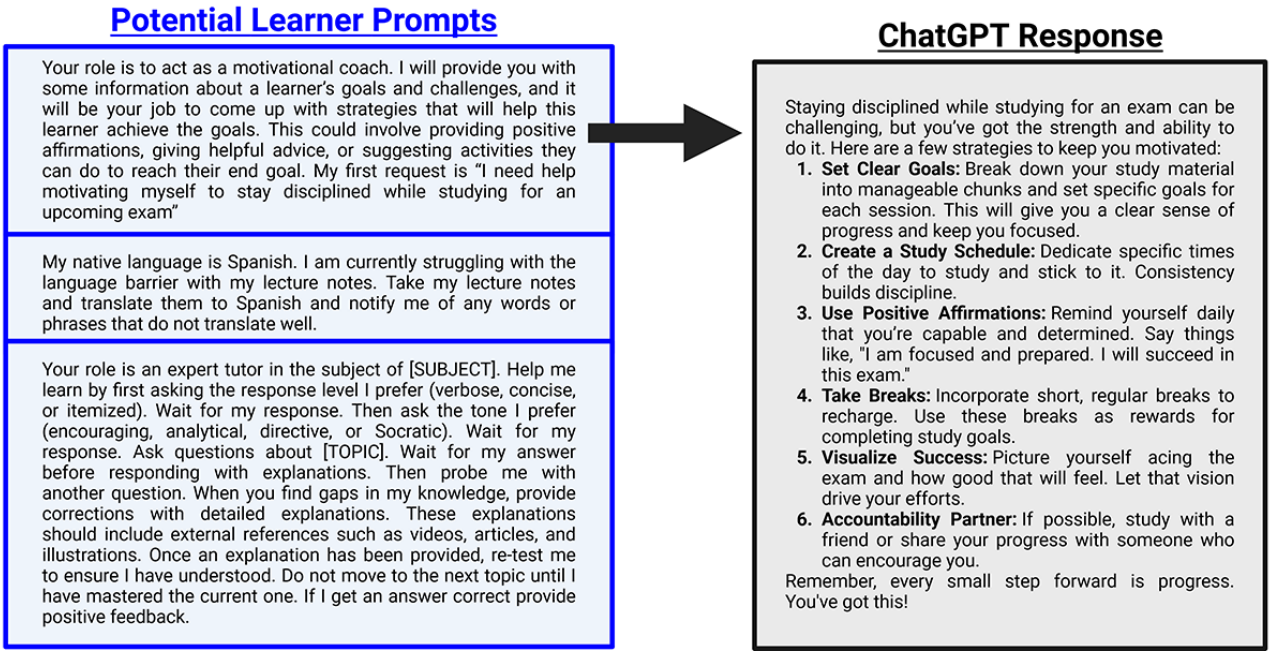
Technology is a powerful means to facilitate collaboration between learners and instructors. Learning management systems (eg, Canvas or D2L) are an example of this point, leveraging technology to facilitate learning and retention. In this sense, bringing GenAI into classroom settings is an evolutionary step with clear emerging data that it enhances learner engagement, motivation, and personalized learning in a self-directed manner. The pragmatic meaning is that interactive collaboration can be extended from instructor-learner or learner-learner to include learner-GenAI where the scope and implementation of learner-GenAI interactions is defined by tools being used, prompt design (Figure 2), and personalized needs (Figures 1 and 3). This approach fosters learner motivation as a key driver for positive outcomes [12]. In addition, learning is more effective when it's relevant, engaging, and contextualized to real-life scenarios (eg, team-based learning or clinical applications) in accordance with adult learning theory. Cognitive load is reached when germane, intrinsic, and extraneous factors become unmanageable [13]. Incorporating GenAI into the educational framework can simplify the intrinsic load, reduce the extraneous load and in turn maximize the germane load. This is consistent with our observations using GenAI to foster collaborative interactions in clinical courses. To maintain learner motivation, one must account for both intrinsic and extrinsic factors [13]. Intrinsic factors include self-efficacy, self-determination, curiosity, cognitive engagement, emotional well-being, professional well-being, and innate interest in the material presented. Extrinsic factors include pedagogical approach, peer interactions, assessment methods, learning environment, curriculum design, and quality and scope of feedback (both peers and instructors). A unique aspect to the learner-GenAI interaction is that it impacts both intrinsic and extrinsic motivational elements for learners. For instance, GenAI can be implemented as a personal tutor or study partner that encourages conversations and positive feedback in a low activation barrier environment (eg, compared with instructor office hours). Engaging GenAI "chatbots" like ChatGPT can also be conversational for learners, like interacting with a human counterpart (see "Current Limitations and Future Hurdles" section below). Thus, leveraging the learner-GenAI interaction provides agency to learners which increases autonomy and motivation [14].

Figure 2. Prompt design. General overview for developing prompts with clearly defined role (red), bounds (green), and input (blue). Each component is essential for effective generative artificial intelligence use toward minimizing refinement while helping ensure output is optimized toward specific needs.



XSL•FO
RenderX

Figure 3. Example learner prompts for leveraging generatice artificial intelligence. Three example learner input prompts are shown (blue box) with representative ChatGPT (accessed on December 10, 2024) output (black box) for the top prompt (black arrow). In these instances, the resultant output can change significantly with minor changes in the input prompt provided. This is also true when using the same input prompt across different generatice artificial intelligence tools. Thus, specificity and clear instructions are key to effect desired output.



Beyond motivation, retrieval practice is an essential component for learner ownership over new material and, for example, applying knowledge in a clinical setting where integration and rapid access is often required [13]. Indeed, learned information is rapidly forgotten without reinforcement [15]. This is a core consideration of the "Desirable Difficulties" theory by Bjork and Bjork [15] suggesting that one should introduce challenges (eg, spacing or testing effects and varied practice) to enhance long-term memory retention of new information. Thus, retrieval practice requires active effort by learners to bolster information recall. This active engagement promotes deeper processing and understanding to facilitate ownership. A common example of retrieval practice in medical learner training is leveraging the Socratic method in clinical rounds, case discussions, simulations, journal clubs, team-based learning, and even mortality and morbidity conferences. In these scenarios, clinicians are pushed to understand, integrate, and verbalize knowledge under immediate critique and assessment. This moves beyond simple passive recall or reading to test true understanding and identify areas where learners assume ownership of knowledge but fail accurate retrieval or application [16]. In simplistic terms, retrieval practice is a common element to most curricula through formative (common in direct clinical training) and summative (common in formalized classroom instruction) assessments. Incorporating learner-GenAI methods into the curriculum provides a dynamic, ongoing, personalized, and iterative method to facilitate retrieval practice for learners outside of formal, instructor-based, course design. The learner-GenAI axis is instructor-independent in this instance. GenAI can generate adaptive quizzes and assessments while customizing difficulty level and content based on learner proficiency (eg, Figure 3). As learners progress and improve in retrieval practice, GenAI can dynamically adjust question complexity, ensuring continued adaptive learning. These tools analyze users learning patterns, preferences, performance data, and needs to personalize content. In this instance, GenAI recommends specific retrieval or practice exercises and intervals to drive memory consolidation. Learner-GenAI natural language interactions can efficiently manage spaced repetition schedules based upon individual learning patterns and needs to adapt timing and frequency of review sessions, ensuring learners revisit information at optimal intervals for memory retention. Finally, it's important for instructors to consider that learners do not enter courses on equitable footing in knowing how to access, use, and leverage GenAI tools. Initial training with pragmatic examples, discussion of prompt engineering, setting up accounts if needed, and reviewing available tools and associated strengths and weaknesses is strongly advised for courses that allow GenAI use.

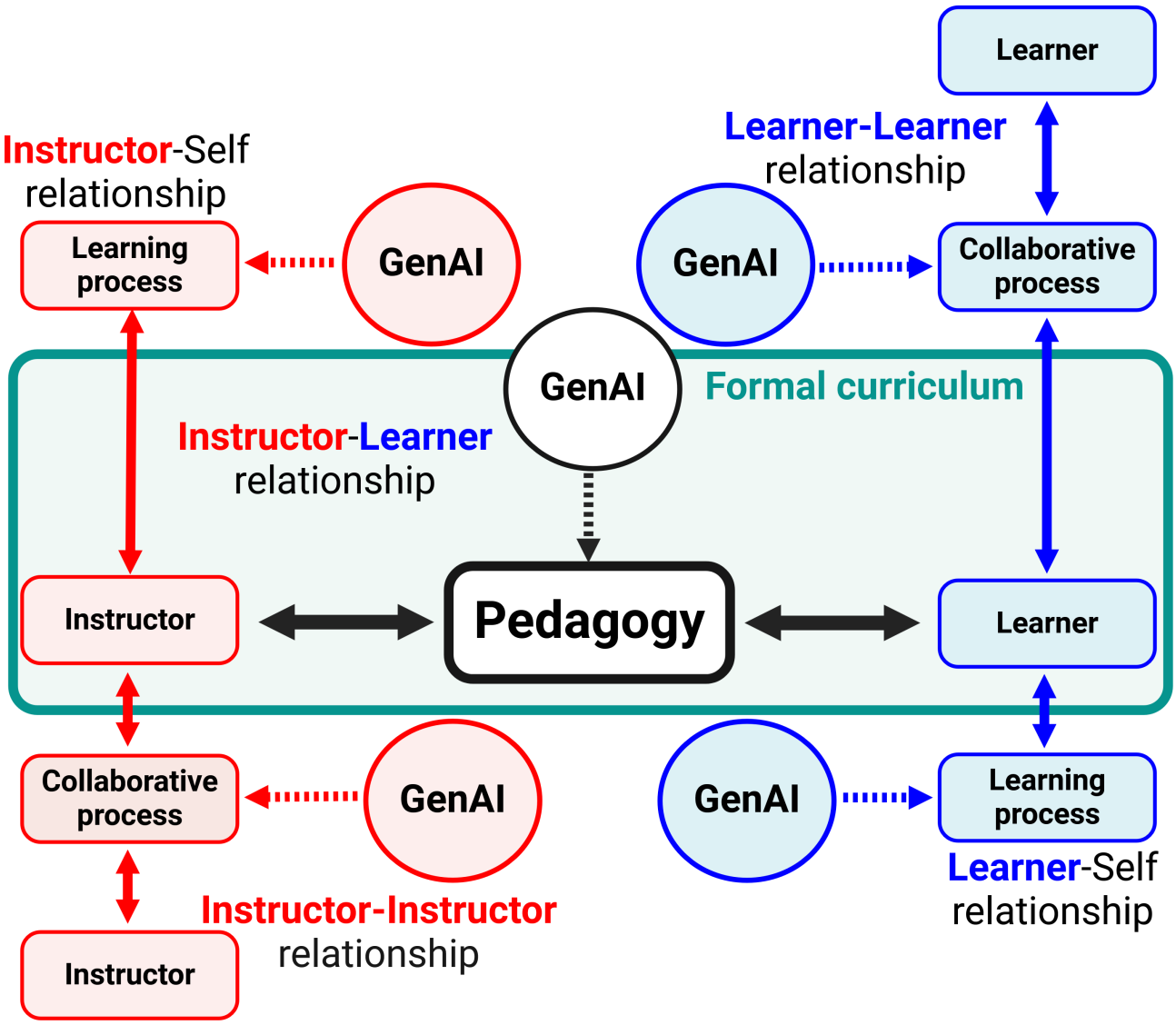
Instructor Perspective

Instructors, faculty, and programs within higher education and clinical training settings are primary determinants of motivational factors for learners [12-14]. These include accreditation and departmental oversight (meaning "static" curriculum), designing and structuring assessment, determining feedback mechanisms, managing learning environment, defining expectations, and even implementing recognition and reward systems for positive outcomes or performance. Thus, in our experience, learner motivation is impacted in significant ways before the first day of class. With this backdrop in place, what role can GenAI play in higher education and clinical training from the instructors' perspective? This question has interesting parallels to the "great calculator debate." These include questions of access and equity, learners using these tools outside class regardless of policies set by schools and instructors, learners not gaining essential skills, or knowledge, through their use, modifications required for existing curricula with a major

XSL•FO
RenderX

shift away from algorithms and "rules" toward meaning, concepts, and applications, and ability to trust accuracy of answers produced from novel technologies. Yet, even with the rapidly evolving current landscape surrounding GenAI, we posit that GenAI has significant benefits for instructors. Thus, the instructors' role in harnessing GenAI as an educational tool is multifaceted and includes instructional design, creating dynamic learning content, and even streamlining administrative tasks (Figures 1 and 4), all of which are predicated on training and learning about a rapidly evolving field with new tools appearing almost daily.

Figure 4. Role of generative artificial intelligence in educational relationships and processes. Generative artificial intelligence intersects with and supports relational dynamics between instructors (red) and learners (blue). Four primary interactions are shown, instructor-self, instructor-learner, learner-self, and learner-learner, where generative artificial intelligence serves as a pivotal tool for enhancing learning processes. Generative artificial intelligence's contribution to the pedagogical framework is central, mediating and enriching explicit curriculum delivery and assimilation. Bidirectional arrows between actors and generative artificial intelligence signify a feedback loop allowing for continuous improvement of educational strategies. It underscores generative artificial intelligence's potential to facilitate collaborative processes as well as promoting self-directed learning and peer-to-peer engagement. GenAI: generative artificial intelligence.



GenAI integration by instructors can lower activation barriers to create dynamic, engaging, personalized, and efficient learning environments that optimize learner outcomes (inclusive of motivation and retrieval). We approach this using a flipped classroom (Figure 1), content gamification, streamlined workflows, team-based learning, knowledge gap analysis, and consistent feedback using "exit tickets," all facilitated using GenAI tools. The concept of classroom "flipping" has gained attention in recent years as an approach to instructional delivery. Flipping involves inverting traditional curriculum structure with learners acquiring, or at least engaging, new content outside of structured class time and using active learning methods during class to reinforce, expand upon, and use retrieval practice to reinforce and learn content [17]. All of which is instructor-guided as part of the instructor-learner core axis (Figure 4) [18]. If done well, this approach allows instructors to focus on providing targeted and personalized feedback, requires higher-order critical assembly and thinking skills, and facilitates meaningful discussion or reinforcement during class time for learners. Core tenants of this approach are engaging motivation for learners and expanded retrieval practice outside of high-stress graded assessments like quizzes or exams. Though

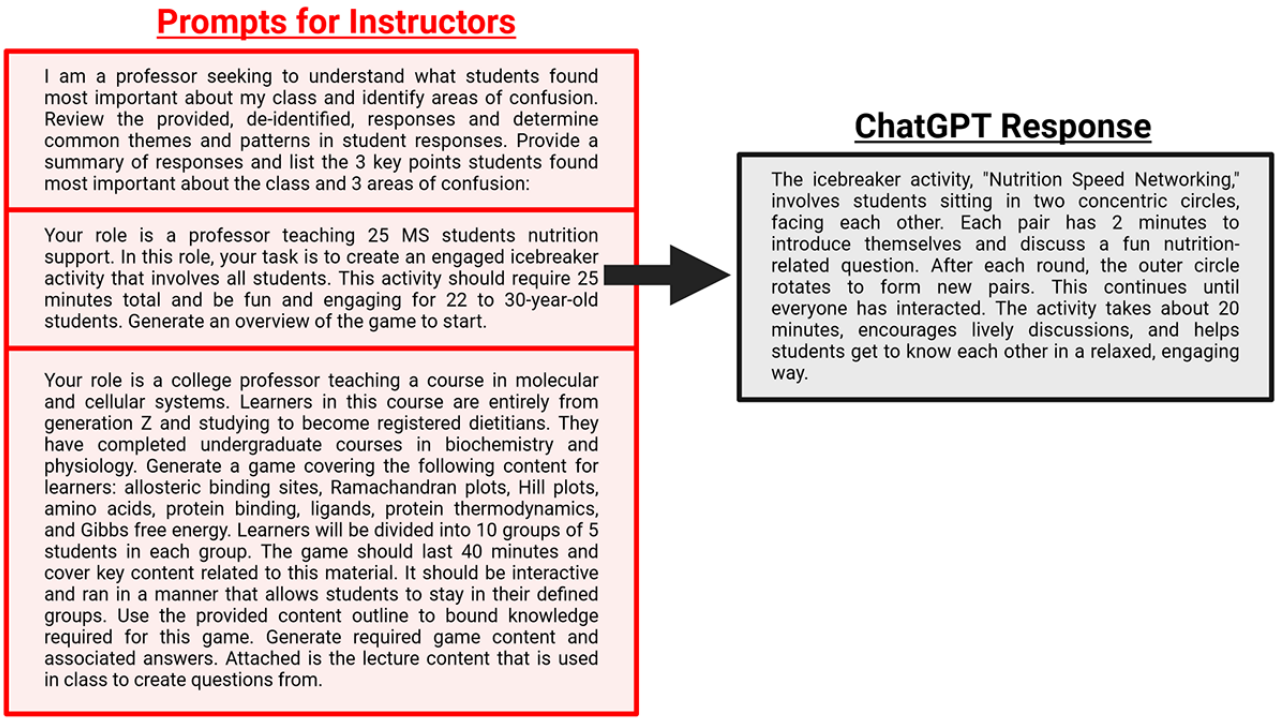
XSL•FO
RenderX

our experience with the above approach involves class sizes ranging from 5‑40 learners, others have successfully implemented flipped classroom methods with more than 200‑300 learners [19]. Great examples of this dynamic approach are “metabolic melodies” in which the instructor, Dr Kevin Ahern from Oregon State University, uses complex biochemistry content to generate songs set to popular music such as “Yellow Submarine” by the Beatles. In this instance, Dr Ahern is extremely creative and a musician with an affinity for the Beatles. GenAI empowers instructors, even those lacking creative brilliance, to turn complex content into interactive and dynamic content such as games, clinical case scenarios, creative narratives, or even music and images. Thus, leveraging GenAI in a flipped classroom environment can reduce instructor workload while improving learning outcomes.

Gamification of course content is one mechanism toward merging GenAI tools with positive learner outcomes. GenAI can quickly generate games (eg, Kahoot!, bingo, Jeopardy!, crossword puzzles, or quiz show questions) using content outlines, slides, or even lecture as input (eg, Figure 5). Games are a low stress means of retrieval practice while promoting an interactive and engaged classroom experience [20]. GenAI can also generate complex clinical practice scenarios with rich hypothetical patient details. These scenarios require learners to use critical thinking and diverse knowledge outputs in a similar method to group-based socratic questioning used in medical education. One area we have had great success in using GenAI is developing gap analysis surveys for learners to assess knowledge levels upon course entry, midcourse for progression, and end of course for effectiveness relative to course objectives. GenAI can quickly generate gap assessments using course learning objectives, prerequisite course content outlines and turnkey instructor needs to provide immediate input on needed content modifications when introducing new content. This is an effective approach to identify knowledge gaps or needs that an instructor may assume are covered in prior courses, or were covered but not retained. Finally, GenAI can be leveraged to streamline workflows before, during, and after content delivery. This includes using templates to generate individualized self-assessment tools for learners, providing narrative feedback for learners on correct and incorrect responses, turning content outlines into focused assessments, integrating lecture modalities under core and redundant learning objectives, and statistical analysis on batched learner responses to identify learning gaps post-content delivery (Figures 1 and 6) [21]. Thus, improving learner outcomes using GenAI tools shows significant promise even with clear limitations, discussed below, and rapidly evolving tools. A central theme for instructors when considering GenAI integration into preplanning, execution, and postanalysis is to balance promoting motivation with opportunities for knowledge retrieval. Course guidelines and instructor expectations must be very clearly defined as to acceptable GenAI use during a given course. This is important considering the current environment where broad institutional or district policies may be lacking, or nonexistent, and variability in what is acceptable between different instructors and courses. Effective communication and clear policies and procedures remain the most important means to avoid academic misconduct or malfeasance. Adapting retrieval strategies to accommodate different learning styles, while ensuring inclusive and personalized learning experiences, is important yet challenging to implement in practice. GenAI holds promise for instructors as these tools provide opportunities to reduce activation barriers (eg, time constraints) toward delivering more effective content and meaningful assessments.

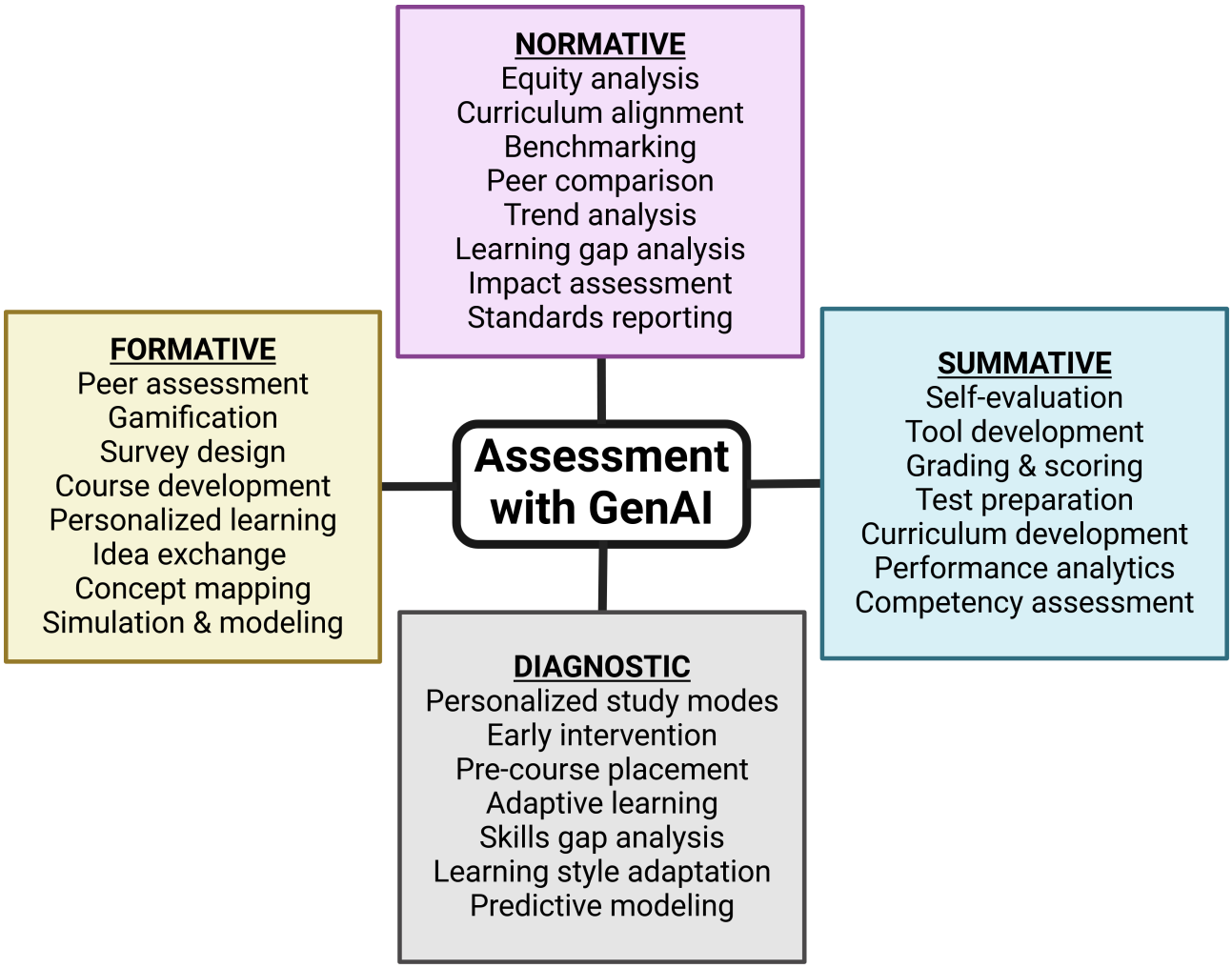
XSL•FO
RenderX

Figure 5. Example instructor prompts for leveraging generative artificial intelligence. Three example instructor input prompts are shown (red box) with representative ChatGPT (accessed on December 10, 2024) output (black box) for the middle prompt (black arrow). A key aspect for instructors is to clearly define the level of instruction being provided, type of learner being instructed, with narrowly defined content scope relative to learning objectives. The latter can be accomplished by inputting content outlines, lecture slides, or narrative summaries.



XSL•FO
RenderX

Figure 6. Generative artificial intelligence integration across assessment types. Generative artificial intelligence use across assessment modalities is shown. In formative assessment, generative artificial intelligence aids in creating interactive content and personalized feedback mechanisms. For summative assessment, generative artificial intelligence capabilities extend to evaluating overall learning achievements and generating comprehensive exams. Normative assessment with generative artificial intelligence focuses on establishing benchmarks and evaluating learning outcomes against standards, while diagnostic assessment leverages generative artificial intelligence for early detection of learning gaps and customization of learning experiences. The central position of "Assessment with GAI" emphasizes its role as a centralized tool, facilitating a comprehensive and integrated approach to educational assessment. GenAI: generative artificial intelligence.



Current Limitations and Future Hurdles

Embracing GenAI tools holds promise for both learners and instructors, yet significant hurdles remain, and user caution is warranted in a rapidly evolving environment of GenAI capabilities, tools, ethics, and acceptable use policies and procedures. Accuracy of GenAI output is a critical aspect that requires careful consideration and diligence, especially when used as a training tool for future clinicians and scientists. LLMs are trained by large datasets and leverage analytics to produce predictions, not logic-derived integrations linking informed input to informed output [22]. Thus, a randomness can exist between prompt design and output obtained. Learners and instructors must both carefully assess and evaluate output from GenAI tools to detect "hallucinations" (ie, incorrect GenAI output that presents as correct). Outside of local application programming interface iterations, general releases of these GenAI models are trained on large datasets that may include unvalidated data from the internet. Thus, even if effort is made to include reliable and authoritative sources, these large training datasets may contain misinformation, biases, or outdated information from uncurated data. We have observed this on several occasions when implementing GenAI as an educational tool, where output sounds entirely factual and even referenced only to be completely incorrect with nonexistent references (this improved substantially in GPT-4o and Claude-3.5 Sonnet, accessed on April 11, 2024). As of this submission, caution is still warranted even with significant improvements in model quality. One effective strategy to minimize accuracy issues is uploading content, such as lecture outlines or even slides, and designing focused prompts working from provided content. If possible, use an institutionally firewalled GenAI tool where content is not shared beyond immediate use. This works well to develop focused learner assessments. This leads to another limitation and hurdle: the rapid pace of GenAI tool development, improvement, and deployment has created an environment where time-limited instructors and digital native learners are increasingly overwhelmed with determining best practices, tools, methods, or even workflows. We have responded to this reality by developing open training courses (eg, on Canvas Learning Management System from Instructure) with frequent

XSL•FO
RenderX

updates and ongoing informational seminars, often targeting instructors, to raise awareness of AI changes as they relate to clinical practice and pedagogy. While this rapidly evolving landscape of tools and capabilities is a challenge, it's also an opportunity to leverage new features and expand impact.

In addition, a caveat to implementing "chatbot" style GenAI like ChatGPT in educational frameworks is the input prompt. The relationship between input prompt and output produced is so integrated that "prompt engineering" is a growing career emerging alongside GenAI [23]. Prompt engineering is the careful construction of input prompts or instructions for GenAI models to influence content, style, voice, depth, and even accuracy of resultant output. How input prompts are constructed is a primary determinant of what models produce, even down to small changes like omitting single words, changing adverbs, or using commas versus numbers for a list [24]. These small changes can produce vastly different results with biased, misleading, or inaccurate information [24]. Truly effective prompts are often complex and descriptive, striking a balance between specificity and openness [25]. Meaning, a prompt that is too specific may limit the ability to generate diverse or creative responses, and a prompt that is too open-ended may result in ambiguous or irrelevant output. The approach we use is constructing "Role+ Bounds+Inputs" style input prompts. In this formula, "Role" involves assigning the chatbot a specific job or identity for the analysis (eg, college professor teaching a specific course and learner type), "Bounds" will establish limitations and constraints for model operation (eg, academic context, subject matter, and level of expected answer), and "Inputs" includes relevant contextual information (eg, content outlines, rubrics, or slide summaries). Using a structured prompt guides GenAI models to produce more focused, accurate, and relevant responses. In addition, one can use a scaffolding approach with iterative prompts building toward a common theme or objective. The complexities of prompt engineering have been discussed elsewhere and leveraging the resources provides a deeper discussion of the benefits of a productive prompt while leveraging GenAI [26,27].

GenAI tools such as ChatGPT are proving transformative, impactful, and hold immense potential for enhancing the educational experience for both learners and instructors. Yet, this new reality is problematic considering the lack of transparency on training data content; ethical and socioeconomic implications; quality control and model accuracy; and the potential to perpetuate bias, loss of voice, and agency for both learners and instructors [21]. Learners should be empowered to make informed decisions regarding their participation in GenAI-related or driven activities, and instructors should encourage learners to ask questions, express concerns, and participate in shaping ethical guidelines related to GenAI use in curricula [11,21]. Furthermore, instructors and administrators have an expectation to proactively define clear policies and procedures for GenAI use in educational settings that provide flexibility for both learners and instructors. A major benefit to GenAI use in educational settings is its collaborative potential to rapidly generate personalized and dynamic content, yet this requires equity in understanding and use. Considering GenAI's rapid evolution, this equity has always been absent in courses we have started since ChatGPT was released in November 2022. It's incumbent on instructors to ensure learners are familiar with GenAI tools being used and available. Data-driven insights from GenAI analytics enable instructors to provide targeted support to individual learners as skills develop and courses progress. This can facilitate multimodal learning experiences, incorporating various media formats such as videos, audio, and interactive simulations mentioned above. Significant hurdles remain regarding GenAI use in higher education. These include data ownership and privacy, output accuracy, linking learner needs with instructor resources, and ensuring sufficient training to avoid equity and GenAI skills gaps when being used. Academic institutions are increasingly looking at internal, data-protected and firewalled, GenAI resources (eg, Microsoft CoPilot) yet there remain limited options for analyzing course content and metrics. Finally, ethical concerns associated with leveraging GenAI by both learners and instructors is a major (probably primary at most institutions) topic of discussion.

Conclusions

Our experiences in leveraging GenAI across 5 academic semesters have, overall, been very positive. This implementation is from the perspective of informing and training both learners and instructors; establishing clear policies and procedures relating to academic misconduct at the department, college, and institutional levels; ensuring equity and ability in use; and constant vigilance regarding content accuracy and limitations. LLMs and GenAI have evolved through decades of iterative research across multiple disciplines including statistics, mathematics, and computer science data science [28]. They are not new technologies. Development of the transformer architecture in 2017 was a key transition point for the emergence of current "chatbots" gaining momentum in popular media and use [29]. This technology continues to evolve with diffusion, attention mechanism variants, and retrieval-enhanced transformer mechanisms being new examples of how GenAI technology is rapidly evolving [30]. Through leveraging large datasets with high-demand computational needs, current GenAI models show significant promise. The important point being, these models (eg, ChatGPT or Claude) excel at pattern recognition yet struggle with defining logical connections between training data and outputs produced ("reasoning"). This is an important caveat for application in higher educational settings focused on critical thinking, developing advanced knowledge and skills within specific disciplines, clinical training, and scientific discovery. As such, it's essential for instructors, administrators, policymakers, institutions, districts, and learners to collaborate and communicate toward what this future will look like as GenAI models evolve. Through this collaboration, GenAI use in educational settings can be leveraged while minimizing negative aspects like potential misconduct, data privacy, algorithmic bias, accuracy, and equity concerns. We argue that GenAI can play a valuable role in higher education settings to improve learner motivation and knowledge retrieval while facilitating workflows and content generation for instructors. This viewpoint explores GenAI's potential as an educational tool including alignment with learning theories (eg, behaviorism and cognitive load theory),

XSL•FO
RenderX

implications for learners and instructors (eg, flipped classrooms and self-directed assessments), responsible implementation (eg, bias and equity), and evolving challenges (eg, hallucinations and misconduct).

Acknowledgments

The authors would like to extend appreciation to professional and graduate students from the Department of Nutritional Sciences and College of Medicine Graduate Program in Biological Sciences for providing feedback and insights during spring and fall 2023 courses as we developed material presented.

Conflicts of Interest

None declared.

References

1. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023 Mar 19;11(6):887. [doi: 10.3390/healthcare11060887] [Medline: 36981544]
2. Arif TB, Munaf U, Ul-Haque I. The future of medical education and research: Is ChatGPT a blessing or blight in disguise? Med Educ Online 2023 Dec;28(1):2181052. [doi: 10.1080/10872981.2023.2181052] [Medline: 36809073]
3. Stretton B, Kovoor J, Arnold M, Bacchi S. ChatGPT-based learning: generative artificial intelligence in medical education. Med Sci Educ 2024 Feb;34(1):215-217. [doi: 10.1007/s40670-023-01934-5] [Medline: 38510403]
4. Gilbert TK. Generative AI and generative education. Ann N Y Acad Sci 2024 Apr;1534(1):11-14. [doi: 10.1111/nyas.15129] [Medline: 38512308]
5. Rodriguez DV, Lawrence K, Gonzalez J, et al. Leveraging generative AI tools to support the development of digital solutions in health care research: case study. JMIR Hum Factors 2024 Mar 6;11:e52885. [doi: 10.2196/52885] [Medline: 38446539]
6. Raza MM, Venkatesh KP, Kvedar JC. Generative AI and large language models in health care: pathways to implementation. NPJ Digit Med 2024 Mar 7;7(1):62. [doi: 10.1038/s41746-023-00988-4] [Medline: 38454007]
7. Montazeri M, Galavi Z, Ahmadian L. What are the applications of ChatGPT in healthcare: Gain or loss? Health Sci Rep 2024 Feb;7(2):e1878. [doi: 10.1002/hsr2.1878] [Medline: 38361810]
8. Tessler I, Wolfovitz A, Livneh N, et al. Advancing medical practice with artificial intelligence: ChatGPT in healthcare. Isr Med Assoc J 2024 Feb;26(2):80-85. [Medline: 38420977]
9. Aydın Ö, Karaarslan E. Is ChatGPT leading generative AI? What is beyond expectations? SSRN Journal 2023. [doi: 10.2139/ssrn.4341500]
10. Weidener L, Fischer M. Teaching AI ethics in medical education: a scoping review of current literature and practices. Perspect Med Educ 2023;12(1):399-410. [doi: 10.5334/pme.954] [Medline: 37868075]
11. Adams C, Pente P, Lemermeyer G, Rockwell G. Ethical principles for artificial intelligence in K-12 education. Comput Artif Intell 2023;4:100131. [doi: 10.1016/j.caeai.2023.100131]
12. Meece JL, Anderman EM, Anderman LH. Classroom goal structure, student motivation, and academic achievement. Annu Rev Psychol 2006;57:487-503. [doi: 10.1146/annurev.psych.56.091103.070258] [Medline: 16318604]
13. Morris LS, Grehl MM, Rutter SB, Mehta M, Westwater ML. On what motivates us: a detailed review of intrinsic v. extrinsic motivation. Psychol Med 2022 Jul;52(10):1801-1816. [doi: 10.1017/S0033291722001611] [Medline: 35796023]
14. Han K. Fostering students' autonomy and engagement in EFL classroom through proximal classroom factors: autonomy-supportive behaviors and student-teacher relationships. Front Psychol 2021;12:767079. [doi: 10.3389/fpsyg.2021.767079] [Medline: 34744946]
15. Bjork RA, Bjork EL. Desirable difficulties in theory and practice. J Appl Res Mem Cogn 2020;9(4):475-479. [doi: 10.1016/j.jarmac.2020.09.003]
16. Roediger HL, Agarwal PK, McDaniel MA, McDermott KB. Test-enhanced learning in the classroom: long-term improvements from quizzing. J Exp Psychol Appl 2011 Dec;17(4):382-395. [doi: 10.1037/a0026252] [Medline: 22082095]
17. Uzunboylu H, Karagozlu D. Flipped classroom: a review of recent literature. WJET 2015;7:142-147 [FREE Full text] [doi: 10.18844/wjet.v7i2.46]
18. Schiff D. Out of the laboratory and into the classroom: the future of artificial intelligence in education. AI Soc 2021;36(1):331-348. [doi: 10.1007/s00146-020-01033-8] [Medline: 32836908]
19. Ahern K. Teaching biochemistry online at Oregon State University. Biochem Mol Biol Educ 2017 Jan 2;45(1):25-30. [doi: 10.1002/bmb.20979] [Medline: 27228905]
20. Lee JJ, andHammer J. Gamification in education: What, how, why bother? Academic exchange quarterly 2011;15(2):1-5.
21. Dave M, Patel N. Artificial intelligence in healthcare and education. Br Dent J 2023 May;234(10):761-764. [doi: 10.1038/s41415-023-5845-2] [Medline: 37237212]
22. Kassab J, El Dahdah J, Chedid El Helou M, et al. Assessing the accuracy of an online chat-based artificial intelligence model in providing recommendations on hypertension management in accordance with the 2017 American College of

XSL•FO
RenderX

Cardiology/American Heart Association and 2018 European Society of Cardiology/European Society of Hypertension Guidelines. Hypertension 2023 Jul;80(7):e125-e127. [doi: 10.1161/HYPERTENSIONAHA.123.21183] [Medline: 37190998]
23. Heston TF, Khun C. Prompt engineering in medical education. Medicine and Pharmacology Preprint posted online on 2023. [doi: 10.20944/preprints202307.0813.v1]
24. Shunyu Yao DY, Zhao J, Shafran I, Griffiths TL, Cao Y, Narasimhan K. Tree of thoughts: deliberate problem solving with large language models. arXiv. Preprint posted online on 2023. [doi: 10.48550/arXiv.2305.10601]
25. Ray PP. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems 2023;3:121-154. [doi: 10.1016/j.iotcps.2023.04.003]
26. Zaghir J, Naguib M, Bjelogrlic M, Névéol A, Tannier X, Lovis C. Prompt engineering paradigms for medical applications: scoping review. J Med Internet Res 2024 Sep 10;26:e60501. [doi: 10.2196/60501] [Medline: 39255030]
27. Abhari S, Fatahi S, Saragadam A, Chumachenko D, Pelegrini Morita P. A road map of prompt engineering for ChatGPT in healthcare: a perspective study. Stud Health Technol Inform 2024 Aug 22;316:998-1002. [doi: 10.3233/SHTI240578] [Medline: 39176959]
28. Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. Gastrointest Endosc 2020 Oct;92(4):807-812. [doi: 10.1016/j.gie.2020.06.040] [Medline: 32565184]
29. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomes AN, et al. Attention is all you need computation and language. arXiv. URL: https://arxiv.org/abs/1706.03762v7
30. Shamshad F, Khan S, Zamir SW, et al. Transformers in medical imaging: a survey. Med Image Anal 2023 Aug;88:102802. [doi: 10.1016/j.media.2023.102802] [Medline: 37315483]

Abbreviations

AI: artificial intelligence
GAI: generative artificial intelligence

Edited by B Lesselroth; submitted 05.04.24; peer-reviewed by A Das, C Lian; revised version received 11.11.24; accepted 02.01.25; published 11.03.25.

Please cite as:
Monzon N, Hays FA
Leveraging Generative Artificial Intelligence to Improve Motivation and Retrieval in Higher Education Learners
JMIR Med Educ 2025;11:e59210
URL: https://mededu.jmir.org/2025/1/e59210
doi:10.2196/59210

© Noahlana Monzon, Franklin Alan Hays. Originally published in JMIR Medical Education (https://mededu.jmir.org), 11.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Designing Personalized Multimodal Mnemonics With AI: A Medical Student's Implementation Tutorial

Noor Elabd[1*]; Zafirah Muhammad Rahman[1*]; Salma Ibrahim Abu Alinnin[1]; Samiyah Jahan[1]; Luciana Aparecida Campos[2], PhD; Ovidiu Constantin Baltatu[1,2], MD, PhD

[1]College of Medicine, Alfaisal University, PO Box 50927, Riyadh, Saudi Arabia

[2]Center of Innovation, Technology, and Education, Anhembi Morumbi University, Sao Jose dos Campos, Brazil

[*]these authors contributed equally

Corresponding Author:
Ovidiu Constantin Baltatu, MD, PhD
College of Medicine, Alfaisal University, PO Box 50927, Riyadh, Saudi Arabia

Abstract

Background: Medical education can be challenging for students as they must manage vast amounts of complex information. Traditional mnemonic resources often follow a standardized approach, which may not accommodate diverse learning styles.

Objective: This tutorial presents a student-developed approach to creating personalized multimodal mnemonics (PMMs) using artifical intelligence tools.

Methods: This tutorial demonstrates a structured implementation process using ChatGPT (GPT-4 model) for text mnemonic generation and DALL-E 3 for visual mnemonic creation. We detail the prompt engineering framework, including zero-shot, few-shot, and chain-of-thought prompting techniques. The process involves (1) template development, (2) refinement, (3) personalization, (4) mnemonic specification, and (5) quality control. The implementation time typically ranges from 2 to 5 minutes per concept, with 1 to 3 iterations needed for optimal results.

Results: Through systematic testing across 6 medical concepts, the implementation process achieved an initial success rate of 85%, improving to 95% after refinement. Key challenges included maintaining medical accuracy (addressed through specific terminology in prompts), ensuring visual clarity (improved through anatomical detail specifications), and achieving integration of text and visuals (resolved through structured review protocols). This tutorial provides practical templates, troubleshooting strategies, and quality control measures to address common implementation challenges.

Conclusions: This tutorial offers medical students a practical framework for creating personalized learning tools using artificial intelligence. By following the detailed prompt engineering process and quality control measures, students can efficiently generate customized mnemonics while avoiding common pitfalls. The approach emphasizes human oversight and iterative refinement to ensure medical accuracy and educational value. The elimination of the need for developing separate databases of mnemonics streamlines the learning process.

(JMIR Med Educ 2025;11:e67926) doi:10.2196/67926

KEYWORDS

medical education; personalized learning; prompt engineering; multimodal learning; memory techniques; dual-coding theory; student-centered approach; student-centered; large language model; natural language processing; NLP; machine learning; AI; ChatGPT; medical student; digital literacy; health care professional

Introduction

Problem Statement

Medical education presents students with the challenge of managing vast amounts of complex information. Mnemonics, memory techniques using associations and patterns, have demonstrated efficacy in improving the encoding and retrieval of medical knowledge [1]. These aids enhance learning and recall by transforming information into more memorable formats through elaborative encoding, retrieval cues, and imagery [2]. However, traditional standardized approaches often fail to accommodate diverse learning preferences, necessitating flexible applications that cater to individual needs.

Theoretical Framework

Paivio's dual-coding theory provides the theoretical foundation for this tutorial, supporting the integration of multimodal tools in education. By encoding both verbal and visual information through separate but interconnected pathways, students' understanding of academic vocabulary can be enhanced [3]. This theory underpins the potential effectiveness of multimodal mnemonics in medical education, particularly when combined

XSL•FO
RenderX

with personalization. Research indicates that personalized mnemonic techniques yield superior recall performance compared to standard strategies, with students using self-generated mnemonics demonstrating better performance on recall tasks [4,5].

Current State of Artificial Intelligence in Medical Education

Recent advances in generative artificial intelligence (AI) technologies have created new opportunities for personalized learning aid creation [6]. AI tools such as ChatGPT and DALL-E have demonstrated proficiency in generating creative and personalized content [7]. ChatGPT, a large language model, uses natural language processing to understand context and generate human-like text responses [8]. It can create diverse textual outputs, including various types of mnemonics. DALL-E, on the other hand, is an AI model designed to generate images from textual descriptions [9]. However, most AI applications in medical education currently focus on data analysis and pattern recognition rather than creative content generation for learning support [10].

Tutorial Aims and Target Audience

This tutorial aims to provide medical students with practical guidance for creating personalized multimodal mnemonics (PMMs) using AI tools. Through a systematic approach, we detail the step-by-step process of generating text and visual mnemonics using ChatGPT and DALL-E, incorporating templates and examples for effective prompt engineering. The tutorial shares strategies for personalizing mnemonics based on individual learning preferences while addressing common challenges and their solutions in AI-assisted mnemonic creation.

The primary audience includes medical students seeking to enhance their learning through personalized AI-assisted mnemonics. Secondary audiences include medical educators interested in implementing these tools in their teaching practice and students in other health care fields who can adapt these methods to their specific needs. By following this tutorial, readers will learn to create personalized learning aids that combine text and visual elements, potentially improving their ability to retain and recall complex medical information. The approach emphasizes practical implementation while maintaining academic rigor, making it accessible to both novice and experienced users of AI tools in educational settings.

Methods

PMMs: Tool Selection and Rationale

This tutorial uses ChatGPT and DALL-E 3 as the primary AI tools for creating PMMs. ChatGPT (GPT-4 model) was selected for text mnemonic generation due to its advanced natural language processing capabilities and ability to generate diverse textual outputs [11]. DALL-E 3 was chosen for visual mnemonic creation based on its proficiency in generating detailed, concept-relevant images from textual descriptions [12]. These tools were selected for their complementary strengths in producing textual and visual content, respectively, allowing for the creation of comprehensive, multimodal mnemonics. Both tools are accessible through OpenAI's platform, with ChatGPT available through both free and paid subscriptions, and DALL-E 3 using a credit-based system.

Configuration Settings and Access

For optimal results in medical mnemonic generation, we recommend using ChatGPT's GPT-4 model with a temperature setting of 0.7, which provides an effective balance between creativity and accuracy in medical content generation. For DALL-E 3, the high-quality setting ensures maximum detail and clarity in visual representations. While both tools offer free tiers, a professional subscription is recommended for consistent access to the latest model versions and enhanced capabilities.

Prompt Engineering Framework and Quality Assessment

The implementation of PMMs requires a systematic approach to prompt engineering and quality assessment. The process begins with zero-shot learning, where we provide clear instructions without examples, allowing the AI to generate mnemonics based purely on prompt structure. When initial results require refinement, we implement few-shot learning by providing 1 to 2 successful examples to guide the AI. For complex medical concepts, we use chain-of-thought prompting to break down the mnemonic creation process into logical steps.

Zero-shot prompting (providing direct instructions without examples) allows the AI to generate outputs based on its pretrained knowledge [13]. For example, a simple prompt like "Create a memorable mnemonic for the Krebs cycle intermediates" tests the AI's baseline capabilities without additional guidance.

Few-shot prompting (including 1 to 2 successful examples before the target prompt) helps guide the AI by demonstrating desired outputs [13,14]. For example, showing a successful biochemical pathway mnemonic before requesting one for the Krebs cycle improves output quality by providing clear examples of the expected format and style.

Chain-of-thought prompting breaks complex tasks into logical steps, improving accuracy through structured reasoning [15]. For example, "First, list intermediates. Then, identify key features. Finally, create a mnemonic." This systematic approach helps ensure comprehensive and accurate outputs, particularly for complex medical concepts.

The implementation followed a five-stage iterative process: (1) template development, in which adaptable prompt templates for text and visual mnemonics are created; (2) refinement, which optimizes prompts through testing of various structures and keywords; (3) personalization, which integrates learning preferences and personal associations and adds options for imagery, humor, and clinical relevance [16]; (4) mnemonic specification, in which prompts for various mnemonic types (acronyms, phrases, rhymes) are created, with corresponding visual representations; and (5), quality control, which is based on peer-to-peer review discussions.

The mnemonics generated through this process were evaluated in peer-to-peer discussions among student authors, facilitated and guided by mentors. This iterative feedback process, integral to quality control, not only ensured medical accuracy and

XSL•FO
RenderX

educational value by leveraging student insights and expert guidance, but also trained students to critically assess the quality, accuracy, and effectiveness of AI-generated content. Identified inaccuracies or areas for improvement directly informed subsequent prompt adjustments.

The medical concepts used in this study were carefully selected from ongoing medical courses, ensuring immediate relevance to current learning needs. This selection process focused on identifying complex topics that students found challenging to memorize while ensuring diverse representation of medical subjects and validation against standard medical resources.

For text mnemonic generation, we developed a basic template structure that incorporates medical accuracy, memorability, and personalization elements: "Create a memorable [mnemonic type] for [medical concept]. Focus on [key aspects]. Make it [characteristics: funny/clinical/etc]. Include [specific elements] that relate to [learning context]."

For visual mnemonic creation, the template emphasizes clarity and medical accuracy: "Generate a [style] image depicting [mnemonic content] for [medical concept]. Emphasize [key visual elements]. Ensure medical accuracy and clarity." These templates serve as starting points and can be customized based on individual learning preferences and specific medical concepts.

Common Challenges and Solutions

Through our implementation process, we identified several common challenges. Inaccurate medical terminology can be addressed by including specific medical terms in prompts. Unclear visual representations are improved by specifying anatomical or clinical details. When mnemonics become overly complex, requesting step-by-step breakdowns helps maintain clarity and usability. These solutions emerged from practical experience and continue to evolve as we refine the process.

Ethical Considerations

This educational methodology development did not require formal ethical review as it did not involve human subjects research, collected no personal data, and used only publicly available AI tools as part of regular educational activities.

Results

The implementation of PMMs using AI tools demonstrated both potential and limitations across a range of medical concepts. Through systematic testing and refinement, we identified key performance metrics, quality assessment outcomes, and practical implementation challenges.

Generation Performance

The PMM generation process exhibited consistent performance characteristics. Text mnemonic generation via ChatGPT consistently required 2 to 3 minutes per concept. Generating corresponding visual mnemonics with DALL-E 3 required 3 to 5 minutes per concept. Reaching a satisfactory mnemonic typically involved 1 to 3 iterative attempts. The initial success rate, defined as achieving acceptable output on the first attempt, was 85%. After applying quality control and refinement procedures, this success rate increased to 95%. These data suggest that generating PMMs is feasible within a reasonable timeframe, particularly when incorporating iterative refinement.

Quality Assessment and Examples

Table 1 presents 6 examples of AI-generated PMMs, illustrating the range of mnemonic types generated and highlighting observed successes and areas for improvement. The table includes the target medical concept, the text mnemonic generated by ChatGPT, the visual prompt provided to DALL-E 3, the resulting visual mnemonic, and specific limitations encountered.

Through systematic documentation of the implementation process, we observed that achieving satisfactory results typically required 1 to 3 iterations per concept. Initial attempts often needed refinement in areas such as the precision of medical terminology, visual clarity, or text-image alignment. The examples in Table 1 showcase both successful implementations and instances where refinement was necessary, providing practical insights for users of this tutorial.

The development of a mnemonic for bacterial cell wall antibiotics serves as a representative example. The initial AI-generated mnemonic incorrectly used the term "harbipenems" instead of "carbapenems." This error, corrected in subsequent iterations, demonstrates the need for human validation, specifically by the medical student or a medical professional. Other examples in Table 1 illustrate the variety of generated mnemonic types, including acronyms, rhymes, and short phrases, paired with their visual counterparts. While generally effective, some visual mnemonics presented challenges in terms of clarity, anatomical accuracy, and alignment with the corresponding text mnemonic.

XSL•FO
RenderX

Table . Artifical intelligence (AI)–generated personalized multimodal mnemonics.

Medical concept	Prompts used and results	Text mnemonic	Visual and text mnemonic	Observations and limitations
Antibiotics targeting cell wall	First prompt: "Make a mnemonic for antibiotics that target the cell wall, you can group the antibiotics by their mechanism and make it memorable." Second prompt: "Using the mnemonic, make a memorable photo." Result: photo was not memorable. Third prompt: "Retry this time make a funny mnemonic." Result: the cat image.	Mnemonic: "very big purple cats bake happy eggs." • V: vancomycin • B: beta-lactams • P: penicillins • C: cephalosporins • B: bacitracin • H: harbapenems		Creative and memorable representation, but incorrect text
Types of leukocytes	Prompt: "Develop a mnemonic to remember the different types of leukocytes. Also, create a funny and cartoonish visual that represents this mnemonic in a memorable way. The illustration should effectively depict the mnemonic."	Mnemonic (for leukocytes): "never let monkeys eat bananas." • N: neutrophils • L: lymphocytes • M: monocytes • E: eosinophils • B: basophils		Engaging visualization, but incorrect text
Life cycle of malaria	Prompt: "Create a mnemonic to help remember the malaria life cycle, and design a humorous, cartoon-style visual to represent it in a memorable way. The illustration should clearly depict the mnemonic for easy recall."	Mnemonic (each word stands for a stage of the malaria parasite's life cycle): "salty mice bite liver, chilling red blood cells." • S: sporozoites (transmitted via mosquito bite) • M: migrate to the liver • B: burst out of the liver as merozoites • L: liver stages • C: chills and fevers (as the parasites enter red blood cells) • Red blood cells: parasites invade and replicate in red blood cells		The AI was occasionally unable to fully adhere to detailed prompts in image generation. Additionally, the text included in the image was incorrect.
Renin-angiotensin-aldosterone system (RAAS)	First prompt used: "Generate a funny mnemonic for Renin-Angiotensin-Aldosterone System (RAAS) but make it simple and easy to remember." Second prompt: "Create simple and memorable image of this mnemonic 'Keen Elephants Always Raise Blood.' Use less texts, clear and correct spelling. Place the mnemonic as a title."	Mnemonic: "keen elephants always raise blood." • K: kidneys release renin when blood pressure or sodium levels are low. • E: enzyme (renin) converts angiotensinogen to angiotensin I. • A: ACE (angiotensin-converting enzyme) converts angiotensin I to angiotensin II. • R: raises blood pressure by vasoconstriction (angiotensin II constricts blood vessels). • B: blood volume increases as aldosterone causes sodium and water retention.		This mnemonic emphasizes the kidney's keen response to blood pressure regulation. Text is not clear.

XSL•FO
RenderX

Medical concept	Prompts used and results	Text mnemonic	Visual and text mnemonic	Observations and limitations
Non-selective COX inhibitors	Prompt used: Make a funny memorable picture mnemonic for me to remember the *Non-Selective COX Inhibitors*. They include: Salicylates: Aspirin; Pyrazalone derivatives; Propionic acid derivatives: Ibuprofen, Naproxen; Indole Derivatives: Indomethacin; Anthranilic acid derivative; Aryl acetic acid derivative: Diclofenac; Oxicam derivative; Pyrrole derivatives."	Mnemonic: "a perfect brew is delightfully aromatic, outstandingly pleasant." • A: aspirin • P: pyrazalone derivatives • B: ibuprofen • I: indole derivatives • D: diclofenac • A: anthranilic acid derivative • O: oxicam derivative • P: pyrrole derivatives		The coffee-themed elements like cups, beans, and mugs may be too abstract and unrelated to the actual topic. The picture lacks clarity and is not memorable.
Naeglaria fowleri	First prompt used: "Make an image to remember naeglaria fowleri, brain eating heat loving free living with a mneumonic for its major facts." Second prompt: "Make it funny: >" Response: "I couldn't generate the image because the content did not align with the guidelines for image creation. If you'd like, we can adjust the concept or explore a different approach. Let me know how you'd like to proceed!" Prompt: "Try again" Response: "The guidelines for image creation restrict certain content related to how medical or potentially dangerous organisms are portrayed, especially if humor or exaggerated characteristics could be seen as trivializing the topic. Creating a funny depiction of a 'brain-eating' amoeba might fall under those restrictions."	__[a]		The AI refused to generate the content, highlighting challenges in discerning educational needs from potentially inappropriate content, emphasizing the need for human oversight and refinement.

[a]Not applicable.

Generic Templates for Prompt Engineering

For text mnemonic generation (with ChatGPT), the following generic prompt template was developed: "Create a memorable sentence mnemonic for [medical concept]. The mnemonic should be [characteristic 1: funny/simple/easy to remember] and [characteristic 2: relevant to clinical practice/focused on key steps/highlight main components]. Each word or part of the sentence should represent a key aspect of the concept. If possible, incorporate [optional element: wordplay/alliteration/vivid imagery]. Make it relatable to [personal preference: a specific scenario/everyday objects/animals]."

For creation of the visual mnemonic (with DALL-E 3), the following generic prompt template was used: "Generate a [style: cartoon/funny/medical illustration] depicting the mnemonic '[text mnemonic]' for [medical concept]. The image should be [characteristic 1: visually engaging/humorous/clear] and [characteristic 2: memorable/related to the mnemonic words]. Incorporate [specific visual elements: anthropomorphized objects/exaggerated features/relevant symbols]. Ensure any text is minimal, clear, and correctly spelled. Place the mnemonic sentence as a title."

These templates were iteratively refined based on the quality and relevance of the AI-generated outputs. Examples of specific prompts based on these templates are in Table 1. These templates and examples provide a framework for creating diverse and engaging sentence mnemonics while allowing for customization based on the specific medical concept and desired learning outcomes.

Implementation Challenges and Refinement Strategies

Three key challenges emerged during implementation, leading to the development of targeted refinement strategies. The first

XSL•FO
RenderX

challenge was medical accuracy. Maintaining medical accuracy necessitated continuous verification against established medical resources. Initial outputs occasionally exhibited terminology errors or incomplete conceptual coverage. These issues were addressed by incorporating specific medical terminology in the prompts and implementing a systematic review process involving medical experts.

Second, achieving consistent visual clarity and anatomical accuracy in the AI-generated images presented challenges. Some images lacked clarity or contained inconsistencies between textual and visual elements. We improved visual quality through prompt refinement, including more precise anatomical descriptions and requests for simplified representations of complex concepts.

The third challenge was in content integration. Ensuring seamless integration between the text mnemonic and visual representation required careful prompt design and quality control. A structured review process was implemented to verify that both components effectively reinforced the target medical concept and functioned synergistically to enhance learning.

These findings offer practical observations for educators considering the use of AI-assisted mnemonic generation. While the PMM approach holds promise for personalized learning, our results underscore the essential role of human oversight, domain expertise, and iterative refinement in ensuring accuracy, clarity, and educational value.

Discussion

Principal Findings and Implications

This tutorial demonstrates a practical approach to generating PMMs using readily available AI tools. Our findings highlight the feasibility of creating customized mnemonics within a reasonable timeframe (2-5 minutes per concept, with 1-3 iterative attempts). The combination of text and visual elements aligns with dual-coding theory [3,17], potentially enhancing learning and recall. However, challenges related to medical accuracy, visual clarity, and content integration underscore the crucial role of human oversight and domain expertise. The “harbipenems” error, for example, emphasizes the need for medical professionals to validate AI-generated content. These findings suggest that AI-assisted PMM generation can be a valuable tool for personalized learning, but careful attention to quality control and prompt refinement is essential.

Comparison to the Literature

This tutorial’s approach aligns with the growing interest in applying AI for personalized learning in medical education. While much of the current research focuses on AI for tasks like data analysis [10,18], this tutorial explores the relatively novel application of AI for generating personalized learning content. Our emphasis on multimodal learning resonates with the principles of dual-coding theory [3], which suggests that combining visual and textual representations can enhance learning and memory. Furthermore, the challenges we encountered regarding accuracy and clarity in AI-generated content echo broader concerns in the literature about the need for human oversight in AI-driven educational applications [8,19].

Strengths and Limitations

This tutorial provides a practical, step-by-step guide for generating PMMs using AI, offering readily adaptable prompt templates and illustrative examples. The student-centered perspective offers valuable insights into the practical challenges and potential benefits of this approach.

This tutorial has several limitations. First, the AI models used may exhibit biases, potentially limiting the diversity and novelty of generated PMMs. Second, inaccuracies in visual representations, such as misspellings or mismatches with the text mnemonic, require careful review and correction. Third, current AI models may refuse to generate content for sensitive medical topics, necessitating alternative strategies or manual content creation. Finally, the lack of a formal evaluation with medical students limits the generalizability of our findings and prevents definitive conclusions about the effectiveness of PMMs on learning outcomes.

Future Directions

Future research should investigate the effectiveness of PMMs on learning outcomes through controlled studies comparing PMMs to traditional learning methods. Such studies should use objective measures of learning, such as recall accuracy, learning efficiency, and student satisfaction. Further research should also explore the long-term impact of PMMs on knowledge retention and application. The scalability and adaptability of the PMM approach across diverse medical subjects and educational settings warrant investigation. Additionally, future work should address the ethical considerations surrounding AI-generated educational content, including data privacy, bias, and overreliance on technology [8]. Developing guidelines for the ethical and effective use of AI in mnemonic creation and medical education more broadly will be crucial as this field evolves [19].

Conclusion

This tutorial presents a practical approach to generating PMMs for medical education using the AI tools ChatGPT and DALL-E 3. This approach emphasizes AI as a tool to enhance, rather than replace, traditional learning methods. Originating from medical students seeking to improve their own learning, this tutorial describes a step-by-step process involving prompt engineering, iterative refinement, and quality assessment, illustrated with examples for 6 medical concepts. The personalized nature of the mnemonics, coupled with the multimodal approach, demonstrates potential for enhancing student engagement and facilitating the retention of complex medical concepts. We also highlight key challenges related to medical accuracy, visual clarity, and content integration, underscoring the importance of human oversight and domain expertise in refining AI-generated content. This student-led exploration offers practical guidance and a valuable starting point for educators and students alike interested in leveraging AI for personalized learning in medical education.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Rosi-Schumacher M, DeGiovanni JC. Using the lessons of learning science to improve medical education in otolaryngology. Ear Nose Throat J 2022 Nov;101(9_suppl):16S-19S. [doi: 10.1177/01455613231160509] [Medline: 36825609]
2. Mocko M, Lesser LM, Wagler AE, Francis WS. Assessing effectiveness of mnemonics for tertiary students in a hybrid introductory statistics course. J Stat Educ 2017 Jan 2;25(1):2-11. [doi: 10.1080/10691898.2017.1294879]
3. Soper K, Geske JA, Bronner L, Godfrey M. Improving student understanding of academic assessment vocabulary words using visual cues: a collaborative effort. J Community Engagem Scholarsh 2022 Jul 29;15(1):1. [doi: 10.54656/jces.v15i1.52] [Medline: 36081415]
4. Dirette DP. A comparison of self-generated versus taught internal strategies for working memory. NeuroRehabilitation 2015;36(2):187-194. [doi: 10.3233/NRE-151206] [Medline: 25882198]
5. Wheeler RL, Gabbert F. Using self-generated cues to facilitate recall: a narrative review. Front Psychol 2017;8:1830. [doi: 10.3389/fpsyg.2017.01830] [Medline: 29163254]
6. Altintas L, Sahiner M. Transforming medical education: the impact of innovations in technology and medical devices. Expert Rev Med Devices 2024 Sep;21(9):797-809. [doi: 10.1080/17434440.2024.2400153] [Medline: 39235206]
7. Waikel RL, Othman AA, Patel T, et al. Generative methods for pediatric genetics education. medRxiv. 2023 Aug 2. [doi: 10.1101/2023.08.01.23293506] [Medline: 37790417]
8. Leng L. Challenge, integration, and change: ChatGPT and future anatomical education. Med Educ Online 2024 Dec 31;29(1):2304973. [doi: 10.1080/10872981.2024.2304973] [Medline: 38217884]
9. Adams LC, Busch F, Truhn D, Makowski MR, Aerts H, Bressem KK. What does DALL-E 2 know about radiology? J Med Internet Res 2023 Mar 16;25:e43110. [doi: 10.2196/43110] [Medline: 36927634]
10. Tolentino R, Baradaran A, Gore G, Pluye P, Abbasgholizadeh-Rahimi S. Curriculum frameworks and educational programs in AI for medical students, residents, and practicing physicians: scoping review. JMIR Med Educ 2024 Jul 18;10:e54793. [doi: 10.2196/54793] [Medline: 39023999]
11. Shoham OB, Rappoport N. MedConceptsQA: Open source medical concepts QA benchmark. Comput Biol Med 2024 Nov;182:109089. [doi: 10.1016/j.compbiomed.2024.109089] [Medline: 39276611]
12. Temsah MH, Alhuzaimi AN, Almansour M, et al. Art or artifact: evaluating the accuracy, appeal, and educational value of AI-generated imagery in DALL·E 3 for illustrating congenital heart diseases. J Med Syst 2024 May 23;48(1):54. [doi: 10.1007/s10916-024-02072-0] [Medline: 38780839]
13. Sivarajkumar S, Kelley M, Samolyk-Mazzanti A, Visweswaran S, Wang Y. An empirical evaluation of prompting strategies for large language models in zero-shot clinical natural language processing: algorithm development and validation study. JMIR Med Inform 2024 Apr 8;12:e55318. [doi: 10.2196/55318] [Medline: 38587879]
14. Drori I, Zhang S, Shuttleworth R, et al. A neural network solves, explains, and generates university math problems by program synthesis and few-shot learning at human level. Proc Natl Acad Sci U S A 2022 Aug 9;119(32):e2123433119. [doi: 10.1073/pnas.2123433119] [Medline: 35917350]
15. Khademi S, Palmer C, Dimaguila GL, Javed M, Buttery J. Exploring large language models for detecting online vaccine reactions. Stud Health Technol Inform 2024 Sep 24;318:30-35. [doi: 10.3233/SHTI240887] [Medline: 39320177]
16. Tuttle JJ, Moshirfar M, Garcia J, Altaf AW, Omidvarnia S, Hoopes PC. Learning the Randleman criteria in refractive surgery: utilizing ChatGPT-3.5 versus internet search engine. Cureus 2024 Jul;16(7):e64768. [doi: 10.7759/cureus.64768] [Medline: 39156271]
17. Clark JM, Paivio A. Dual coding theory and education. Educ Psychol Rev 1991 Sep;3(3):149-210. [doi: 10.1007/BF01320076]
18. Tozsin A, Ucmak H, Soyturk S, et al. The role of artificial intelligence in medical education: a systematic review. Surg Innov 2024 Aug;31(4):415-423. [doi: 10.1177/15533506241248239] [Medline: 38632898]
19. Franco D'Souza R, Mathew M, Mishra V, Surapaneni KM. Twelve tips for addressing ethical concerns in the implementation of artificial intelligence in medical education. Med Educ Online 2024 Dec 31;29(1):2330250. [doi: 10.1080/10872981.2024.2330250] [Medline: 38566608]

Abbreviations

AI: artificial intelligence
PMM: personalized multimodal mnemonic

XSL•FO
RenderX

Edited by B Lesselroth; submitted 24.10.24; peer-reviewed by M Alshiekh, R Sajja; revised version received 20.03.25; accepted 06.04.25; published 08.05.25.

Please cite as:
Elabd N, Rahman ZM, Abu Alinnin SI, Jahan S, Campos LA, Baltatu OC
Designing Personalized Multimodal Mnemonics With AI: A Medical Student's Implementation Tutorial
JMIR Med Educ 2025;11:e67926
URL: https://mededu.jmir.org/2025/1/e67926
doi:10.2196/67926

© Noor Elabd, Zafirah Muhammad Rahman, Salma Ibrahim Abu Alinnin, Samiyah Jahan, Luciana Aparecida Campos, Ovidiu Constantin Baltatu. Originally published in JMIR Medical Education (https://mededu.jmir.org), 8.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Creation of the ECHO Idaho Podcast: Tutorial and Pilot Assessment

Ryan Wiet*, PhD; Madeline P Casanova*, PhD; Jonathan D Moore*, PhD; Sarah M Deming*, PhD; Russell T Baker Jr*, PhD, DAT

WWAMI Medical Education Program, Idaho Office of Underserved and Rural Medical Research, University of Idaho, 875 Perimeter Drive, Moscow, ID, United States

*all authors contributed equally

Corresponding Author:
Russell T Baker Jr, PhD, DAT
WWAMI Medical Education Program, Idaho Office of Underserved and Rural Medical Research, University of Idaho, 875 Perimeter Drive, Moscow, ID, United States

Abstract

Background: Project ECHO (Extension for Community Health Outcomes) is an innovative program that uses videoconferencing technology to connect health care providers with experts. The model has been successful in reaching health care providers in rural and underserved areas and positively impacting clinical practice. ECHO Idaho, a replication partner, has developed programming that has increased knowledge and confidence of health care professionals throughout the state of Idaho, United States. Although the ECHO model has a demonstrated ability to recruit, educate, and train health care providers, barriers to attending Project ECHO continuing education (CE) programs remain. The asynchronous nature of podcasts could be used as an innovative medium to help address barriers to CE access that health care professionals face. The ECHO Idaho "Something for the Pain" podcast was developed to increase CE accessibility to rural and frontier providers, while upscaling their knowledge of and competence to treat and assess substance use disorders, pain, and behavioral health conditions.

Objective: This paper describes the creation and preliminary assessment of the ECHO Idaho "Something for the Pain" podcast.

Methods: Podcast episodes consisted of interviews with individuals as well as didactic lectures. Audio from these recordings were edited for content and length and then professionally reviewed by subject matter experts (eg, featured episode speakers). Target audiences consisted of health care providers and community members interested in behavioral health and substance use disorders. Metrics on podcast listeners were assessed using SoundCloud's RSS feed, continuing education survey completion, and iECHO.

Results: The ECHO Idaho "Something for the Pain" podcast's inaugural season comprised 14 episodes with 626 minutes of CE material. The podcast series received a total of 2441 listens from individuals in 14 different cities across Idaho, and 63 health care providers listened and claimed CE credits. The largest professional group was social workers (n=22; 35%).

Conclusions: We provide preliminary evidence that podcasts can be used to provide health care providers with opportunities to access CE material. Health care providers listened to and claimed CE credits from the ECHO Idaho "Something for the Pain" podcast. Project ECHO programs should consider creating podcasts as an additional platform for disseminating ECHO material.

(JMIR Med Educ 2025;11:e55313) doi:10.2196/55313

KEYWORDS

Project ECHO; ECHO Idaho; medical education; medical training; medication teaching; medical knowledge; rural health care; rural medicine; underserved population; underserved people; substance use; substance use disorder; SUD; drug abuse; drug use; alcoholism; addiction; pain; behavioral health; podcast; webinar

Introduction

Project ECHO (Extension for Community Health Outcomes), founded by the University of New Mexico, provides accessible education and specialty training to rural and underserved professionals through an all-teach, all-learn model that uses interactive video conferencing (ie, Zoom technology) [1]. In 2018, through a collaboration between the University of Idaho WWAMI (Washington, Wyoming, Alaska, Montana, Idaho) Regional Medical Education Program and the Idaho Department of Health and Welfare Statewide Healthcare Innovation Plan, ECHO Idaho became the only Project ECHO replication partner in the state of Idaho [2]. Since 2018, ECHO Idaho has provided rural and underserved communities with educational opportunities and cultivated a community of health care professional learners across the state. Researchers assessing the Project ECHO model have found that attendees increase their

XSL•FO
RenderX

knowledge, clinical skills, and confidence, thus increasing the quality of care for patients in rural and underserved communities [1-10]. Health care professionals participating in the ECHO Idaho series have similarly reported increases in both their knowledge and in their confidence to provide specialty care for their patients [2,11].

Despite the myriad documented benefits of the Project ECHO model, challenges relevant to maximizing attendance and increasing overall reach of the program (eg, recruiting participants from rural areas) still exist [2-4,12]. For example, scheduling conflicts and time constraints have been reported as barriers to attending ECHO sessions [5,10,13]. Although some hubs have modified the ECHO model (eg, changed session lengths), no conclusive changes have been recommended and minimal research has been conducted to assess how modifications to the model could positively impact and enhance provider training [10,13]. Other models of continuing education (CE) (eg, e-learning) have been reported to be effective in changing knowledge, self-efficacy, and skills [14]. However, the effect of such models on professional and clinical practice, as well as patient outcomes, remains largely unknown.

Thus, exploring and researching novel or innovative platforms to disseminate information and increase clinical skills for health care providers, while subsequently decreasing potential barriers to attendance, are necessary. One proposed approach is the use of a podcast platform to deliver Project ECHO materials. A podcast may be an effective, interactive, and cost-efficient way to deliver Project ECHO programming while addressing some barriers reported by participants. For example, the asynchronous nature of podcasts would allow providers the flexibility to engage with the material on their own time, and to pause and resume listening when needed. Research has reported that podcasts can be an impactful model to train and enhance professional development [15-17]. Additionally, due to the COVID-19 pandemic dramatically changing the landscape of health care and education [18], the use of podcasts to deliver medical education curricula has increased [19-22]. Researchers have found podcasts increase participant knowledge and self-efficacy while also indirectly impacting professional and clinical practice [19-22]; however, the full scope of the impact of educational podcasts remains unknown [16].

To test the efficacy of the podcast model for delivering medical education, ECHO Idaho developed the “Something for the Pain” podcast, with support from the Idaho WWAMI Medical Education Program and in partnership with the Valley County Opioid Response Project [23]. The podcast was meant to increase CE accessibility for rural and frontier providers, while upscaling their knowledge of and competence to treat and assess behavioral health conditions. The purpose of this paper was to present the development and implementation process of the ECHO Idaho “Something for the Pain” podcast’s inaugural season, as well as to discuss the preliminary findings regarding the reach of the podcast, lessons learned, and proposed future uses of this innovative platform.

Methods

Program Development and Implementation

Program Description

Due to the increased need for more specialty training related to behavioral health, opioids, pain, and substance use disorder (SUD), in 2021, the ECHO Idaho “Something for the Pain” podcast was developed as an innovative approach to disseminate ECHO Idaho materials. The inaugural season contained 14 episodes that presented best-practices and resources for behavioral health, opioid use disorder, and SUD prevention, treatment, and recovery specific to Idaho. The podcast was free to access, and eligible providers were able to obtain no-cost CE credits for listening to an episode.

Episode Development

The ECHO model learning framework typically includes two sections: (1) didactic lectures and (2) a case-based presentation and discussion. To mirror the framework of the ECHO model, the 14 podcast episodes were primarily broken down into two parts: (1) interviews with individuals and (2) didactic presentations. The interviews were with health care professionals and community members who had first-hand experience and knowledge of evidence-based practices and resources pertinent to the Idaho context. The didactic presentations were taken from previously recorded ECHO Idaho sessions across 4 series (ie, Behavioral Health in Primary Care; Opioids, Pain and Substance Use Disorders; Counseling Techniques for Substance Use Disorders; and Viral Hepatitis and Liver Care). Audio from these recordings were edited for content and length and then professionally reviewed by subject matter experts. The subject matter experts also assisted with generating CE-eligible episode assessment questions that could be used to gauge audience engagement.

Audience Recruitment

The primary intended audience for the podcast was health care providers and community members with an interest in SUD and behavioral health. Initial recruitment occurred by using ECHO Idaho’s network of prior ECHO Idaho session attendees, which, at the time of the first episode’s release in May 2021, consisted of approximately 3000 diverse Idaho health care professionals (eg, those who held an MD/DO, PA, NP, MCSW, or LCPC). Members of the ECHO Idaho staff invited participants from the pre-existing ECHO network to listen to the podcast as an additional means of earning CE credit at their convenience. Additionally, ECHO staff developed a targeted marketing campaign that involved personal and bulk emails, newsletter announcements, weekly announcements, paid advertisements, social media posts, print bulk postcard mailings, as well as advertising on other podcast programs. The podcast was housed on SoundCloud, which increased its accessibility through user search engines and algorithms.

Podcast Data Metrics

Listener engagement was tracked using SoundCloud’s RSS feed. The feed provided click metrics (ie, how many listens occurred) for each podcast episode consumed within a specified

XSL•FO
RenderX

timeframe. Eeds, an electronic CE management system, was used to track CE credits claimed following episode and assessment completion. Additionally, to gain insight into listener experience, participants were required to answer a user experience question in Eeds. Lastly, individuals interested in listening to the podcast could register on the ECHO Idaho website. Registration information collected included demographic questions like profession, credentials, primary practice location, sex, and age. iECHO, a web-based proprietary program and management software database, was used to manage participant demographic data (ECHO Institute, University of New Mexico) entered on the registration form.

Data Analysis

Data were exported from Eeds and iECHO and a descriptive statistical analysis was performed using SPSS, version 28 (IBM Corp).

Ethical Considerations

The project was certified as exempt (protocol #23‑150) by the Institutional Review Board at the University of Idaho. Data were deidentified for the purposes of analysis, and informed consent was obtained from all participants prior to their involvement in the study.

Results

RSS Feed

The first season of the ECHO Idaho “Something for the Pain” podcast included 14 episodes (released May 2021 to June 2022); 13 episodes were related to perinatal SUD, and 1 bonus episode gave a brief history of Project ECHO and Vandal Theory (Table 1). The average length of an episode was 45 (SD 11.9) minutes and the average number of listens per episode was 188 (SD 46.5) (Table 1). The season provided a total of 626 minutes of CE material available for perpetual access. The podcast was released on the streaming platforms SoundCloud, Apple Podcasts, Google Podcasts, Spotify, Sticher, and iHeartRadio. As of April 19, 2023, the initial season of the podcast had garnered over 2000 listeners from various parts of the United States, with most of the listeners based in Idaho.

XSL•FO
RenderX

Table . Information on each episode including release date, title, length, and total listens of each episode.

Release date	Episode	Length (minutes)	Listens
5/11/2021	Episode 1: Framework for Addiction as Disease (feat. Craig Lodis, PhD)	31	286
5/17/2021	Episode 2: State of Substance Use in Idaho (feat. Amy Jeppesen, LCSW, ACADC)	47	212
6/3/2021	Bonus Episode: Project ECHO Origin Story & The Vandal Theory (feat. Sanjeev Arora, MD)	37	N/A[a]
7/1/2021	Episode 3: De-escalation Techniques and the Valley County Court Services' Diversion Program (feat. Abby Abbondondalo and Skip Clapp)	45	159
7/11/2021	Episode 4: Harm Reduction and Valley County's Opioid Response Project (feat. Brenda Hoyt, NP, Courtney Boyce and Shelly Hitt)	47	165
7/23/2021	Episode 5: Motivational Interviewing and Donnelly's The Change Clinic (feat. Deb Thomas and Barbara Norton)	60	191
8/9/2021	Episode 6: LaDessa Foster Talks Levels of Care in Substance Use Disorder Treatment (feat. LaDessa Foster)	32	164
8/23/2021	Episode 7: Monica Forbes Talks SMART Recovery, Stigma and Reentering Society Post-Incarceration (feat. Monica Forbes)	35	178
9/1/2021	Episode 8: Marjorie, Wilson Talks Idaho's Syringe Service Programs (feat. Marjorie, Wilson and Ian Trosoyer, DNP)	51	185
9/13/2021	Episode 9: LaDessa Foster Talks Managing Clinical Services for Patients and Providers (feat. LaDessa Foster)	30	274
3/2/2022	Episode 10: Lindsay Brown Talks Peer Recovery Supports (feat. Lindsay Brown)	55	168
3/25/2022	Episode 11: Talking Telehealth in SUD Care | McCall Mobile Medicine (feat. Drew Holliday, MSW)	45	188
5/4/2022	Episode 12: Deborah Seltzer Talks Coding and Billing for Substance Use Disorders (feat. Deborah Seltzer)	51	122
6/13/2022	Episode 13: SUD Treatment for Justice-Involved Patients | Day One Program (feat. Radha Sadacharan, MD, MPH and Rebecca Lee)	60	149

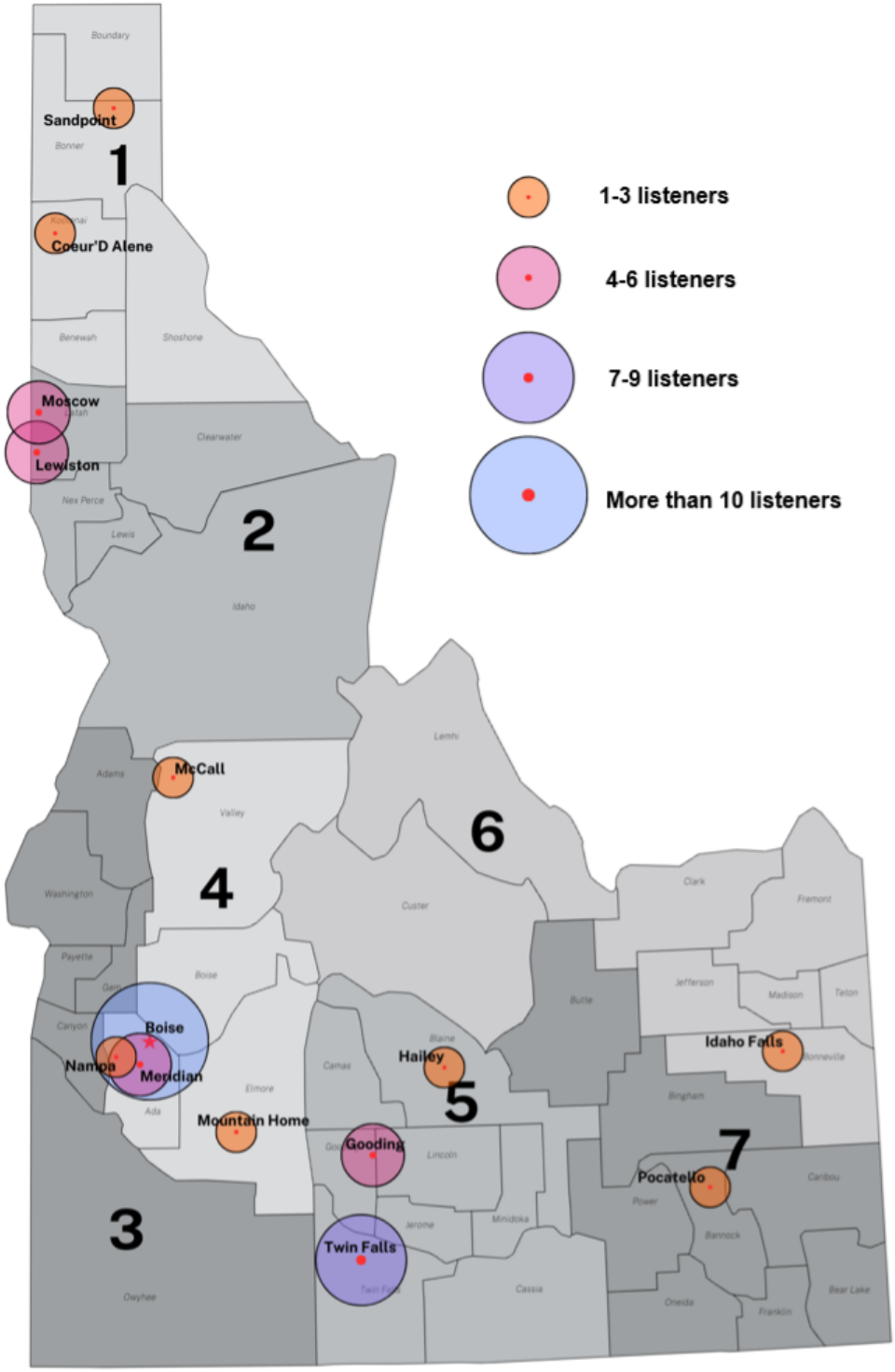
[a]Not available.

XSL•FO
RenderX

CE Credit

A total of 63 unique health care professionals representing 14 distinct cities and spanning all 7 public health districts within the state of Idaho, as well as out-of-state listeners, claimed CE credit for season 1 of the podcast (Figure 1). The podcast drew a diverse array of professionals from various disciplines, with social workers being the most widely represented group among those who claimed CE credits (Multimedia Appendix 1). Three questions were generated for each of the 13 episodes eligible for CE credits. The average percentage of correct answers for all episodes was 83% (SD 11.6%) (Multimedia Appendix 2), thus providing preliminary evidence related to audience engagement and potential impact on learning outcomes.

Figure 1. Health care professionals claiming continuing education credits from the ECHO Idaho "Something for the Pain" podcast within the state of Idaho. Values 1-7 represent Idaho's 7 public health districts (map adapted from MapChart).



XSL•FO
RenderX

Discussion

Principal Findings

ECHO Idaho's "Something for the Pain" podcast included 14 episodes providing best practices and resources for behavioral health, opioid use disorder, and SUD prevention, treatment, and recovery specific to Idaho. The podcast included interviews with health care professionals and community members who had first-hand experience and knowledge of evidence-based practices, as well as didactic presentations taken from previously recorded ECHO Idaho sessions. Audience recruitment primarily focused on health care providers and community members interested in the behavioral health and SUD field. Metrics such as click metrics, CE credits, and user experience questions were used to track listener engagement. Data collected from Eeds and iECHO systems were analyzed using SPSS, providing a descriptive statistical analysis.

The ECHO Idaho podcast effectively engaged a diverse group of health care professionals throughout Idaho, demonstrating the utility of podcasts as a versatile tool for professional development and outreach. This highlights the possibility of leveraging podcasts to not only provide CE but also to serve as a means of attracting more health care professionals to Project ECHO sessions. The widespread accessibility of podcasts suggests their potential as a potent recruitment tool for future Project ECHO initiatives.

This report presents a preliminary assessment of an innovative platform to engage with health care professionals, as well as community members, by providing insightful information and professional development material via the medium of podcasts.

Limitations

Although the initial season of the podcast engaged with diverse health care providers across the state, the podcast's impact on clinical practice remains unknown. Previous literature has suggested internet-based continuing medical education can increase knowledge, change physician behavior, and indirectly impact clinical practice [16,24]. Therefore, future research should collect feedback from nonlisteners and previous listeners to understand their preferences for this innovative platform as well as assess the direct impact of podcast-based CE on listeners' knowledge and professional and clinical practice through quantitative and qualitative analyses.

Additionally, although several data analytics methods were used, some challenges remain. The RSS data records clicks (ie, listens) on each episode, but there is no way to track how long someone listened or if they completed the full episode. Therefore, interpretation of those data points should be made with caution. Similarly, due to data constraints, it is difficult to assess the benefits of the ECHO podcast compared to the more traditional ECHO program model. As a result, we were unable to fully evaluate the podcast's impact, particularly in terms of reach, impact on learning outcomes, and its influence on clinical practice. Future research should focus on assessing these variables, as they are crucial for understanding the effectiveness of this type of CE material. Lastly, the effectiveness of our recruitment strategies remains undetermined. Future research should prioritize evaluating these strategies to optimize the return on investment and enhance the overall impact of the podcast.

Lessons Learned

Overall, the goal of the podcast was to provide clinicians across the state with an easily accessible CE program. We learned that podcasts could be used to engage Idaho health care providers and provide access to CE opportunities. Although we provide evidence of successfully reaching listeners, a robust marketing campaign was not used. Furthermore, the topics were selected by a team of subject matter experts without input from listeners or previous ECHO Idaho attendees. In the future, assessing additional indicators (eg, listener preferences related to episode length, topic selection, impact on clinical practice, perceived barriers to listening to podcasts) could ensure that new podcast series and episodes meet the needs of current and potential listeners.

Conclusions

The ECHO Idaho "Something for the Pain" podcast was developed as a new and innovative way of providing CE for health care providers. We provide evidence that this approach was successful in its efforts to engage a sizable audience of listeners who claimed CE credits for participation. The podcast had listeners from diverse health care professions representing cities from across Idaho and the United States. Future research efforts should include the collection of additional information such as listener preferences, knowledge change, professional and clinical impact, and patient outcomes to guide the implementation of Project ECHO information into an effective CE podcast format.

Acknowledgments

ECHO (Extension for Community Health Outcomes) Idaho's "Something for the Pain" (episodes 1-13) podcast was created by ECHO Idaho and the University of Idaho, supported by the WWAMI Medical Education Program, in partnership with the Valley County Opioid Response Project. ECHO Idaho's "Something for the Pain" (episodes 1-13) podcast was made possible by a grant from the US Department of Health and Human Services Health Resources and Services Administration (HRSA; grant GA1RH39585). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of Central District Health or HRSA. The authors would like to thank Samuel Steffen for his work with podcast development and implementation.

XSL•FO
RenderX

Authors' Contributions

RW was involved in the concept and design, data analysis, data interpretation, manuscript writing, and approval of the final manuscript. MPC was involved in the concept and design, data interpretation, manuscript writing, and approval of the final manuscript. JDM, SMD, and RTB were involved in the concept and design, manuscript writing, and approval of the final manuscript. All authors agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Unique health care professionals claiming continuing education credit from the ECHO Idaho "Something for the Pain" podcast.
[DOCX File, 22 KB - mededu_v11i1e55313_app1.docx]

Multimedia Appendix 2
Percentage of correct answers for questions of continuing education credit–eligible episodes.
[DOCX File, 22 KB - mededu_v11i1e55313_app2.docx]

References

1. Komaromy M, Duhigg D, Metcalf A, et al. Project ECHO (Extension for Community Healthcare Outcomes): a new model for educating primary care providers about treatment of substance use disorders. Subst Abus 2016;37(1):20-24. [doi: 10.1080/08897077.2015.1129388] [Medline: 26848803]
2. Baker RT, Casanova MP, Whitlock JN, Smith LH, Seegmiller JG. Expanding access to health care: evaluating project Extension for Community Health Care Outcomes (ECHO) Idaho's tele-education behavioral health program. Journal of Rural Mental Health 2020;44(4):205-216. [doi: 10.1037/rmh0000157]
3. Katzman JG, Qualls CR, Satterfield WA, et al. Army and Navy ECHO pain telementoring improves clinician opioid prescribing for military patients: an observational cohort study. J Gen Intern Med 2019 Mar;34(3):387-395. [doi: 10.1007/s11606-018-4710-5] [Medline: 30382471]
4. Shimasaki S, Bishop E, Guthrie M, Thomas JFF. Strengthening the health workforce through the ECHO stages of participation: participants' perspectives on key facilitators and barriers. J Med Educ Curric Dev 2019;6:2382120518820922. [doi: 10.1177/2382120518820922] [Medline: 30729170]
5. Salvador J, Bhatt S, Fowler R, et al. Engagement with Project ECHO to increase medication-assisted treatment in rural primary care. Psychiatr Serv 2019 Dec 1;70(12):1157-1160. [doi: 10.1176/appi.ps.201900142] [Medline: 31434561]
6. Dubin RE, Flannery J, Taenzer P, et al. ECHO Ontario chronic pain & opioid stewardship: providing access and building capacity for primary care providers in underserviced, rural, and remote communities. Stud Health Technol Inform 2015;209:15-22. [doi: 10.3233/978-1-61499-505-0-15] [Medline: 25980700]
7. Katzman JG, Comerci G Jr, Boyle JF, et al. Innovative telementoring for pain management: project ECHO pain. J Contin Educ Health Prof 2014;34(1):68-75. [doi: 10.1002/chp.21210] [Medline: 24648365]
8. Mazurek MO, Brown R, Curran A, Sohl K. ECHO Autism. Clin Pediatr (Phila) 2017 Mar;56(3):247-256. [doi: 10.1177/0009922816648288] [Medline: 27169714]
9. Arora S, Kalishman S, Thornton K, et al. Expanding access to hepatitis C virus treatment--Extension for Community Healthcare Outcomes (ECHO) project: disruptive innovation in specialty care. Hepatology 2010 Sep;52(3):1124-1133. [doi: 10.1002/hep.23802] [Medline: 20607688]
10. Agley J, Delong J, Janota A, Carson A, Roberts J, Maupome G. Reflections on project ECHO: qualitative findings from five different ECHO programs. Med Educ Online 2021 Dec;26(1):1936435. [doi: 10.1080/10872981.2021.1936435] [Medline: 34076567]
11. Casanova MP, Nelson MC, Blades KC, Smith LH, Seegmiller JG, Baker RT. Evaluation of an opioid and addiction treatment tele-education program for healthcare providers in a rural and frontier state. J Opioid Manag 2022;18(4):297-308. [doi: 10.5055/jom.2022.0725] [Medline: 36052928]
12. Furlan AD, Zhao J, Voth J, et al. Evaluation of an innovative tele-education intervention in chronic pain management for primary care clinicians practicing in underserved areas. J Telemed Telecare 2019 Sep;25(8):484-492. [doi: 10.1177/1357633X18782090] [Medline: 29991316]
13. Shea CM, Gertner AK, Green SL. Barriers and perceived usefulness of an ECHO intervention for office-based buprenorphine treatment for opioid use disorder in North Carolina: a qualitative study. Subst Abus 2021;42(1):54-64. [doi: 10.1080/08897077.2019.1694617] [Medline: 31809679]
14. Rouleau G, Gagnon MP, Côté J, et al. Effects of e-learning in a continuing education context on nursing care: systematic review of systematic qualitative, quantitative, and mixed-studies reviews. J Med Internet Res 2019 Oct 2;21(10):e15118. [doi: 10.2196/15118] [Medline: 31579016]

XSL•FO
RenderX

15. Block J, Lerwick P. 1064: educational preferences among residents in the ICU. Crit Care Med 2019;47(1):509. [doi: 10.1097/01.ccm.0000551809.99859.12]
16. Malecki SL, Quinn KL, Zilbert N, et al. Understanding the use and perceived impact of a medical podcast: qualitative study. JMIR Med Educ 2019 Sep 19;5(2):e12901. [doi: 10.2196/12901] [Medline: 31538949]
17. Purdy E, Thoma B, Bednarczyk J, Migneault D, Sherbino J. The use of free online educational resources by Canadian emergency medicine residents and program directors. CJEM 2015 Mar;17(2):101-106. [doi: 10.1017/cem.2014.73] [Medline: 25927253]
18. Grafton-Clarke C, Uraiby H, Gordon M, et al. Pivot to online learning for adapting or continuing workplace-based clinical learning in medical education following the COVID-19 pandemic: a BEME systematic review: BEME Guide No. 70. Med Teach 2022 Mar;44(3):227-243. [doi: 10.1080/0142159X.2021.1992372] [Medline: 34689692]
19. Chartier LB, Hansen S, Lim D, et al. P023: Code Resus - using a quality improvement approach to improve health care provider response during resuscitations. CJEM 2016 May;18(S1):S86-S86. [doi: 10.1017/cem.2016.199]
20. McCarthy J, Porada K. Serving up Peds Soup: podcast-based paediatric resident education. Med Educ 2020 May;54(5):456-457. [doi: 10.1111/medu.14113] [Medline: 32185820]
21. Reid Burks A, Nicklas D, Owens J, Lockspeiser TM, Soranno D. Urinary tract infections: pediatric primary care curriculum podcast. MedEdPORTAL 2016 Aug 5;12:10434. [doi: 10.15766/mep_2374-8265.10434] [Medline: 31008213]
22. Roth J, Chang A, Ricci B, Hall M, Mehta N. Why not a podcast? Assessing narrative audio and written curricula in obstetrical neurology. J Grad Med Educ 2020 Feb;12(1):86-91. [doi: 10.4300/JGME-D-19-00505.1] [Medline: 32089798]
23. ECHO Idaho. Something for the Pain. SoundCloud. URL: https://soundcloud.com/user-658492948 [accessed 2025-03-05]
24. Fordis M, King JE, Ballantyne CM, et al. Comparison of the instructional efficacy of internet-based CME with live interactive CME workshops: a randomized controlled trial. JAMA 2005 Sep 7;294(9):1043-1051. [doi: 10.1001/jama.294.9.1043] [Medline: 16145024]

Abbreviations

CE: continuing education
ECHO: Extension for Community Health Outcomes
SUD: substance use disorder
WWAMI: Washington, Wyoming, Alaska, Montana, Idaho

Edited by B Lesselroth; submitted 08.12.23; peer-reviewed by J Woods, LE Tomedi, Y Lunsky; revised version received 15.08.24; accepted 25.02.25; published 21.03.25.

Please cite as:
Wiet R, Casanova MP, Moore JD, Deming SM, Baker Jr RT
Creation of the ECHO Idaho Podcast: Tutorial and Pilot Assessment
JMIR Med Educ 2025;11:e55313
URL: https://mededu.jmir.org/2025/1/e55313
doi:10.2196/55313

© Ryan Wiet, Madeline P Casanova, Jonathan D Moore, Sarah M Deming, Russell T Baker Jr. Originally published in JMIR Medical Education (https://mededu.jmir.org), 21.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Enhancing Access to Neuraxial Ultrasound Phantoms for Medical Education of Pediatric Anesthesia Trainees: Tutorial

Leah Webb[1*], MD; Melissa Masaracchia[2*], MD; Kim Strupp[1*], MD

[1]Division of Pediatric Anesthesiology, Department of Anesthesiology, Children's Hospital Colorado, University of Colorado, Denver, CO, United States

[2]Northwell, Division of Pediatric Anesthesiology, Department of Anesthesiology, Cohen Children's Medical Center, Zucker School of Medicine at Hofstra/Northwell, New Hyde Park, NY, United States

[*]all authors contributed equally

Corresponding Author:
Leah Webb, MD
Division of Pediatric Anesthesiology, Department of Anesthesiology, Children's Hospital Colorado, University of Colorado, Denver, CO, United States

Abstract

Opportunities to learn ultrasound-guided/assisted (USGA) neuraxial techniques for pediatric patients are limited, given the inherent high stakes and small margin of error in this population. Simulation is especially valuable in pediatrics because it enhances competency and efficiency, without added risk, when learning new skills, specifically those seen with ultrasound-guided regional anesthetic techniques. However, access to simulation opportunities involving the use of phantom models in medical education is limited due to excessive costs. We describe a process for producing ultrasound phantoms by using synthetic ballistic gelatin; these ultrasound phantoms can be used for simulation and are affordable, reproducible, and indefinitely shelf stable. The ultrasound images produced by these phantoms are comparable to those obtained from a real pediatric patient, including the sacral anatomy necessary for caudal epidural blocks, as validated by practicing pediatric anesthesiologists. Phantom models offer a more cost-effective alternative to commercially prepared phantoms, thereby expanding access to realistic simulations for neuraxial ultrasound in pediatric medical education, without the prohibitively high expense.

(JMIR Med Educ 2025;11:e63682) doi:10.2196/63682

KEYWORDS

anesthesiology; pediatric; ultrasound; education; neuraxial ultrasound; medical education; pediatric anesthesia trainees; anesthesia; trainees; ultrasound-guided; neuraxial techniques; pediatric patients; efficiency

Introduction

Opportunities to learn ultrasound-guided/assisted (USGA) neuraxial techniques for pediatric patients are limited, given the inherent high stakes and small margin of error in this population. Simulation is a valuable and effective method for learners—whether used by trainees or experienced clinicians—to enhance their competency, efficiency, and confidence in performing regional anesthetic and neuraxial techniques [1-4]. Ultrasound enhances safety, decreases complications, and improves the efficacy and accuracy of neuraxial blockade in pediatric patients from preterm to adolescence [5-12]. The utility of ultrasound is even more apparent in syndromic children with unusual anatomy, patients who comprise a large subset of the pediatric population that presents for surgery at a young age [13]. Honing pediatric patient–related ultrasound skills in a simulation setting is an ideal scenario for learning without risk. Unfortunately, educational curricula and teaching models lag behind recent advancements in simulation.

Despite efforts to create affordable and reproducible ultrasound phantoms, many lack a realistic appearance, and most are not indefinitely, if at all, shelf stable or portable because they are made of water, agar, gelatin, or other substances or are derived porcine models [3,4,14,15]. The cost of manufactured models that offer all these features can be prohibitively expensive, amounting to several thousand dollars, and these models may generate an inferior simulation experience [14]. Limited access to high-fidelity ultrasound phantoms significantly restricts opportunities for learners to take advantage of low-stakes simulation training and necessitates practicing on live patients, including infants and children, to learn valuable skills—a method with varying degrees of success and much higher stakes. There is a dearth of literature describing spine phantoms that are made with the necessary anatomy to teach pediatric trainees how to use ultrasound to approach the caudal epidural space. We describe a method for creating a realistic, affordable, reproducible, and shelf-stable spine phantom model that allows for the demonstration of key ultrasound images of the spine and caudal anatomy that are required to perform USGA neuraxial techniques on pediatric patients. Furthermore, the synthetic ballistic gelatin used to produce the phantom model can be reclaimed and reused to make “fresh” models for an indefinite

XSL•FO
RenderX

period of time, allowing for multiple practice sessions without additional costs.

Methods

Overview

We present a tutorial describing the construction of an ultrasound phantom of the spine, based on similar previous descriptions [16]. Notably however, our model includes both lumbar anatomy and sacral anatomy, which are lacking in previously published iterations but are essential for learning pediatric-specific neuraxial sonoanatomy. Additionally, we completely submerged our spine model in ballistics gel, creating stable, flat surfaces surrounding the spine to facilitate scanning the model in multiple orientations, which simulates the use of ultrasound for prone, lateral, and sitting positions. Further, an anonymous survey was sent to 10 practicing attending pediatric anesthesiologists to evaluate the similarity between the ultrasound images generated from the phantom and images from a real patient. Three ultrasound views were evaluated for likeness and accuracy on a 5-point Likert scale.

How to Create a Phantom Model

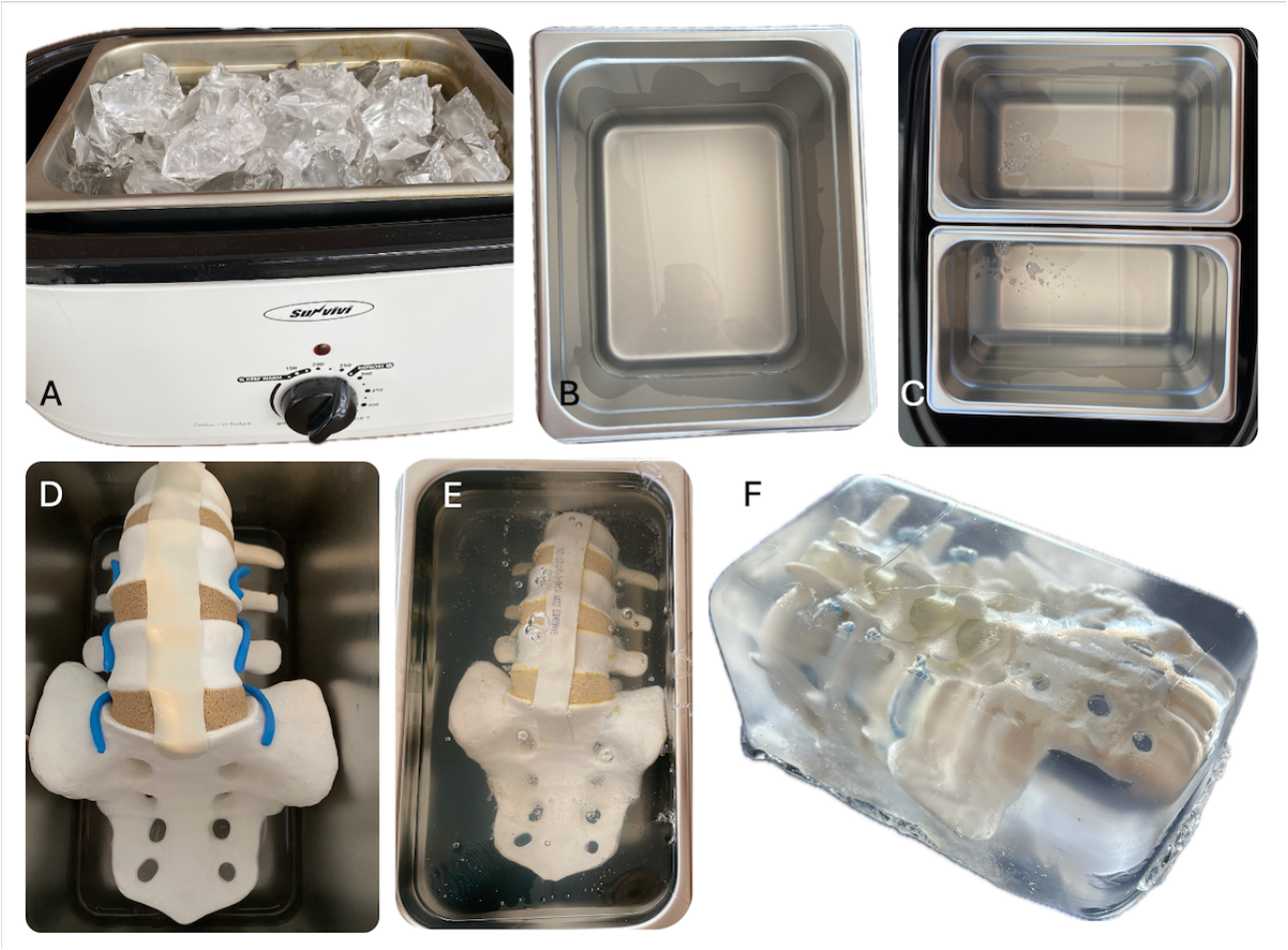
The following stepwise process can be used to create a spine phantom:

- Step 1: Preheat a portable oven to 250-270 °F (121-132 °C). A portable oven is preferable, as it can be used outside to prevent inhalation of the unpleasant smell from melting gel. It is critical to review the manufacturer's guidelines; ensure that the ballistics gel is always melted in well-ventilated areas; and ensure that caution is used to avoid overheating the gel, as it could light on fire.
- Step 2: Cut or tear ballistics gel into smaller pieces for melting.
- Step 3: Place the gel into an oven-safe pan (either a mold pan or an extra container), with the goal of melting the gel to create a 1- to 3-inch gel layer at the bottom of the pan.
    - This layer mimics the soft tissue covering the spinous processes. Add more gel to create a thicker layer, if desiring to create a model with greater depth to the epidural space.
    - Generally, it is preferable to melt the gel in an extra oven-safe container and pour it into a mold pan for each subsequent step; however, for the first layer, the gel can preferentially be melted directly in the mold pan.
- Step 4: Melt the gel in the oven until all bubbles are gone. This step takes about 2 to 4 hours, depending upon the amount of gel melted. It is critical to minimize bubbles in this layer, as this is the surface that will be scanned with ultrasound.
- Step 5: Allow the bubble-free layer to cool significantly (approximately 30-60 min). Then, place the spine model into the pan, with spinous processes facing down toward the bottom of pan and touching the gel, and press it very gently into the gel. Hold or secure the spine model in place until the gel sets and the model is not moving in the pan (10-20 min).
    - During the first cooling period, the gel should be cool enough to touch and be starting to firm up, with some resistance to pressure from a fingertip, but it should be soft enough to envelop the tips of the spine model's spinous processes.
    - If free-pouring gel into a mold pan (rather than melting gel directly in it, as is preferable), aim to pour toward one side/corner of the pan to minimize air bubbles. A ladle can be used to pour gel into the mold pan, again aiming to pour into one side/corner of the pan (rather than fanning) to minimize bubbles.
    - If many large bubbles are present, consider placing the mold pan back into the oven and cooking further until bubbles are gone (a few small bubbles are generally not problematic).
    - Hot gel will soften the spine model and result in curved spinous processes. Therefore, press only hard enough on the model to make slight contact between the tips of spinous processes and the gel; this contact secures the model to hold it in place in future steps.
- Step 6: Allow gel in mold pan (containing 1- to 3-inch gel layer) and spine model to cool completely.
- Step 7: Once cool, pour more melted gel (see step 3 for melting instructions) into the mold pan until the spine model is completely covered.
    - Pour quickly and to one corner or side at the coccyx-end of the model.
    - Bubbles in this step are not as concerning because this surface will be placed facing down and will not be scanned.
    - Bubbles will continue to rise to the top for several minutes as the gel cools; large bubbles can be popped/opened to create a flatter surface on this side of the phantom, though it is not necessary to do so.
- Step 8: Allow the mold pan (which should now contain the gel-covered spine model) to cool completely, preferably overnight, until the gel is solid.
- Step 9: Remove phantom from pan, using firm but gentle traction on the gel.
    - It may help to run an offset spatula (or another flat, thin tool, such as a butter knife) along the edges of the phantom and pan to help separate the phantom from the mold pan.
    - Once loosened, it can be helpful to stand the pan upright on the short side and slide fingers between the gel and pan as deep as possible to fully free the top side of the gel. Then, firmly push down, while continuously pulling out, on the gel until it releases from the pan.
- Step 10: Store phantom at room temperature, with spinous process side up. To clean, use water and a lint-free towel.

Figure 1 shows correlated pictures of the stepwise process and final phantom model.

XSL•FO
RenderX

Figure 1. Stages of phantom production, with A to C showing steps 1 to 4, D showing steps 5 and 6, E showing steps 7 and 8, and F showing step 9. (A): Cut gel in an extra container placed inside a portable oven set to 270 °F (132 °C). (B): Melted, bubble-free gel in the extra container. (C): Two mold pans. (D): Spine model, with anterior side up, placed in cooled, bubble-free layer. (E): Spine model submerged completely in gel and cooled. (F): Completed spine phantom that has cooled completely and has been removed from the mold pan.



Materials

Multiple options exist for the materials that are used to create a phantom model for practicing neuraxial ultrasound skills. Table 1 outlines those used by the authors, along with purchase sites and prices. Each phantom is composed of a spine model embedded in ballistics gel. Additional necessary items are reusable for multiple production cycles. Supplies include an oven that can sustain 250-270 °F (121-132 °C; US $119 for a portable oven), an oven-safe mold (US $9), and an extra oven-safe container (US $11). Optional items include a ladle or another heatproof tool for scooping melted gel. The ladle can be useful for more precision in transferring the gel into the pan. It does potentially create more bubbles than pouring directly; however, bubbles are mitigated by placing the pan back into the oven. The ladle is also useful for ensuring that the anterior side of the model is completely covered with gel and that any bubbles remaining on the anterior side do not interfere with ultrasound scanning. Furthermore, because the spine model is completely submerged in ballistics gel, an offset spatula may be helpful for loosening and releasing the phantom from its mold.

XSL•FO
RenderX

Table . List of materials, where to purchase, costs, and notes on pertinent information.

Item and description		Purchase site	Cost	Notes
Oven (portable)				
	"Sunvivi 22-Quart Roaster Oven"	Amazon.com (ASIN[a]: B07K25WBZ4)	US $119	Any oven that can sustain 250 - 270 °F (121-132 °C).
Ballistics gel				
	"10% FBI Gel Block"	Clearballistics.com (SKU[b]: 852844007000)	US $76 + shipping	Makes ≥4 phantoms.
Spine model				
	"Spine, Lumbar Vertebrae with Nerve Roots and Ligamenta Flava, L3-Sacrum, Solid Foam"	Sawbones.com (SKU: 1340-1)	US $161 + shipping	Preferred.
	"Medical Human Lumbar Spine Demonstration Model Anatomical Model Lumbar Vertebrae Sacrum & Coccyx, with Herniation Disc,for Science Classroom Study Display Teaching Medical Model 15 Inch Hight"	Amazon.com (ASIN: B074JCS4SC)	US $34	Less expensive. Requires removal of some vertebrae to fit recommended oven-safe mold. Alternative option is 3D printed model.
Oven-safe mold				
	"1/4 size 6" Deep Steam Table Pan"	Webstaurantstore.com (item number: 4070469	US $9	Any oven-safe receptacle that is similar in size to spine model. Can be purchased from local restaurant supply store.
Extra oven-safe container				
	"1/2 Size 6" Deep Steam Table Pan"	Webstaurantstore.com (item number: 4070269)	US $11	Used to melt bigger volume of gel.
Gel dye				
	"Tone dye"	Humimic Medical [17]	US $35	Used to opacify gel; comes in a variety of skin tone colors.

[a]Amazon Standard Identification Number.

[b]Stock keeping unit.

Cost

The cost for our preferred spine model is approximately US $161, but a more cost-effective version with fewer vertebrae can be purchased for US $34. The more expensive model is preferred due to the ease of placement in the mold, image quality on ultrasound scans, and representation of more neuraxial structures (spinal nerves and ligamentum flavum). Newer 3D printing technology allows for the printing of customizable and cost-effective pediatric spine models that could alternatively be used in our phantom. Ballistics gel priced at US $76 allows for the production of 4 or more phantoms. The additional items previously mentioned can amount to a cost between US $139 and US $150.

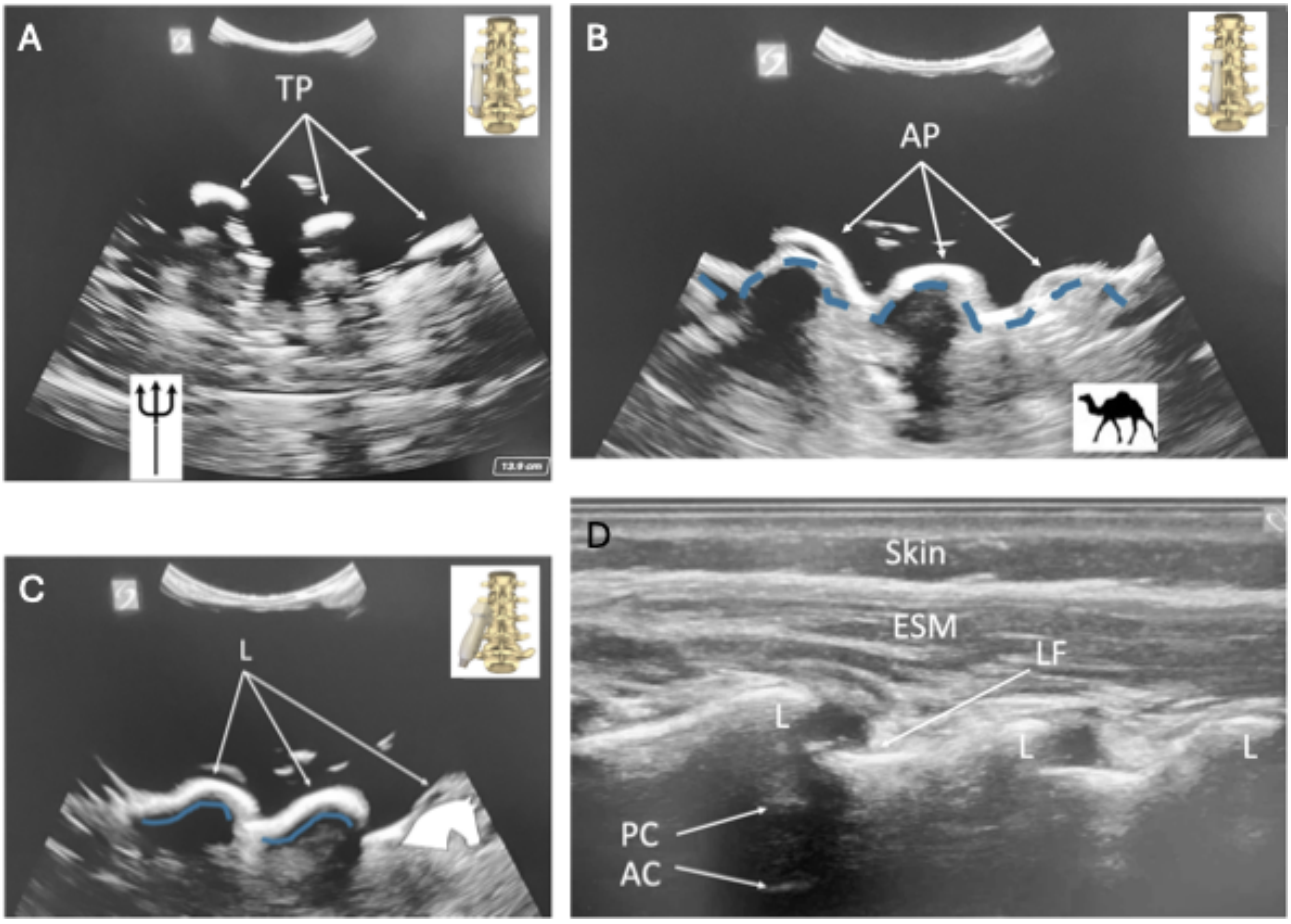
Results

There are 6 views that are critical to performing USGA neuraxial procedures; each is easily obtained from the ultrasound phantom:

- Parasagittal views (Figure 2): transverse process ("trident" sign), articular process ("camel hump" sign), and oblique interlaminar ("horse head" or "sawtooth" sign) views
- Transverse midline views (Figure 3): spinous process, interspinous process/interlaminar ("bat" or "flying bat" sign), and sacral cornua ("frog" or "frog eye" sign) views

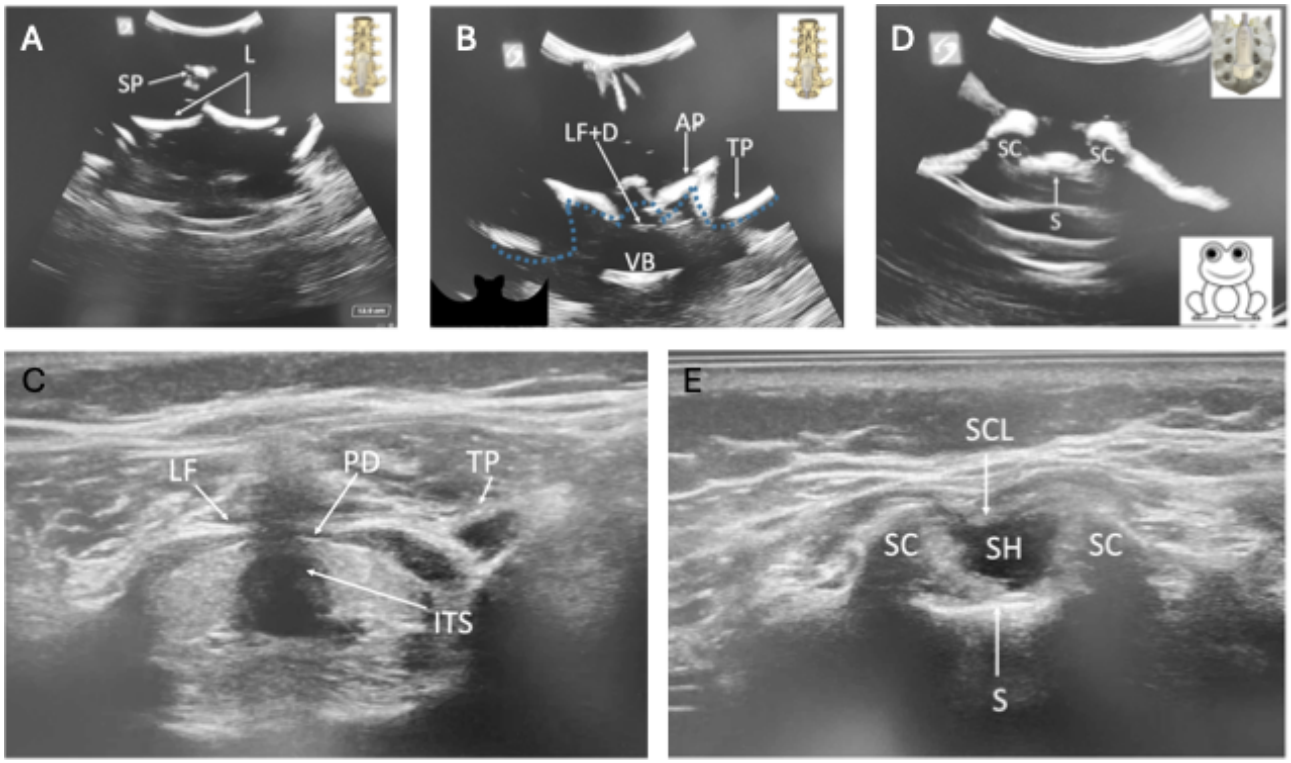
XSL•FO
RenderX

Figure 2. Parasagittal images from phantom (**A, B, and C**), with probe placement relevant to bony anatomy and ultrasound indicator oriented cephalad, and patient (**D**), with ultrasound indicator oriented caudad. (A): Parasagittal TP view ("trident" sign). (B): Parasagittal AP view ("camel hump" sign; dashed blue line shows "camel hump" outline). (C): Parasagittal oblique interlaminar view ("horse head" or "sawtooth" sign; blue line shows "horse head" outline) in phantom. (D): Parasagittal oblique interlaminar view in patient. AC: anterior complex (interface of anterior dura and vertebral body); AP: articular process; ESM: erector spinae muscle; L: lamina; LF: ligamentum flavum; PC: posterior complex (interface of ligamentum flavum and posterior dura); TP: transverse process.



XSL•FO
RenderX

Figure 3. Transverse images from phantom (**A, B, and D**), with probe placement relevant to bony anatomy, and patient (**C and E**); ultrasound probe indicator is oriented left in all images. (A): Transverse midline SP view from phantom. (B): Transverse interspinous (interlaminar) view ("bat" or "bat wing" sign; dotted line shows "bat wing" outline) from phantom. (C) Transverse interspinous (interlaminar) view from patient. (D): Transverse SC view ("frog" or "frog eye" sign) from phantom. (E): Transverse SC view from patient. AP: articular process; ITS: intrathecal space; L: lamina; LF: ligamentum flavum; LF+D: ligamentum flavum+dura (ie, posterior complex); PD: posterior dura; S: sacrum; SC: sacral cornua; SCL: sacrococcygeal ligament; SH: sacral hiatus; SP: spinous process; TP: transverse process; VB: vertebral body.



Each of these views is demonstrated in Figures 2 and 3, with an additional image of the ultrasound probe placement in relation to the bony landmarks of the lumbosacral spine and patient sonoanatomy, where available. Several spine images from a 6-month-old male infant (written consent obtained from parent) are shown (Figures 2D, 3C, and 3E) beside the phantom images for comparison. Because the spine model lacks certain elements, not all structures appear on the phantom scan, and some cannot be obtained.

Phantom images were evaluated for likeness and accuracy via comparison to actual patient images by 10 practicing attending anesthesiologists. Each image was graded on a 5-point Likert scale for how similar it appeared to the actual patient image. All 10 respondents agreed or strongly agreed that the transverse sacral cornua view ("frog" sign) and parasagittal oblique interlaminar view ("horse head" sign) were similar to those of real patients. Of the 10 respondents, 8 agreed or strongly agreed that the transverse interspinous view ("bat wing" sign) was similar between the phantom and real patient images, 1 respondent was neutral, and 1 respondent somewhat disagreed.

Discussion

Unlike previous phantoms described by Morrow et al [16], Mashari et al [14], and others, our spine phantom generates ultrasound images and views that closely replicate the sonoanatomy of a pediatric patient (Figures 2D, 3C, and 3E). A key advancement in our design is the incorporation of the sacrum and sacral hiatus—critical structures needed for visualizing the caudal space, which is a technique that is often used in accessing the neuraxis in pediatric patients. Furthermore, by fully submersing the spine model in ballistics gel, our phantom offers superior stability during scanning and allows for repositioning to simulate sitting, lateral, and prone patient orientations. The enhanced design ensures a more realistic training experience, thereby helping practitioners develop the precise skills necessary for pediatric neuraxial techniques.

Practicing pediatric anesthesiologists overall found our phantom's ultrasound images comparable to ultrasound images of real pediatric anatomy, particularly for the transverse sacral cornua ("frog" sign) and parasagittal oblique interlaminar ("horse head" sign) views. However, while responses for the transverse interspinous (interlaminar) view ("bat wing" sign) were generally positive, some noted minor discrepancies between the phantom images and those of real patients. Given that the phantom lacked several ligaments and the spinal canal seen in real patients, this feedback provides an opportunity for improvement in future phantom models, which could be addressed by the techniques described by Morrow et al [16] (spinal canal) and Mashari et al [14] (ligaments).

Ultrasound has been used to identify anatomical landmarks for epidural or spinal neuraxial procedures and to identify placement of catheters that are inserted in the caudal space and threaded to the lumbar or thoracic space in pediatric patients [5,6,9-12]. The creation of ultrasound phantoms, as described in this paper, can increase access to ultrasound simulation and enhance opportunities for learning critical procedural skills in a

XSL•FO
RenderX

low-stakes environment [1-4,14,16]. To meet these needs, we created a phantom that is indefinitely shelf stable, reproducible, and cost-effective (approximately US $92 to US $219 per phantom, including the materials listed plus the reusable materials). By modifying the previous technique described by Morrow et al [16], our phantom was specifically designed to image the sacral cornua and to easily scan in the prone or lateral positions, which are essential features for training anesthesia clinicians in pediatric neuraxial sonoanatomy.

The use of spine phantoms was previously limited by their costs; however, budget-friendly spine phantoms created with readily available materials produce a realistic feel when palpating for anatomic landmarks [14] and generate many of the views required to perform neuraxial USGA procedures [14,16]. These phantoms also replicate sonoanatomy with high fidelity, as demonstrated by Mashari et al [14], who actually found that their low-cost model resulted in superior fidelity for ultrasound imaging when compared to an expensive, commercially available task trainer.

There are some limitations to the phantom described herein, of which many can be attributed to the absence of more complex anatomical structures. Although some views and sonoanatomy cannot be identified without these structures, easy solutions are available if needed. For example, our model has a fused sacrum, as is common in adults; therefore, scanning of the sacrum in the sagittal plane—a useful technique for performing in-plane USGA caudal epidural blocks in infants and children—is futile. This problem can be relieved by obtaining an anatomically correct spine model that is reflective of infants or young children, either through purchase or through 3D printing [14,18]. Models of pediatric spines with both normal anatomy and abnormal anatomy could be made via 3D printing, enhancing the pediatric-specific simulation experience; however, access can be limited and may be costly when considering the initial monetary investment in a 3D printer.

The phantom described also lacks contents of the spinal canal, rendering it inadequate for simulating access to the intrathecal space for spinal blockade. Inserting fluid-filled tubing into the empty spinal canal (a technique described by Morrow et al [16]) prior to pouring melted gel on the model could provide a potential solution. However, while this added feature can present itself as another useful learning tool, we found that needling the phantom degrades the image quality over time and should be considered when deciding whether to include a spinal canal in future models. Other potential options for making a more complete model include adding a ligamentum flavum by using silicone paste [14]. This technique may be useful for creating a sacrococcygeal ligament, which is an important landmark when performing USGA caudal blocks while using the transverse sacral cornua view (“frog” sign).

Of further note, we chose to use clear ballistics gel for our phantoms, since it allows for the direct visualization of spine model structures, which can be very helpful for early learners but is not realistic or comparable to scanning live patients. Products used to opacify the gel can be purchased on the internet (Table 1) if a more realistic option is desired.

Future directions for the use of our spine phantom model center on teaching critical skills and assessing knowledge of and comfort with high-stakes procedures in novice trainees. Further evaluation of our phantom should focus on the effectiveness of the phantom as a teaching tool.

By constructing a reproducible, affordable, and shelf-stable spine phantom that can be scanned to generate images and sonoanatomy of the infant and child neuraxis, trainees can be provided with a low-stakes environment in which they can learn how to perform high-stakes regional anesthesia blocks. By addressing the limitations of previous models, our phantom provides an affordable, high-fidelity tool that enhances access to realistic neuraxial ultrasound training for pediatric trainees.

Acknowledgments

Funding for the materials used to create the phantoms was provided by the University of Colorado Department of Anesthesiology. The funder was not involved in the study design; in the collection, analysis, and interpretation of the data; in the writing of the report; or in the decision to submit the paper for publication.

Conflicts of Interest

None declared.

References

1. Moore DL, Ding L, Sadhasivam S. Novel real-time feedback and integrated simulation model for teaching and evaluating ultrasound-guided regional anesthesia skills in pediatric anesthesia trainees. Paediatr Anaesth 2012 Sep;22(9):847-853. [doi: 10.1111/j.1460-9592.2012.03888.x] [Medline: 22612411]
2. Chen XX, Trivedi V, AlSaflan AA, et al. Ultrasound-guided regional anesthesia simulation training: a systematic review. Reg Anesth Pain Med 2017;42(6):741-750. [doi: 10.1097/AAP.0000000000000639] [Medline: 28759501]
3. Karmakar MK. The “Water-based Spine Phantom”? - A small step towards learning the basics of spinal sonography. Br J Anaesth 2009 Mar 12;103(eLetters Supplement). [doi: 10.1093/bja/el_4114]
4. Li JW, Karmakar MK, Li X, Kwok WH, Ngan Kee WD. Gelatin-agar lumbosacral spine phantom: a simple model for learning the basic skills required to perform real-time sonographically guided central neuraxial blocks. J Ultrasound Med 2011 Feb;30(2):263-272. [doi: 10.7863/jum.2011.30.2.263] [Medline: 21266566]

XSL•FO
RenderX

5. Boretsky KR, Camelo C, Waisel DB, et al. Confirmation of success rate of landmark-based caudal blockade in children using ultrasound: a prospective analysis. Paediatr Anaesth 2020 Jun;30(6):671-675. [doi: 10.1111/pan.13865] [Medline: 32267040]
6. Cristiani F, Henderson R, Lauber C, Boretsky K. Success of bedside ultrasound to identify puncture site for spinal anesthesia in neonates and infants. Reg Anesth Pain Med 2019 Jul 3:rapm-2019-100672. [doi: 10.1136/rapm-2019-100672] [Medline: 31273065]
7. Sidiropoulou T, Christodoulaki K, Siristatidis C. Pre-procedural lumbar neuraxial ultrasound-a systematic review of randomized controlled trials and meta-analysis. Healthcare (Basel) 2021 Apr 17;9(4):479. [doi: 10.3390/healthcare9040479] [Medline: 33920621]
8. Rubin K, Sullivan D, Sadhasivam S. Are peripheral and neuraxial blocks with ultrasound guidance more effective and safe in children? Paediatr Anaesth 2009 Feb;19(2):92-96. [doi: 10.1111/j.1460-9592.2008.02918.x] [Medline: 19207895]
9. Kil HK, Cho JE, Kim WO, Koo BN, Han SW, Kim JY. Prepuncture ultrasound-measured distance: an accurate reflection of epidural depth in infants and small children. Reg Anesth Pain Med 2007;32(2):102-106. [doi: 10.1016/j.rapm.2006.10.005] [Medline: 17350519]
10. Kil HK. Caudal and epidural blocks in infants and small children: historical perspective and ultrasound-guided approaches. Korean J Anesthesiol 2018 Dec;71(6):430-439. [doi: 10.4097/kja.d.18.00109] [Medline: 30086609]
11. Tsui BCH, Suresh S. Ultrasound imaging for regional anesthesia in infants, children, and adolescents: a review of current literature and its application in the practice of neuraxial blocks. Anesthesiology 2010 Mar;112(3):719-728. [doi: 10.1097/ALN.0b013e3181c5e03a] [Medline: 20179511]
12. Ponde VC, Bedekar VV, Desai AP, Puranik KA. Does ultrasound guidance add accuracy to continuous caudal-epidural catheter placements in neonates and infants? Paediatr Anaesth 2017 Oct;27(10):1010-1014. [doi: 10.1111/pan.13212] [Medline: 28795472]
13. Park SK, Bae J, Yoo S, et al. Ultrasound-assisted versus landmark-guided spinal anesthesia in patients with abnormal spinal anatomy: a randomized controlled trial. Anesth Analg 2020 Mar;130(3):787-795. [doi: 10.1213/ANE.0000000000004600] [Medline: 31880632]
14. Mashari A, Montealegre-Gallegos M, Jeganathan J, et al. Low-cost three-dimensional printed phantom for neuraxial anesthesia training: development and comparison to a commercial model. PLoS One 2018 Jun 18;13(6):e0191664. [doi: 10.1371/journal.pone.0191664] [Medline: 29912877]
15. Cole JH, Fishback JE, Hughey SB. Cadaveric porcine spines as a model for the human epidural space. Comp Med 2019 Aug 1;69(4):308-310. [doi: 10.30802/AALAS-CM-18-000133] [Medline: 31340882]
16. Morrow DS, Cupp JA, Broder JS. Versatile, reusable, and inexpensive ultrasound phantom procedural trainers. J Ultrasound Med 2016 Apr;35(4):831-841. [doi: 10.7863/ultra.15.04085] [Medline: 26969595]
17. Dye. Humimic Medical. URL: https://humimic.com/product-category/molds-and-accesories/dye/ [accessed 2025-04-25]
18. Marsh-Armstrong BP, Ryan JF, Mariano DJ, Suresh PJ, Supat B. Building affordable, durable, medium-fidelity ballistic gel phantoms for ultrasound-guided nerve block training. J Vis Exp 2024 Feb 9(204). [doi: 10.3791/66194] [Medline: 38407270]

Abbreviations

USGA: ultrasound-guided/assisted

Edited by B Lesselroth; submitted 26.06.24; peer-reviewed by JG Bailey, S Narayanasamy; revised version received 28.02.25; accepted 19.03.25; published 12.05.25.

Please cite as:
Webb L, Masaracchia M, Strupp K
Enhancing Access to Neuraxial Ultrasound Phantoms for Medical Education of Pediatric Anesthesia Trainees: Tutorial
JMIR Med Educ 2025;11:e63682
URL: https://mededu.jmir.org/2025/1/e63682
doi:10.2196/63682

© Leah Webb, Melissa Masaracchia, Kim Strupp. Originally published in JMIR Medical Education (https://mededu.jmir.org), 12.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Using Web-Based Continuing Education to Improve New Diagnoses of Alzheimer Disease in Claims Data: Retrospective Case-Control Study

Katie Lucero[1], PhD; Thomas Finnegan[1], PhD; Soo Borson[2], MD

[1]Medscape, 283-299 Market Street, 2 Gateway Center - 4th Floor, Newark, NJ, United States

[2]Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Katie Lucero, PhD
Medscape, 283-299 Market Street, 2 Gateway Center - 4th Floor, Newark, NJ, United States

Abstract

Background: Alzheimer disease (AD) presents significant challenges to health care systems worldwide. Early and accurate diagnosis of AD is crucial for effective management and care to enable timely treatment interventions that can preserve cognitive function and improve patient quality of life. However, there are often significant delays in diagnosis. Continuing medical education (CME) has enhanced physician knowledge and confidence in various medical fields, including AD. Notably, web-based CME has been shown to positively influence physician confidence, which can lead to changes in practice and increased adoption of evidence-based treatment selection.

Objective: This study investigated the impact of a targeted, web-based CME intervention on health care providers' confidence, competence, and real-world outcomes in diagnosing early AD.

Methods: The study employed a 2-phase design. Phase I used a pre-post assessment to evaluate immediate changes in knowledge and confidence before and after CME participation. Phase II involved a retrospective, matched case-control study to examine the impact of CME on AD diagnoses in claims data.

Results: A 1-way ANOVA showed a significant effect of CME regarding change in the volume of AD diagnoses (F_{1900}=5.50; P=.02). Compared to controls, CME learners were 1.6 times more likely to diagnose AD, resulting in an estimated net increase of 7939 new diagnoses annually. Post-CME confidence was associated with a greater likelihood of diagnosing AD (odds ratio 1.64; 95% CI 0.92-2.92; P=.09; n=219).

Conclusions: Web-based CME participation is associated with increased real-world AD diagnoses. Findings offer a mechanism to explain the changes in clinical practice seen as a result of the CME intervention, which improves skills and confidence.

(JMIR Med Educ 2025;11:e72000) doi:10.2196/72000

KEYWORDS

Alzheimer disease diagnosis; continuing medical education; real-world outcomes; physician confidence; web-based CME; CME; self-efficacy

Introduction

Alzheimer disease (AD) is a progressive neurodegenerative disorder that poses significant challenges to health care systems worldwide. AD affects more than 6.0 million persons in the United States, 7.9 million in Europe, and at least 50 million people worldwide [1,2]. The risk of AD increases with age. By 2050, the number of affected persons 65 years and older is expected to reach 12.7 million in the United States and over 152 million worldwide [1]. As the global population ages, the prevalence of AD has risen dramatically and will continue to rise, bringing new urgency to addressing the widespread lags in diagnosis that impede effective patient management and care.

Early diagnosis is vital for maintaining quality of life, delaying institutionalization, and improving treatment outcomes. Monoclonal antibodies that target plaque work best in the early stages of AD when pathologic changes are still relatively mild [3,4]. Early diagnosis allows for early initiation of treatment, which can help preserve patients' functional abilities and cognitive function, thereby improving quality of life [5]. Early diagnosis can also reduce caregiver burden by helping patients and caregivers access culturally competent care and support services to improve quality of life [6].

Significant delays are common in the diagnosis and management of patients with AD. Physician practice patterns across several countries, including the United States, reveal that while approximately half of patients globally receive an AD diagnosis

XSL•FO
RenderX

within 6 months of initial presentation, a significant number of patients remain undiagnosed for several months after initially presenting to a physician [7-9]. Misdiagnosis of AD in primary care settings is exceptionally high, with as many as two-thirds of patients being misdiagnosed [6]. While primary care physicians (PCPs) are typically first to see patients with mild cognitive impairment (MCI) and early AD [10], physician suspicion accounts for only 20% of AD diagnoses globally, with caregivers often serving as the primary impetus for seeking medical attention [8]. Overall, referral rates for specialist care are also low (14% - 23%) [8].

Recent updates to diagnostic and staging criteria for AD are based on biological indicators versus clinical syndrome [11]. In practice, AD is often diagnosed through the evaluation of cognitive symptoms, which is highly dependent on a clinician's experience and skill [6]. In the absence of a single diagnostic test for AD, physicians rely on physical and neurological examination, mental status tests, imaging, and biomarkers for diagnostic purposes [12]. However, by the time a patient starts showing signs of cognitive impairment, underlying pathologic changes have likely been happening for a decade or longer [13].

Gaps in physician knowledge and insufficient specialized training partly drive challenges in the diagnostic process [12]. Physicians often struggle to distinguish normal aging from dementia, and between various types of dementia [14], and demonstrate limited awareness of early cognitive impairment indicators [8]. Notably, many physicians lack self-efficacy in diagnostic abilities, including their skills to detect signs of MCI and differentiate MCI from AD [12]. Physicians also lack self-efficacy to use and interpret cognitive testing and neuroimaging [5]. While specialists are more likely to use magnetic resonance imaging, PCPs often rely on computed tomography scans, which are less informative [12]. This lack of self-efficacy can lead to delayed diagnoses, hindering timely interventions and optimal patient outcomes.

Continuing medical education (CME), including web-based CME, has shown promise in improving physician knowledge and self-efficacy across various medical domains, including in AD diagnosis [15-17]. However, the relationship between improving knowledge, competence, self-efficacy, and real-world outcomes (RWOs) in AD diagnosis remains understudied [10,16-18]. Improving knowledge does not guarantee its application in practice. Rather, improving self-efficacy is an essential intermediary between knowledge and practice change. Self-efficacy, a motivational construct also known as confidence, empowers physicians to act upon their knowledge and implement learned skills (also known as competence) [19]. However, the relationship is not strictly linear. Improvements in and reinforcement of knowledge and competence can also increase self-efficacy, which in turn influences practice change [20,21]. These relationships suggest that clinicians with a greater sense of self-efficacy following CME activities demonstrate a stronger intention to change their practice, regardless of whether they improved their knowledge [22,23].

This study investigates whether a targeted, web-based CME intervention can improve physicians' self-efficacy, competence, and RWOs in diagnosing early AD. The study addressed the following hypotheses: (1) competency scores for HCPs will increase; (2) the proportion of HCPs who are confident will increase; and (3) the volume of new AD diagnoses will increase for the CME group compared with the matched control group.

Methods

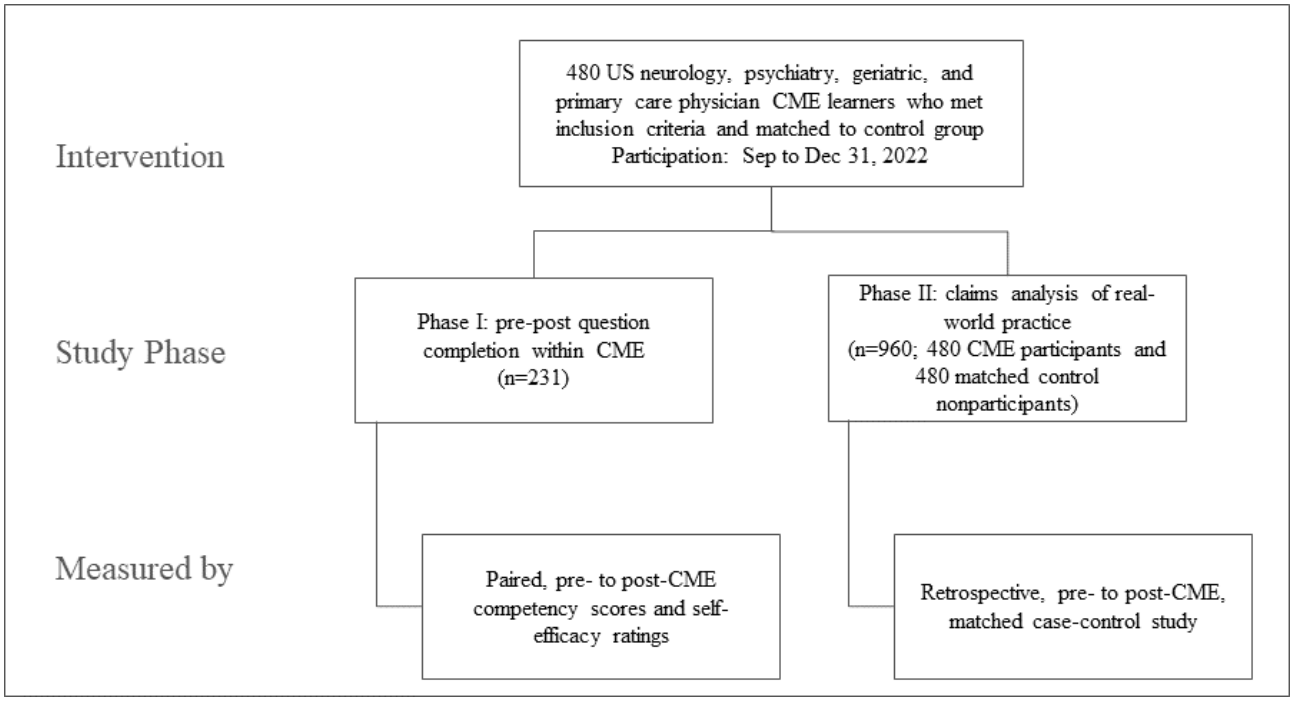
Study Design

Overview

We conducted 2 study phases using the outcomes assessment framework by Moore et al [24] to assess leading indicators (changes in competency scores and confidence ratings) and lagging indicators of success (changes in real-world performance, specifically the volume of new AD diagnoses). Phase I focused on AD diagnosis within CME activities. Phase II focused on AD diagnosis in the real world (Figure 1).

XSL•FO
RenderX

Figure 1. Study design. CME: continuing medical education.



Phase I: Educational Assessment

We employed a paired, pre-post design to assess the impact of CME activities on knowledge and competency scores and confidence ratings immediately before and after the point of learning in the activity for learners from September 13 to December 31, 2022.

Phase II: Real-World Outcomes

We conducted a retrospective, matched case-control study from March 2022 to June 2023 to evaluate the impact of CME activities on diagnosing patients with AD. The intervention period spanned from September 13 to December 31, 2022, with data collection extending from March 2022 to June 2023, assessing practice 6 months before and 6 months after the CME participation date (“index date”).

CME Intervention

The intervention consisted of a web-based CME initiative for PCPs and neurologists designed to improve competence and confidence in early recognition and diagnosis of AD. The first activity focused on best practices in delivering care for patients with AD in primary care and neurology (released September 13, 2022, through September 13, 2023) [25]. The activity was valid for a maximum of 0.50 American Medical Association Physician Recognition Award (AMA PRA) Category 1 Credit. Topics included triaging and assessing patients with cognitive impairment in primary care, coordinating with neurologists, and treatment goals. A subsequent series of 3 simulated online office visits focused on increasing clinicians’ ability to identify and communicate with patients experiencing early cognitive impairment (released October 14, 2022, through October 14, 2023, valid for a maximum of 0.25 AMA PRA Category 1 Credit) [26]. Each office visit centered around an interactive patient-physician vignette for which a decision on screening and evaluation was required. Vignettes included White and Black patients with cognitive impairment due to MCI, cognitive impairment due to early AD, and cognitive impairment not due to dementia. Participants previewed the chief concern for each patient vignette via a landing page with an interactive, graphic table of contents. The activity required participants to investigate cognitive complaints and select diagnostic tests. Faculty feedback was included in each activity.

Inclusion Criteria

Physicians were included if they participated (that is, viewed the content, after the front matter and disclosures) in at least 1 activity in the study period, practiced in the United States, had at least 1 patient who met the inclusion criteria, and had complete claims data available for the study period. Patients were included if they were at least 60 years of age and saw the learner during the study period as evidenced by at least 1 *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* (*ICD-10*) code or prescription by the learner.

Sample

Medscape member registration provides the country of residence, profession, specialty, and National Provider Identifier (NPI; if applicable) number for health care providers (HCPs). A total of 1725 US physicians in the target specialties with a valid NPI were learners between September 13 and December 31, 2022. Of these, 1310 had matches with claims data with at least 1 patient who met the inclusion criteria. The RWO intervention group (RWO learners) comprised 480 physicians who met the full inclusion criteria, participated in the CME activities during the intervention period, and were matched to a nonparticipant control (287/480, 60% PCPs; 100/480, 21% neurologists; 71/480, 15% psychiatrists; and 22/480, 5% geriatric specialists). Of the 480, 231 fully completed an activity.

XSL•FO
RenderX

Matching Process

An equal number of 480 HCPs who did not participate in CME served as the control group. We used a 1:1 matching ratio to pair cases with controls. The matching criteria included: (1) number of patients with AD, (2) number of patients diagnosed with AD by the HCP, (3) profession, (4) specialty, and (5) the first 2 digits of the HCP's ZIP code. Match-It in R was used to match the propensity score by the number of patients with AD and the number of patients diagnosed with AD by the HCP. Exact matching was used for profession, specialty, and the first 2 digits of the ZIP code. An independent samples *t* test showed no statistically significant difference in the number of patients with AD seen in the preperiod by the CME and matched control groups (t=−0.24; P=.81).

Real-World Data Collection

We sourced real-world data from Medscape licensed claims data, accessed in November 2023. This dataset provided comprehensive information on patient visits, diagnoses, and procedures performed by the participating HCPs. Data were aggregated at the patient level, and patient counts were aggregated at the HCP level. See Multimedia Appendix 1 for the codes accepted as indicators of diagnosis.

Measures

Three primary outcome measures assessed the effectiveness of the CME intervention: (1) competency score, (2) confidence rating, and (3) number of patients newly diagnosed with AD by the HCP. Additionally, we examined whether there was an association between being confident following CME and real-world diagnoses.

Competency Score

Competency scores were assessed by asking 3 case-based vignette questions pre- and postpoint of learning in the CME activity. Scores were aggregated at either the learning topic or activity level and then to the HCP. Scores ranged from 0% to 100%. Questions were developed to assess learning against learning objectives related to cognitive assessment or differentiation between MCI and AD. Questions were developed by content experts and reviewed by an expert AD HCP, an outcomes assessment specialist, and a copyeditor.

Self-Efficacy Rating

A question assessed self-efficacy on a 5-point Likert-type scale, with higher scores indicating greater self-efficacy (eg, "How confident are you right now in your ability to assess patients for cognitive impairment?"). Learners were deemed "confident" if they rated themselves as a 4 or 5. We used the term "confident" because it is more easily understood by the respondent than self-efficacy and use "confident" in the methodology and results for ease of interpretation when referring to those who select a 4 or 5.

New AD Diagnoses

New AD diagnoses were assessed by examining patient-level *ICD-10* data. Patients with an AD *ICD-10* code by the HCP of interest in the timeframe of interest were identified. Then, the patient's history 2.5 years prior was examined to assess whether they had previously received any AD *ICD-10* codes by any HCP. If there was no history of receiving an AD *ICD-10* code previously, the patient's *ICD-10* code was considered a new AD diagnosis by the HCP of interest. The count of patients who met these criteria was aggregated at the HCP level.

Statistical Analysis

Phase I: Educational Assessment

We assessed immediate changes in learner competency scores and confidence via 4 matched pair questions before and after CME participation. The McNemar test evaluated change in the proportion of learners who rate themselves as confident. The McNemar test was chosen because it measures differences in paired proportions against the null hypothesis. A paired samples *t* test was conducted to measure mean differences in paired samples. Overall competency changes were assessed using paired samples *t* tests. Statistical significance was set at P<.05 for all tests.

Phase II: Real-World Outcomes

The relationship between CME participation and the postintervention volume of new AD diagnoses was assessed with a 1-way ANOVA. One-way ANOVA was chosen because we wanted to examine the change in volume of AD diagnosis from pre– to post–index date (dependent variable) and whether being a learner was associated with this change (independent variable with 2 independent groups). The dependent variable was the change in the volume of AD diagnoses from pre– to post–index date, and the independent variable was CME participation (versus control). In a secondary analysis, we explored the association between postintervention confidence (confident=1, not confident=0) and AD diagnosis (diagnoser=1, nondiagnoser=0) via logistic regression.

The association between being confident and diagnosing AD was explored because the more we know about mechanisms for change that we can immediately measure at scale within a web-based CME activity, the more effective our education can be. A dichotomous independent variable and dichotomous dependent variable were selected because the research question focused on whether confidence predicts being a diagnoser. More confidence should not equate to diagnosing more because how many patients get diagnosed depends on the types of patients an HCP sees. Previous research used this same dichotomy and found similar results [25].

Statistical significance was set at P<.05 for all tests, and analyses were performed using SAS version 9.4 (SAS Institute).

Ethical Considerations

The Sterling Institutional Review Board deemed this study exempt under the terms of the US Department of Health and Human Service's Policy for Protection of Human Research Subjects at 45 CFR §46.104(d) [27]. The ethical standards of the Declaration of Helsinki were applied to all research procedures. As the study was exempt, there was no requirement for informed consent. The institutional review board approval covered secondary analysis without additional consent. The data were deidentified prior to analysis to safeguard participant information. No compensation was provided to participants.

XSL•FO
RenderX

Results

Phase I: Competency and Confidence

"Completers" answered all linked questions within at least 1 of the CME activities, representing 48% (231/480) of the larger learner population. After participation, RWO learners demonstrated a 34 percentage point increase in correct answers for competency in the diagnosis of AD (33% prescore to 67% postscore; P=.008) and a 16 percentage point pre- to postactivity increase in the proportion of those who were confident in assessing cognitive function and diagnosing AD (75/231 preactivity to 99/231 postactivity; P<.001).

Phase II: New AD Diagnoses

ANOVA showed a significant effect of CME regarding change in the volume of AD diagnoses (F_{1900}=5.50; P=.02). The 6-month postactivity increase in new AD diagnoses was 160% greater for the CME group than the control group, as verified by claims data. RWO learners diagnosed 239 more patients after education (487 diagnoses pre-education vs 726 diagnoses posteducation). Control-group learners diagnosed 91 more patients after education (517 diagnoses pre-education vs 608 diagnoses posteducation). Neurologists had the highest increase in new AD diagnoses (1.58 per neurologist), while psychiatrists had the lowest (0.10 per psychiatrist). The logistic regression model showed a trend within the CME group toward a significant positive relationship between being confident in AD assessment post-CME and diagnosing AD in the real world in the 6 months following CME (odds ratio [OR] 1.64, 95% CI 0.92-2.92; P=.09; n=219). Table 1 summarizes the RWO learners and the matched control group on key outcomes from claims data.

Table . Number of patients with Alzheimer disease (AD) and number of patients newly diagnosed with AD before and 6 months after the activity.

	Patients with AD, mean (SD)		Patients newly diagnosed with AD, mean (SD)	
	Pre	Post	Pre	Post
CME[a] (n=480)	2.16 (7.51)	2.33 (8.35)	1.01 (3.60)	1.51 (5.59)
Geriatric specialists (n=22)	2.27 (5.23)	2.64 (5.95)	1.41 (2.92)	2.05 (4.36)
Neurologists (n=100)	6.50 (15.03)	7.14 (16.79)	3.04 (7.17)	4.62 (11.29)
PCPs[b] (n=287)	1.12 (2.38)	1.14 (2.43)	0.51 (1.14)	0.71 (1.52)
Psychiatrists (n=71)	0.20 (0.60)	0.27 (0.81)	0.10 (0.38)	0.20 (0.62)
Control (n=480)	2.05 (6.27)	1.92 (6.03)	1.08 (3.33)	1.27 (4.07)
Geriatric specialists (n=22)	1.95 (4.36)	1.86 (2.92)	0.82 (1.62)	1.00 (1.69)
Neurologists (n=100)	5.94 (12.11)	5.50 (11.40)	3.24 (6.40)	3.71 (7.46)
PCPs (n=287)	1.15 (2.56)	1.08 (3.03)	0.57 (1.42)	0.68 (2.31)
Psychiatrists (n=71)	0.24 (0.60)	0.31 (1.04)	0.17 (0.48)	0.27 (1.03)

[a]CME: continuing medical education.

[b]PCP: primary care physician.

Discussion

Principal Findings

This matched case-control study examined the impact of a web-based, vignette-based CME on participants' knowledge, competence, self-efficacy, and RWOs in diagnosing early AD. Participation in CME was associated with a significant (P=.02) increase in the diagnosis of early AD. RWO learners were more likely to be diagnosers than control-group physicians, with a magnitude of increase in AD diagnoses that was 1.6 times higher for RWO learners than control-group physicians. The estimated net increase of 7939 in new AD diagnoses in the year following participation for CME learners through the expiration of the activities for credit indicates a substantial positive impact of education on AD diagnosis rates. RWO learners also improved their confidence in identifying early forms of AD (P<.001). When HCPs were confident after CME, they had a 1.64 greater odds of diagnosing AD.

Comparison with Prior Work

Research suggests that CME can effectively improve physician knowledge, self-efficacy, and competence regarding dementia care in general. A large study in Australia evaluated an accredited CME program on the diagnosis and management of dementia in primary care. Participants who completed the program reported feeling significantly more confident in their knowledge, skills, and ability to provide care for people with dementia [15,16]. Our study not only affirms the impact of CME on real-world AD diagnoses but also offers a mechanism to explain the changes in real-world practice seen as a result of the CME intervention. Previous research shows that improvements in knowledge and competence following CME participation are associated with increased self-efficacy, and posteducation self-efficacy mediates the relationship between knowledge and competence and intention to change [20,21]. A recent secondary analysis of knowledge, competency, self-efficacy, and clinical practice using pre- and postparticipation data from web-based CME interventions in 3

XSL•FO
RenderX

different therapeutic areas combined with medical claims data examined the relationship between knowledge, competency, self-efficacy, and real-world clinical practice [23]. Knowledge and competency (P=.08; OR 1.515, 95% CI 0.953-2.410) and self-efficacy (P<.001; OR 2.768, 95% CI 1.705-4.492) were significant predictors of clinical practice. However, the effect size for self-efficacy was larger, suggesting that clinicians confident in their abilities were more likely to utilize evidence-based treatments. These results suggest that self-efficacy plays a significant mediating role in influencing clinical practice.

Reinforcement of existing knowledge also appears to influence clinical practice. A study that examined the relationships among knowledge, competence, self-efficacy, and intention to change across 57 online oncology-certified education programs published from 2018 to 2020 found that both improvements in and reinforcement of knowledge and competence are significant predictors of changes in self-efficacy [20]. Lucero et al [28] supported this finding. They found that participants who reinforced their knowledge had higher posteducation confidence ratings than participants who improved their knowledge after controlling for posteducation scores. Reinforcement of knowledge also likely explains why neurologists demonstrated the most significant increase in the number of new AD diagnoses.

Limitations

Potential confounding factors could affect the relationship between CME participation and increased AD diagnoses. Physicians who participated in CME may have been more motivated to improve their practice, potentially leading to increased diagnoses regardless of the CME content. Three activities were case-based simulated patient visits and 1 was a video-based discussion on cases. We did not tease out which activities might have been more or less impactful and whether participation in multiple activities was associated with practice change. Concurrent initiatives, such as other AD awareness campaigns or participation in non–study-related CME activities focused on AD or cognitive disorders during the study period, could also have confounded results. The control group also saw an increase of 91 more diagnoses postintervention, suggesting some external factors may have influenced diagnosis rates. Professional learning occurs in many places, given the demands of clinical practice and the requirement to maintain licensure. While changes for the control group were anticipated, matching based on demographic and practice factors helps reduce biases associated with those factors such as opportunity to diagnose, training, types of patients seen, and environment in which one practices.

Despite these limitations, the comprehensive matching strategy minimizes potential confounding factors and ensures group comparability. By matching profession, specialty, and ZIP code, the study controls for some differences in baseline knowledge, experience, and practice patterns associated with different medical specialties, as well as variations in patient demographics and health care access. Matching based on the number of patients with AD controlled for differences in patient population and exposure to AD cases. Using a time-aligned control group helped to control temporal factors, such as concurrent initiatives, that would affect both groups equally. Results were assessed by counting claims nested in patients to better tease out patients with their first AD diagnosis from a learner versus a nonlearner physician. Using paired pre- and postintervention data for individual learners enhances the statistical precision of the analysis, reducing sampling error and providing a robust assessment of the education's impact. Including the matched control group at follow-up increases confidence that changes are associated with education. Future research should explicitly examine how CME interventions affect AD diagnosis rates across different racial and ethnic groups and identify with more detail the mechanisms for change. We identified self-efficacy as a mechanism for practice change, but we should further understand which components in the CME influenced self-efficacy.

Conclusion and Significance

Diagnostic delays contribute to suboptimal patient outcomes in AD. By using a matched case-control design and assessing both immediate educational outcomes and subsequent changes in diagnostic behavior, this study provides evidence for the potential of CME as a tool to increase AD diagnosis. This web-based CME intervention increased participant likelihood of diagnosing AD, led to a greater number of new AD diagnoses than the control group, and fostered a positive relationship between postintervention confidence and diagnosis rates. Building self-efficacy should be a key objective in education interventions with practice-changing potential, alongside improving, reinforcing, and validating existing knowledge. Overall, this study shows the power of real-world data in demonstrating the impact of CME on clinical behavior and offers a first step in identifying CME's impact on dementia care. We are currently conducting a second phase of this initiative. Future directions could include a breakdown of CME engagement levels and learning outcomes by specialty to clarify which provider groups benefit most from this intervention.

Acknowledgments

We thank Garimesh Kumar for his work on pulling and analyzing data; to continuing medical education content faculty Charles Vega, MD; SB; Frances McFarland, MA, PhD; and Sharon Cohen, MD; and Alexandra Howson, PhD, for written contributions. The study was funded by an Independent Medical Education Grant from Lilly.

XSL•FO
RenderX

Data Availability

The datasets generated and analyzed during this study are not publicly available due to Medscape member privacy. Access to the data is restricted to Medscape employees who have permission. Statistical code and study protocol are available from the corresponding author on reasonable request.

Conflicts of Interest

TF and KL are employees of Medscape, LLC. SB receives funding from the Centers for Disease Control, National Institute on Aging, and National Institute of Neurological Disease and Stroke; honoraria as deputy editor of the *Journal of the American Geriatrics Society*; consulting fees for service on clinical and scientific advisory boards for Biogen, Eisai, Novo Nordisk, Abbvie, Lilly, and Linus Health, and as a speaker and content consultant from Medscape.

Multimedia Appendix 1
Codes used for the study.
[DOCX File, 13 KB - mededu_v11i1e72000_app1.docx]

References

1. Alzheimer's Association. 2022 Alzheimer's disease facts and figures. Alzheimer's & Dementia 2022 Apr;18(4):700-789. [doi: 10.1002/alz.12638]
2. Dementia. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/dementia [accessed 2024-10-28]
3. Mintun MA, Lo AC, Duggan Evans C, et al. Donanemab in early Alzheimer's disease. N Engl J Med 2021 May 6;384(18):1691-1704. [doi: 10.1056/NEJMoa2100708] [Medline: 33720637]
4. van Dyck CH, Swanson CJ, Aisen P, et al. Lecanemab in early Alzheimer's disease. N Engl J Med 2023 Jan 5;388(1):9-21. [doi: 10.1056/NEJMoa2212948] [Medline: 36449413]
5. Arvanitakis Z, Shah RC, Bennett DA. Diagnosis and management of dementia: review. JAMA 2019 Oct 22;322(16):1589-1599. [doi: 10.1001/jama.2019.4782] [Medline: 31638686]
6. Sluder KM. Acknowledging disparities in dementia care for increasingly diverse ethnoracial patient populations. Fed Pract 2020 Feb;37(2):69-71. [Medline: 32269468]
7. Judge D, Roberts J, Khandker RK, Ambegaonkar B, Black CM. Physician practice patterns associated with diagnostic evaluation of patients with suspected mild cognitive impairment and Alzheimer's disease. Int J Alzheimers Dis 2019;2019(4942562):4942562. [doi: 10.1155/2019/4942562] [Medline: 30937189]
8. Podhorna J, Winter N, Zoebelein H, Perkins T, Walda S. Alzheimer's diagnosis: real-world physician behavior across countries. Adv Ther 2020 Feb;37(2):883-893. [doi: 10.1007/s12325-019-01212-0] [Medline: 31933051]
9. Bouldin E. BRFSS data on cognitive decline and caregiving: national and state implications. Innov Aging 2024 Dec 31;8(Supplement_1):477-477. [doi: 10.1093/geroni/igae098.1556]
10. Berenbaum R, Dresner J, Maaravi Y, Erlich B, Pivko N, Tziraki C. Translating knowledge into practice at the local level: evaluation of a pilot CME for primary care physicians on dementia early diagnosis and management. Int Psychogeriatr 2020 Dec;32(12):1469-1470. [doi: 10.1017/S1041610219000097] [Medline: 30782229]
11. Jack CR Jr, Andrews JS, Beach TG, et al. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. Alzheimers Dement 2024 Aug;20(8):5143-5169. [doi: 10.1002/alz.13859] [Medline: 38934362]
12. Bernstein A, Rogers KM, Possin KL, et al. Dementia assessment and management in primary care settings: a survey of current provider practices in the United States. BMC Health Serv Res 2019 Nov 29;19(1):919. [doi: 10.1186/s12913-019-4603-2] [Medline: 31783848]
13. Pais M, Martinez L, Ribeiro O, et al. Early diagnosis and treatment of Alzheimer's disease: new definitions and challenges. Braz J Psychiatry 2020 Aug;42(4):431-441. [doi: 10.1590/1516-4446-2019-0735] [Medline: 31994640]
14. Prins A, Hemke F, Pols J, Moll van Charante EP. Diagnosing dementia in Dutch general practice: a qualitative study of GPs' practices and views. Br J Gen Pract 2016 Jun;66(647):e416-e422. [doi: 10.3399/bjgp16X685237] [Medline: 27114209]
15. Lathren CR, Sloane PD, Hoyle JD, Zimmerman S, Kaufer DI. Improving dementia diagnosis and management in primary care: a cohort study of the impact of a training and support program on physician competency, practice patterns, and community linkages. BMC Geriatr 2013 Dec 10;13(134):24325194. [doi: 10.1186/1471-2318-13-134] [Medline: 24325194]
16. Schütze H, Shell A, Brodaty H. Development, implementation and evaluation of Australia's first national continuing medical education program for the timely diagnosis and management of dementia in general practice. BMC Med Educ 2018 Aug 10;18(1):194. [doi: 10.1186/s12909-018-1295-y] [Medline: 30097036]
17. Casey AN, Islam MM, Schütze H, et al. GP awareness, practice, knowledge and confidence: evaluation of the first nation-wide dementia-focused continuing medical education program in Australia. BMC Fam Pract 2020 Jun 10;21(1):104. [doi: 10.1186/s12875-020-01178-x] [Medline: 32522153]

XSL•FO
RenderX

18. Moehead A, DeSouza K, Walsh K, Pit SW. A web-based dementia education program and its application to an Australian web-based dementia care competency and training network: integrative systematic review. J Med Internet Res 2020 Jan 22;22(1):e16808. [doi: 10.2196/16808] [Medline: 32012077]
19. Bandura A. Human agency in social cognitive theory. Am Psychol 1989 Sep;44(9):1175-1184. [doi: 10.1037/0003-066x.44.9.1175] [Medline: 2782727]
20. Lucero KS, Chen P. What do reinforcement and confidence have to do with it? A systematic pathway analysis of knowledge, competence, confidence, and intention to change. J Eur CME 2020 Oct 12;9(1):1834759. [doi: 10.1080/21614083.2020.1834759] [Medline: 33133769]
21. Spyropoulos J, Boutsalis G, Lucero K, Waskelo J, Wilson K, Anderson DR. Improving appropriate use of omega-3 fatty acids for patients with dyslipidemia: effect of online CME. Crit Pathw Cardiol 2021 Dec 1;20(4):208-212. [doi: 10.1097/HPC.0000000000000265] [Medline: 34431820]
22. Lucero KS, Larkin A, Zakharkin S, Wysham C, Anderson J. The impact of web-based continuing medical education using patient simulation on real-world treatment selection in type 2 diabetes: retrospective case-control analysis. JMIR Med Educ 2023 Aug 29;9:e48586. [doi: 10.2196/48586] [Medline: 37642994]
23. Lucero KS, Moore DE Jr. A systematic investigation of assessment scores, self-efficacy, and clinical practice: are they related? J CME 2024;13(1):2420373. [doi: 10.1080/28338073.2024.2420373] [Medline: 39498264]
24. Moore DE Jr, Green JS, Gallis HA. Achieving desired results and improved outcomes: integrating planning and assessment throughout learning activities. J Contin Educ Health Prof 2009;29(1):1-15. [doi: 10.1002/chp.20001] [Medline: 19288562]
25. Best practices in the delivery of care for Alzheimer's disease: from primary care to neurology. Medscape. URL: https://www.medscape.org/viewarticle/980622 [accessed 2025-05-09]
26. Waiting room for patients with cognitive impairment. Medscape. URL: https://www.medscape.org/sites/pwr/cognitive-impairment [accessed 2025-05-09]
27. 45 CFR 46.104 -- exempt research. Code of Federal Regulations. URL: https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-A/part-46/subpart-A/section-46.104 [accessed 2025-05-13]
28. Lucero KS, Williams B, Moore DE Jr. The emerging role of reinforcement in the clinician's path from continuing education to practice. J Contin Educ Health Prof 2023 Nov 14;44(2):143-146. [doi: 10.1097/CEH.0000000000000541] [Medline: 37962911]

Abbreviations

AD: Alzheimer disease
AMA PRA: American Medical Association Physician Recognition Award
CME: continuing medical education
HCP: health care provider
ICD-10: *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision*
MCI: mild cognitive impairment
NPI: National Provider Identification
OR: odds ratio
PCP: primary care physician
RWO: real-world outcome

Edited by T Gladman; submitted 31.01.25; peer-reviewed by C Br, MJ Ebadi, RS Gomaa Mahmoud, SM Shaffi; revised version received 03.04.25; accepted 28.04.25; published 22.05.25.

Please cite as:
Lucero K, Finnegan T, Borson S
Using Web-Based Continuing Education to Improve New Diagnoses of Alzheimer Disease in Claims Data: Retrospective Case-Control Study
JMIR Med Educ 2025;11:e72000
URL: https://mededu.jmir.org/2025/1/e72000
doi:10.2196/72000

© Katie Lucero, Thomas Finnegan, Soo Borson. Originally published in JMIR Medical Education (https://mededu.jmir.org), 22.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring Health Care Professionals' Perspectives on Education, Awareness, and Preferences for Digital Educational Resources to Support Transgender, Nonbinary, and Intersex Care: Interview Study

Sravya Katta[1], MSc; Nadia Davoody[1], MSc, PhD

Health Informatics Centre, Department of Learning, Informatics, Management, and Ethics, Karolinska Institutet, Stockholm, Sweden

Corresponding Author:
Nadia Davoody, MSc, PhD
Health Informatics Centre
Department of Learning, Informatics, Management, and Ethics
Karolinska Institutet
Tomtebodavägen 18 A
S-17177
Stockholm
Sweden
Phone: 46 (0)8 524 864
Email: nadia.davoody@ki.se

Abstract

Background: Health care professionals often face challenges in providing affirming and culturally competent care to transgender, nonbinary, and intersex (TNBI) patients due to a lack of understanding and training in TNBI health care. This gap highlights the opportunity for tailored educational resources to enhance health care professionals' interactions with TNBI individuals.

Objective: This study aimed to explore health care professionals' perspectives on education and awareness of health issues related to TNBI individuals. Specifically, it aimed to identify their needs, challenges, and preferences in accessing and using digital educational resources to enhance their knowledge and competence in providing inclusive and effective care for this population.

Methods: A qualitative research approach was used in this study. In total, 15 health care professionals were recruited via convenience sampling to participate in semistructured interviews. Thematic analysis was applied to identify recurring codes and themes.

Results: The study identified several themes and subthemes related to gender diversity awareness, inclusive communication and understanding the needs of TNBI individuals, societal and structural challenges, regulatory gaps in training and support infrastructure, education and training needs for health care professionals on TNBI care, educational resources and training tools for TNBI care, challenges and design considerations for eHealth tools integrations, and evaluating eHealth impact. Participants identified communication barriers, the need for health care providers to use inclusive language, and gaps in both health care system infrastructure and specialized training for gender-affirming care. In addition, participants expressed a need for comprehensive education on transgender and nonbinary health issues, resources for mental health professionals, user-friendly design, and accessibility features in eHealth tools.

Conclusions: The study revealed substantial deficiencies in health care professionals' knowledge of gender diversity, cultural competency, and the importance of inclusive communication. Addressing the identified barriers and challenges through targeted interventions, such as providing training and support for health care professionals, investing in user-friendly design and data security, and promoting cultural competence in TNBI health care, is essential. Despite integration challenges, eHealth tools have the potential to improve patient–health care professional relationships and access to care.

(JMIR Med Educ 2025;11:e67993) doi:10.2196/67993

XSL•FO
RenderX

KEYWORDS

health care professionals; transgender, nonbinary, and intersex; communication challenges; systematic barriers; information and communication technology

Introduction

Background

The term transgender refers to individuals whose gender identity or expression differs from the sex they were assigned at birth [1]. The concept has expanded over time to include a wide range of gender identities, such as transmen, transwomen, nonbinary individuals, and those who are gender nonconforming [2,3]. Nonbinary individuals may not exclusively identify as male or female. Their gender identity can be fluid, agender, or fall outside the binary spectrum [1,4]. The transgender community is highly diverse, and the understanding of transgender identity varies across different cultures [2]. Intersex individuals are those whose physical sex characteristics do not conform to the traditional binary classification of bodies as strictly male or female [4]. Previous epidemiological and clinic-based investigations have suggested that approximately 0.1% to 2% of the population identifies as transgender or with other noncisgender identities [5-7].

The literature reveals that transgender, nonbinary, and intersex (TNBI) individuals experience disproportionate levels of human rights violations and adverse health outcomes, largely attributed to intersecting forms of social marginalization and legal exclusion. Particularly transgender individuals, especially those from minority ethnic groups, are disproportionately impacted by gender-based hate crimes [8]. In addition, TNBI individuals face multiple challenges in accessing adequate health care services [9-12]. The challenges faced by TNBI individuals in accessing health care services are often rooted in systemic barriers that perpetuate stigma, discrimination, and lack of understanding. These systemic barriers contribute to disparities in health care access, quality, and outcomes for these individuals [9-12]. These barriers may include policies that fail to recognize their gender identity and social factors such as discrimination and stigma from health care providers [9-13]. TNBI individuals often face a range of health disparities, largely driven by a lack of awareness regarding their unique health needs. These disparities manifest in higher rates of mental health issues, such as depression and anxiety, as well as increased risks of sexually transmitted infections, including HIV, and other diseases such as cancer, smoking-related conditions, and cardiovascular disease [14,15]. The elevated risk of acquiring sexually transmitted infections among TNBI individuals is influenced by factors such as limited access to comprehensive sexual health education, prevention measures, and adequate preventive care [16]. Furthermore, TNBI individuals often face significant barriers to accessing gender-affirming health care, which can lead to poorer overall health outcomes [13].

In addition to these health care challenges, TNBI individuals frequently experience widespread stigma, discrimination, and prejudice in various facets of life, including employment, education, housing, and health care. This systemic stigma surrounding gender diversity and nonconformity creates a hostile environment for TNBI individuals even within health care systems [6,17-20]. In a sexual health seminar held in Minnesota, a sample of 181 transgender participants revealed that 66% reported experiencing discrimination based on their gender identity or presentation [21]. Consequently, the denial of essential health care services can intensify feelings of dysphoria and distress among transgender individuals. While TNBI individuals face significant challenges in accessing care, health care professionals also encounter numerous obstacles that hinder their ability to provide effective support and services to these individuals.

Challenges Faced by Health Care Professionals

Health care professionals play a crucial role in caring for patients, acting as key facilitators of essential health care services and resources. This highlights the responsibility of health care providers to promote the health and overall welfare of the populations including TNBI [22]. However, health care professionals face challenges in treating and communicating with TNBI individuals, as many of them receive minimal or no training during their medical education and professional development regarding hormone therapy, gender-affirming surgeries, and mental health support—key aspects that are aimed at aligning an individual's physical appearance and gender identity with their affirmed gender during medical education and professional development [10,23].

In addition, health care professionals often encounter challenges when interacting with TNBI individuals, including issues related to cultural competency, communication barriers, and a lack of knowledge about appropriate care practices [24]. These challenges can result in disparities in health care access and quality for TNBI individuals, leading to negative health outcomes and experiences of discrimination [25]. Without adequate training, health care professionals may lack the necessary knowledge and skills to provide competent and affirming care to TNBI patients. This can result in misdiagnosis and inappropriate treatment for patients [26]. Moreover, health care professionals who are unfamiliar with the specific health needs of TNBI individuals may inadvertently overlook important aspects of care and may fail to provide appropriate interventions. Increasing awareness and understanding of TNBI health issues through education, professional development, and exposure to diverse patient populations can help health care professionals better meet the needs of their TNBI patients [27].

In this study, we focused on 4 countries—Sweden, India, the United Kingdom, and the United States—due to their diverse sociocultural context, legal frameworks, and health care systems. This selection provides valuable insights into the varying approaches to TNBI health care and medical education within each country, which allows for a broad examination of how medical education on TNBI individuals can be improved globally. Sweden's progressive health care policies and a strong emphasis on human rights, including TNBI rights, make it an ideal setting to explore advanced practices and identify areas

XSL•FO
RenderX

for improvement. India, with a complex sociocultural landscape and a vast and diverse population, presents unique challenges and opportunities in providing health care to TNBI individuals. Understanding health care professionals' perspectives in India can reveal the specific needs and barriers faced by TNBI individuals in a resource-limited setting. The United Kingdom has recently experienced significant sociopolitical changes affecting health care policies for TNBI individuals. The United States' diverse health care environment regarding TNBI care, with some states enacting progressive policies and others imposing restrictions on gender-affirming treatments, offers insights into the challenges and successes in providing care to TNBI individuals in such varied regulatory settings. By understanding these varied contexts, we aimed to identify gaps in education and awareness as well as potential best practices to address health care professionals' needs, challenges, and preferences in accessing and using educational resources. This insight will help enhance their knowledge and competence in providing inclusive and effective care for TNBI individuals. The medical education systems in each of these countries present unique structures but share a common challenge: a lack of formal education on TNBI health care needs. In Sweden, despite its progressive stance on gender equality, medical curricula still rarely address the specific health care concerns of gender-diverse individuals. India's medical education, influenced by traditional values and diverse cultural perspectives, similarly lacks comprehensive training on gender diversity, despite growing awareness of TNBI rights [28]. The United Kingdom and the United States have made strides in addressing health care inequalities, yet many health care professionals report limited exposure to TNBI health topics during their training [29,30].

The Importance of Educational Tools and the Role of Information and Communication Technology

Lack of necessary training can result in misunderstandings, misgendering, and insensitive or inappropriate interactions that undermine trust and rapport between patients and health care professionals [31]. To address these challenges, it is essential to explore and develop tailored educational resources and interventions that provide health care professionals with the knowledge and skills needed to interact effectively with TNBI individuals [32]. Promoting equitable health care access for all individuals requires innovative solutions that empower health care professionals to provide supportive and inclusive care to TNBI patients [32-36]. Incorporating information and communication technology (ICT) into training and education has proven highly beneficial, as it enhances learning opportunities, improves communication, and increases accessibility to educational resources [37]. While there are other strategies for training and education, ICT offers unique advantages such as personalized learning, interactive content, and the ability to reach a wider audience [38]. These benefits make ICT a more effective approach for improving various aspects of education and training. ICT plays a transforming role in health education by providing innovative and accessible solutions to train health care professionals effectively. It also enhances learning flexibility, promotes collaborative opportunities, and ensures scalability to meet diverse educational needs in health care settings [39].

In health care, ICT can be leveraged through digital tools to support training by providing diverse and interactive educational resources, facilitating remote learning, and enabling real-time access to up-to-date medical information and best practices. The World Health Organization defines eHealth as the "cost-effective and secure use of information and communications technologies in support of health and health-related fields, including health-care services, health surveillance, health literature, and health education, knowledge and research" [40]. In this study, eHealth tools refer specifically to the digital health tools (eg, mobile apps and web-based platforms) used by health care professionals for health education purposes. By fostering inclusive practices, these tools can enhance patient trust, reduce discrimination, and ultimately lead to better health outcomes for these populations. Given the limited research on this topic for health care professionals [41], understanding health care professionals' needs, challenges, and preferences is vital for developing effective, targeted educational resources that promote more inclusive and effective care for TNBI individuals.

Aim of the Study

This study aimed to explore health care professionals' perspectives on education and awareness of health issues related to TNBI individuals. Specifically, it aimed to identify their needs, challenges, and preferences in accessing and using digital educational resources to enhance their knowledge and competence in providing inclusive and effective care for this population.

Methods

Study Design

A qualitative research approach was chosen to investigate health care professionals' perspectives on education, awareness, and preferences for digital educational resources to support TNBI care. This method aligns with the study's aim by comprehensively understanding their experiences, needs, and challenges in accessing and using educational resources. To further enrich this exploration, a sociotechnical framework [42] was applied, as it provides a structured perspective to examine how social, cultural, and technological factors intersect and influence the delivery of inclusive and effective care for TNBI populations. This framework has been applied to examine how health care professionals engage with TNBI individuals and the potential of eHealth educational tools to enhance these interactions. The strengths of qualitative research lie in the ability to gain profound insights into a problem or necessity by directly engaging with individuals and their contexts where the issue arises [43].

Study Setting and Participants

In total, 15 health care professionals with various health care backgrounds were recruited to participate in this study. The data were collected in Stockholm County, Sweden. While most participant interviews were held via Zoom (Zoom Video Communications), 2 were conducted in person at the participants' workplaces. The participants were selected through convenience sampling, primarily via social media platforms.

XSL•FO
RenderX

The inclusion criteria required participants to be health care professionals with experience in using digital tools in their practice, aged ≥18 years, and proficient in written and spoken English. To capture diverse perspectives, we included professionals across 9 disciplines (physiotherapy, dentistry, pediatrics, general practice, general surgery, gynecology, oncology, psychology, and cosmetic surgery), representing a broad spectrum of patient care. Recruiting participants from different disciplines and countries with varying acceptance, health care system diversity, legal recognition, and social and cultural attitudes toward TNBI individuals enabled us to gather varied insights on the challenges that health care professionals face in interacting with TNBI individuals. Each group of health care professionals in this study plays a crucial role in different aspects of health care, from initial assessments and referrals to specialized care, ongoing support, and mental health services. The inclusion of these varied perspectives was essential for capturing the complexity of care required for TNBI individuals. However, the diversity in the respondent pool also presents challenges, as it can make it more difficult to maintain focused discussions and reach consensus. The exclusion criteria were non–English-speaking health care professionals and individuals without any health care background. These criteria ensured that participants had relevant and thorough experiences, insights, and recommendations related to the research topic, thereby preserving the quality and validity of the study's findings.

In qualitative research, the number of participants is typically determined by reaching data saturation, meaning that further data collection does not yield new insights [44]. In this study, participants were interviewed until no new insights were generated from the interviews. The participants' characteristics are presented in Table 1. The participants had a mean age of 40 (SD 7.3) years and worked in various fields within the health care sector. The study included 4 participants from Sweden, 5 participants from India, and 3 participants each from the United States and the United Kingdom. The variation in participant numbers across countries was due to the use of convenience sampling and the differing availability of health care professionals in each discipline and location. Among the 15 participants, 8 (53%) were female, 6 (40%) were male, and 1 (6%) participant did not disclose their sex. In total, 6 (40%) out of 15 participants had no prior experience working with TNBI individuals.

Table 1. Participants' characteristics.

Participant ID	Age range (y)	Geographic location	Occupation	Experience in health care (y)	Experience working with TNBI[a] individuals
Participant 1	35-45	Sweden	Physiotherapist	5-10	No
Participant 2	30-40	Sweden	Dentist	1-5	No
Participant 3	30-40	Sweden	Pediatrician	5-10	No
Participant 4	30-40	Sweden	Dentist	1-5	Yes
Participant 5	30-40	India	General physician	5-10	Yes
Participant 6	40-50	India	General surgeon	10-15	Yes
Participant 7	50-60	India	Gynecologist	15-20	Yes
Participant 8	50-60	India	Oncologist	1-5	Yes
Participant 9	30-40	India	General practitioner	5-10	No
Participant 10	40-50	United Kingdom	General practitioner	5-10	Yes
Participant 11	30-40	United Kingdom	General surgeon	1-5	No
Participant 12	25-35	United Kingdom	Surgical intern	1-5	No
Participant 13	30-40	United States	Psychologist	5-10	Yes
Participant 14	35-45	United States	General practitioner	5-10	Yes
Participant 15	40-50	United States	Cosmetic surgeon	10-15	Yes

[a]TNBI: transgender, nonbinary, and intersex.

The selected professions were included because they are likely to engage TNBI individuals in their practice. This diversity allows us to capture a wide range of insights and experiences, which are crucial for understanding the multifaceted needs of these patients. Although 6 (40%) out of 15 participants had no prior experience working with TNBI individuals, their contributions were insightful and added significant value to the findings. By including a variety of roles, we aimed to identify common themes and differences across different medical specialties, which can inform more inclusive health care practices.

Data Collection

Overview

Data were collected through semistructured interviews. To formulate our interview script, we followed a systematic approach grounded in the sociotechnical framework by Sittig and Singh [42]. This framework ensures that we comprehensively address the intersection of social, cultural, and technical factors and their impact on providing inclusive and effective care for the TNBI population.

XSL•FO
RenderX

Development of the Interview Schedule

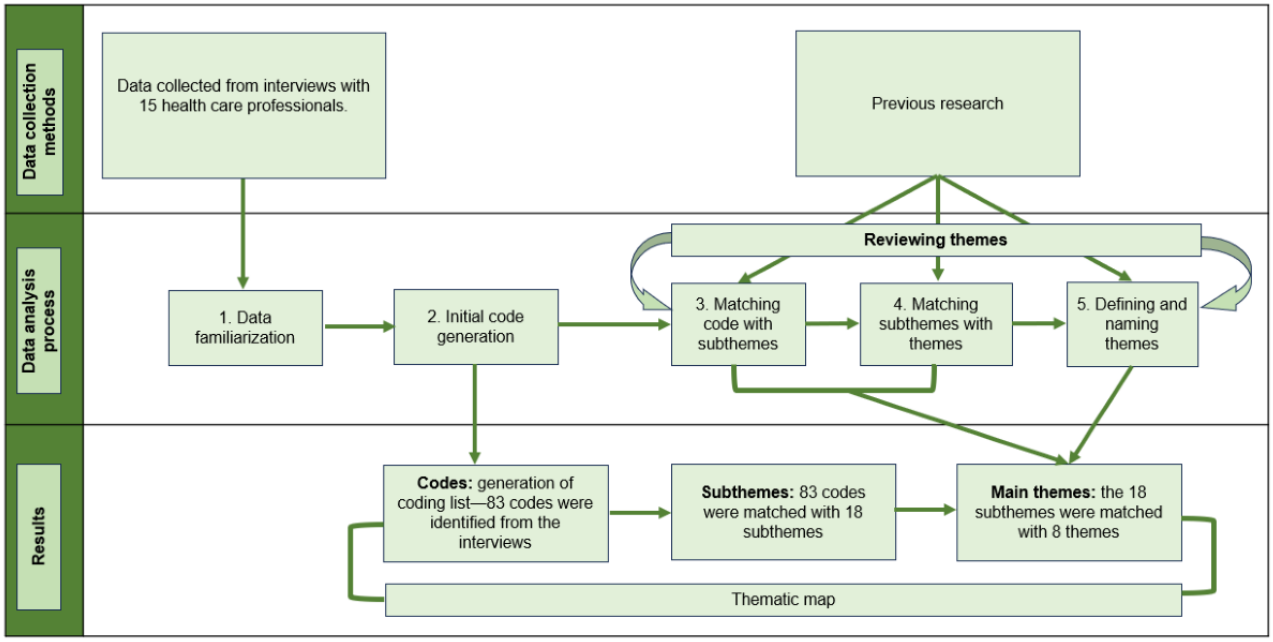
The interview schedule (Multimedia Appendix 1) was developed based on the 8 key dimensions of the sociotechnical framework: hardware and software; clinical content; human-computer interface; people; workflow and communication; internal policies, procedures, and culture; external rules and regulations; and measurement and monitoring [42]. We mapped the objectives of our study to the relevant dimensions of the framework to ensure comprehensive coverage: understanding current interactions and challenges (people and workflow and communication), identifying gaps in resources and training (internal policies, procedures, and culture and clinical content), exploring the potential of eHealth tools (hardware and software and human-computer interface), ensuring usability and integration (human-computer interface and communication), considering regulatory and organizational factors (external rules and regulations and internal policies, procedures, and culture), and evaluating effectiveness and feedback mechanisms (measurement and monitoring). We then continued with developing specific questions. For each dimension, we developed specific questions that align with our research objectives. The final interview script was designed to comprehensively address all aspects of the sociotechnical framework, ensuring the collection of detailed data on health care professionals' perspectives regarding education and awareness of health issues related to TNBI individuals and the potential of digital educational resources to enhance their care delivery. The interview guide was iteratively refined during early interviews to ensure its validity and relevance to the study's aim.

By using semistructured interviews, we could get a deeper understanding of what the participants had experienced during their clinical practice and their thoughts and feelings related to the need for digital educational resources. Semistructured interviews also allowed for follow-up questions to further deepen the understanding of their thoughts. Each interview lasted between 40 and 60 minutes. All the interviews were conducted in English.

Data Analysis

The interviews were audio recorded, transcribed, and analyzed using thematic analysis according to the guidelines by Braun and Clarke [45]. The audio recordings were transcribed verbatim using Microsoft Word and then reviewed for accuracy. The transcribed data were first thoroughly reviewed to gain a comprehensive understanding of the data. The next phase involved coding the text where codes are segments of the text that are linked by content and context, allowing a deeper exploration of the underlying themes and concepts. Descriptive coding was used to capture and summarize the main topics of the text. To build on this, pattern coding was applied to condense meaning units into overarching patterns, grouping the initial codes into broader themes. This approach facilitated the identification and understanding of larger patterns and relationships within the data [46]. SK conducted the initial coding by identifying meaning units, condensing them, and assigning relevant codes. Both authors reviewed the initial codes and grouped them into clusters reflecting emerging subthemes and then examined the relationships between the subthemes. They identified patterns and combined related subthemes into broader themes. The data analysis process was an iterative process in which discrepancies were resolved through several meetings and discussions. The authors analyzed the collected data and continuously revisited and refined the codes, subthemes, and themes as needed to ensure a comprehensive understanding of the data. The data analysis process is presented in Figure 1.

Figure 1. Data analysis process.



XSL•FO
RenderX

Ethical Considerations

This research was conducted in Sweden. According to the Swedish Ethical Review Act, this study does not require ethics approval as it does not handle sensitive personal information (as defined by the European General Data Protection Regulation). However, ethical requirements still apply. Participants were recruited from Sweden, India, the United States, and the United Kingdom. No sensitive personal information (eg, health status, political opinion, or racial or ethnic background) was collected. Prospective study participants were provided with comprehensive information regarding the study's objectives, methodologies, potential risks and benefits, and their right to withdraw at any time. This information was conveyed through both a written consent form, which participants were required to sign, as well as verbal explanations provided by the first author. To maintain confidentiality and privacy, the collected data were anonymized. In addition, any personal or sensitive information shared by participants was excluded from the study. The participants were guaranteed confidentiality and informed about how their data would be handled. Although the study was conducted in Sweden, ethical principles such as confidentiality and respect for participants' autonomy were upheld in line with international research standards, including those applicable in the United States (eg, Common Rule) [47], the United Kingdom (eg, Health Research Authority guidelines) [48], and India (eg, Indian Council of Medical Research guidelines) [49]. These measures ensured the ethical integrity of the research across all participant demographics.

Results

Overview

The analysis of the interviews resulted in 8 themes: gender diversity awareness, inclusive communication and understanding of the needs of TNBI individuals, societal structural challenges, regulatory gaps in training and support infrastructure, education and training needs for health care professionals on TNBI care, educational resources and training tools for TNBI care, challenges and design considerations for eHealth tools integration, and evaluating eHealth impact. In addition to the 8 themes, 18 subthemes and 83 codes were formulated. An overview of the sociotechnical aspects, themes, and subthemes is presented in Table 2.

In this study, the participants predominantly used the term *transgender* as an umbrella term to refer broadly to TNBI individuals.

Table 2. An overview of the subthemes and themes.

Sociotechnical aspects	Subthemes	Themes
People	• Limited understanding of gender diversity	Gender diversity awareness
Workflow and communication	• Acknowledgment of communication barriers and the need for inclusive language • Lack of understanding of TNBI[a] individuals' needs	Inclusive communication and understanding of the needs of TNBI individuals
Internal organizational policies, procedures, and culture	• Suppression of identity due to societal stigma, cultural norms, and societal pressures • Vulnerability arising from societal and political oppression • Limited research on TNBI health issues	Societal and structural challenges
External rules and regulations and pressures	• Lack of awareness among health care professionals and inadequate mental health support • Gaps in specialized training and guidelines for gender-affirming care • Weakness in health care infrastructure for TNBI individuals	Regulatory gaps in training and support infrastructure
Clinical content	• Inadequate training in cultural competency regarding gender diversity • Importance of education on TNBI health issues for health care professionals • Need for tailored resources and training modules designed for mental health professionals	Education and training needs for health care professionals on TNBI care
Hardware and software	• Interactive case studies and peer support forums • Comprehensive training modules workshops and web-based courses • Resources about gender-affirming therapy, trauma, and intersectionality	Educational resources and training tools for TNBI care
Human-computer interface	• Challenges in integrating eHealth tools into regular health care practice, including time constraints and cultural change • Emphasis on user-friendly design, accessibility features, and data security in health care tools	Challenges and design considerations for eHealth tools integration
System measurement and monitoring	• Expectations of improved patient–health care professional relationships using eHealth tools	Evaluating eHealth impact

[a]TNBI: transgender, nonbinary, and intersex.

XSL•FO
RenderX

Sociotechnical Aspect: People (Theme 1: Gender Diversity Awareness, Subtheme: Limited Understanding of Gender Diversity)

A hierarchical structure with the themes at the top, subthemes in the middle, and the corresponding codes at the base is illustrated in Figure 2.

A limited understanding of gender diversity may result in inadequate screening and assessment practices for TNBI patients' health care needs. Health care professionals discussed misunderstanding the unique health risks and concerns facing TNBI individuals, leading to delays in diagnosis, inappropriate treatment recommendations, or suboptimal care outcomes:

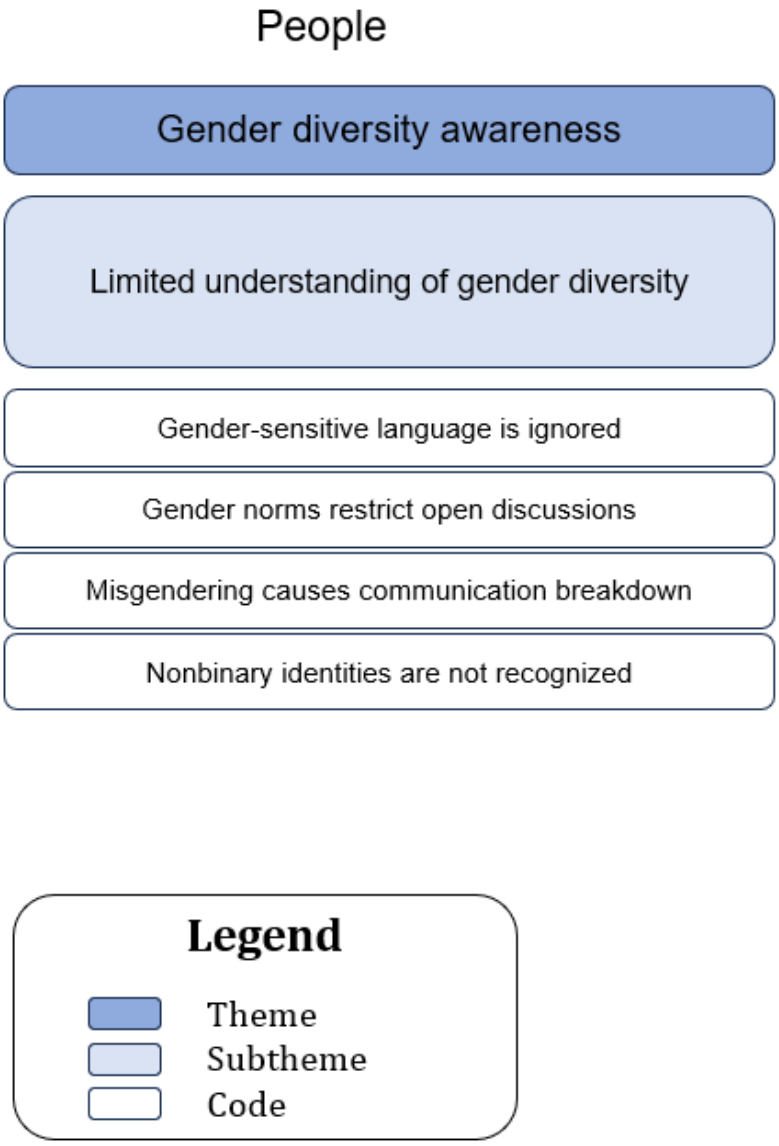
> *When I worked as an intern, we usually strengthened the biological sex or gender, ignoring the psychological stuff, because most of us in our education, are taught to learn the difference between male and female.* [Participant 10]

Some participants expressed that a limited understanding of gender diversity could create institutional barriers to accessing gender-affirming care, further exacerbating disparities in health care access and outcomes for TNBI individuals:

> *From my clinical work, outdated regulations often obstruct the care of transgender patients, a sign of the healthcare system's failure to fully grasp the nuances of gender diversity.* [Participant 11]

> *In my clinical experience, I've observed bureaucratic hurdles, such as outdated policies and procedures that fail to accommodate the unique needs of transgender individuals. These stemmed from a limited understanding of gender diversity.* [Participant 6]

Figure 2. The relationship between the codes and subthemes for theme 1—gender diversity awareness.

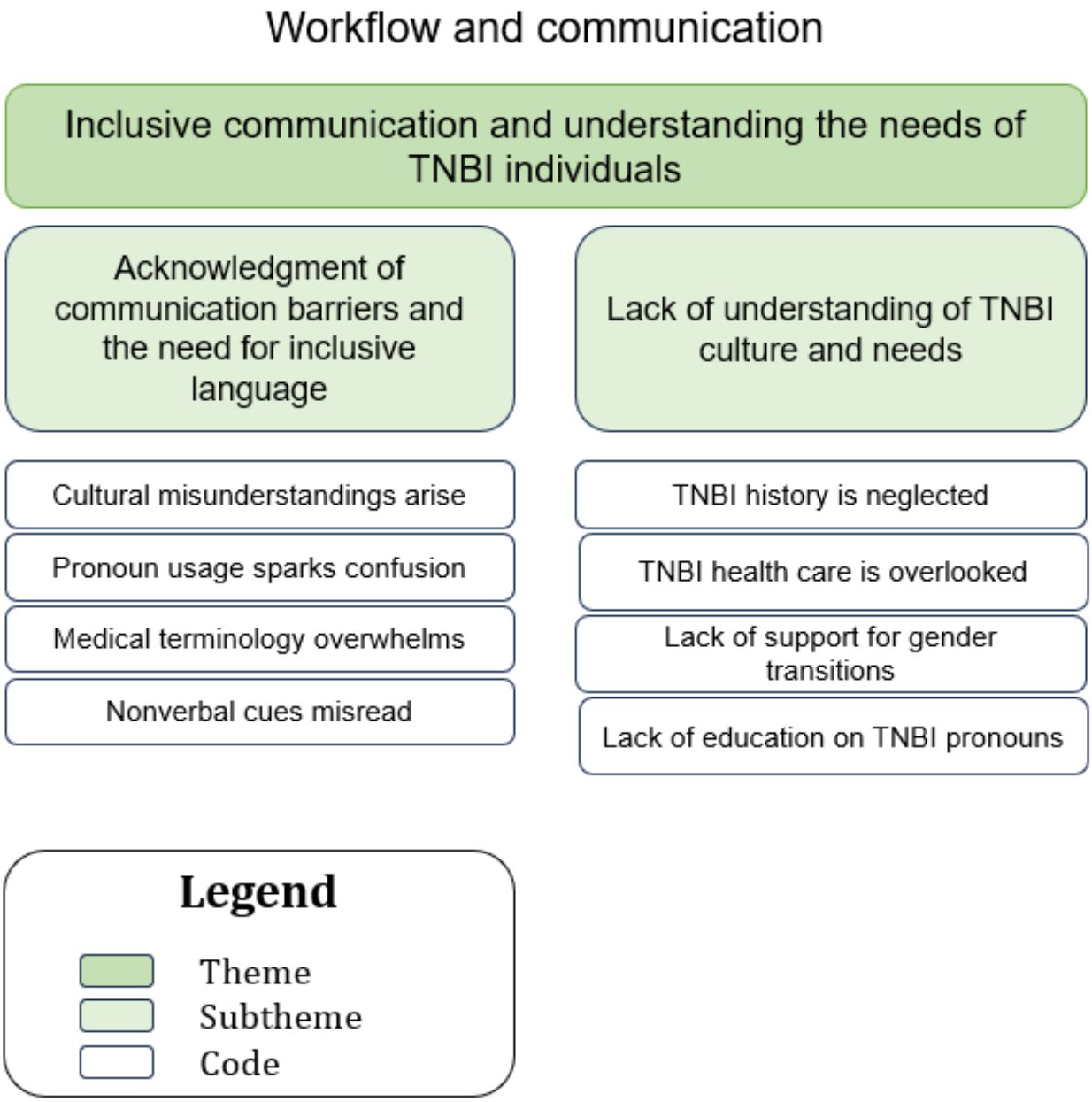


XSL•FO
RenderX

Sociotechnical Aspect: Workflow and Communication (Theme 2: Inclusive Communication and Understanding the Needs of TNBI Individuals)

A hierarchical structure with the themes at the top, subthemes in the middle, and the corresponding codes at the base is illustrated in Figure 3.

Figure 3. The relationship between the codes and subthemes for theme 2—inclusive communication and understanding the needs of transgender, nonbinary, and intersex (TNBI) individuals.



Acknowledgment of Communication Barriers and the Need for Inclusive Language

Participants felt an existing lack of awareness about TNBI health issues, including appropriate language and communication strategies. They also stated that without adequate education on TNBI terminology and cultural competency, they may unintentionally use insensitive or outdated language, leading to misunderstandings and discomfort for TNBI patients:

> *I used inappropriate language for communication with transgender patients and they felt bad for that and some of them were even frustrated for not having adequate knowledge of terms that should be used.* [Participant 7]

> *I've unintentionally used terms that were not affirming, which caused discomfort for my transgender patients, and this has shown me the urgent need for proper education on inclusive language.* [Participant 5]

Most of the participants presented fear of inadvertently misgendering them. This fear of causing harm or disrespect can lead to hesitation or avoidance of discussions related to gender identity, which can hinder effective communication and rapport building with TNBI patients:

XSL•FO
RenderX

> *Sometimes I did avoid discussions regarding gender identity, as I was not sure about it, due to which rapport with the patients was not constructive.* [Participant 4]

Some of the participants mentioned that there are limited resources and guidelines available to health care professionals on best practices for communication with TNBI patients. In the absence of clear guidance, they may struggle to navigate conversations about gender identity and may rely on personal assumptions or biases, which can contribute to communication barriers and misunderstandings:

> *There were moments when my judgment was clouded by my assumptions leading to uncomfortable situations. It is a mistake I have learned from.* [Participant 15]

> *Struggled a lot and also felt embarrassed due to my personal assumptions that led to misunderstandings in a peculiar situation and never did it again.* [Participant 6]

Lack of Understanding of TNBI Culture and Needs

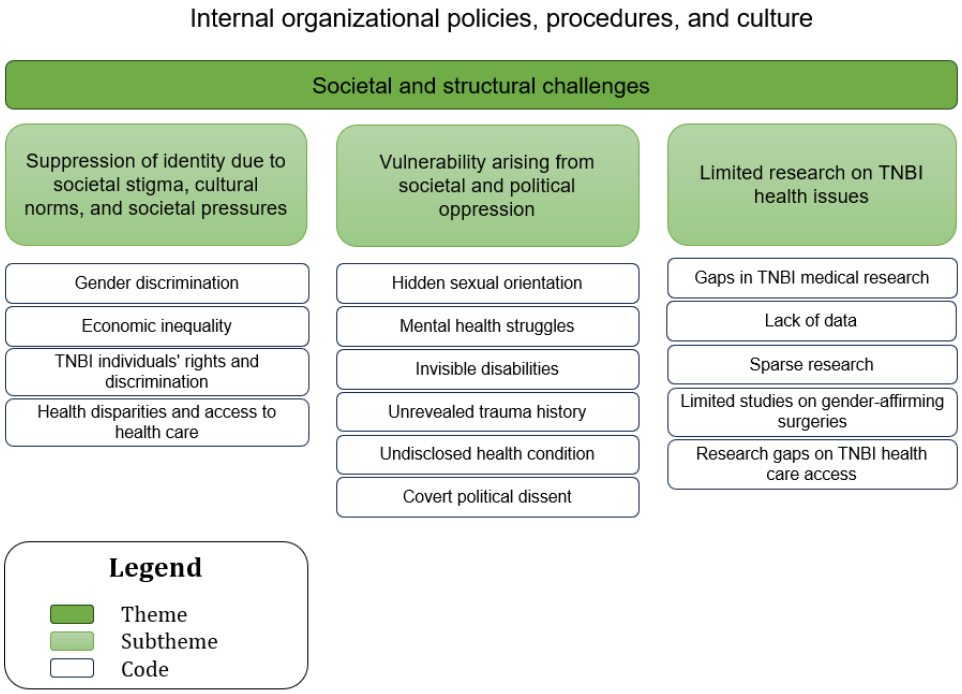
Participants also expressed that they struggled to establish trusting relationships due to a lack of understanding regarding their cultural identities and needs. Therefore, they observed a hindrance in disclosing patient's gender identity, expressing their health care concerns, and seeking support for their health and well-being:

> *In fact, I have no experience of working with transgender, hence building rapport with these patients is tough for me as I don't have a proper idea regarding their needs.* [Participant 2]

Sociotechnical Aspect: Internal Organizational Policies, Procedures, and Culture (Theme 3: Societal and Structural Challenges)

A hierarchical structure with the themes at the top, subthemes in the middle, and the corresponding codes at the base is illustrated in Figure 4.

Figure 4. The relationship between the codes and subthemes for theme 3—societal and structural challenges. TNBI: transgender, nonbinary, and intersex.



Suppression of Identity Due to Societal Stigma, Cultural Norms, and Societal Pressures

Participants highlighted that experiences of discrimination, prejudice, and social rejection increased the vulnerability of TNBI individuals to mental health conditions, such as depression, anxiety, and posttraumatic stress disorder. They also believe that internalizing negative stereotypes and beliefs about their identity can further exacerbate mental health issues, causing feelings of shame, self-doubt, and identity concealment:

> *During my experiences in clinical practice, I have encountered transgender patients who have expressed fear, shame, and hesitation in disclosing their gender identity to healthcare providers.* [Participant 10]

One of the participants said that traumatic experiences such as hate crimes, physical violence, or verbal abuse based on gender identity can lead to symptoms of posttraumatic stress disorder, including intrusive thoughts, hypervigilance, and avoidance behaviors:

> *They cannot say how they recognize themselves as transgender or non-binary in the workplace or even to their families. It's a kind of a secret for them. So, they are emotionally vulnerable and sensitive at the same time.* [Participant 13]

> *Many transgender individuals feel they must conceal their identities both professionally and personally,*

XSL•FO
RenderX

> *which places them under immense emotional strain and leaves them feeling isolated.* [Participant 2]

In addition to that, they also focused on the barriers to accessing health care faced by TNBI individuals because of societal stigma. Participants also revealed that past negative experiences or stories of discrimination within health care settings may lead to mistrust and reluctance to seek medical care.

Another point they mentioned is that the fear of being misgendered, invalidated, or subjected to invasive questioning can create significant barriers to accessing necessary health care services. In addition to that, internalized stigma and shame may also contribute to reluctance to seek health care services, as TNBI individuals may feel unworthy of receiving care or fear being perceived as *different*:

> *The first I think is the culture because trans people and also non-binary people, have their own culture, different than the majority culture and we, health care professionals didn't know that.* [Participant 5]

Some of the participants mentioned that using correct pronouns and respecting chosen names can help mitigate the effects of societal stigma and foster a sense of validation and belonging for TNBI patients:

> *It reminds me of a transgender patient, who did not behave like the gender that person looked like and that causes some confusion. So, it's necessary to ask them first how you recognize yourself and respect their social gender identification.* [Participant 5]

> *Gender identity is personal, and as clinicians, we should always lead with questions, not assumptions to ensure we are providing care that respects who they are.* [Participant 8]

Vulnerability Arising From Societal and Political Oppression

For TNBI individuals, intersectionality exacerbates vulnerability, particularly for those who belong to marginalized racial, ethnic, or socioeconomic groups. This means that TNBI individuals may face compounded discrimination and barriers to health care access due to overlapping forms of oppression:

> *I have seen many cases, who faced added discrimination to healthcare access due to political oppression...I believe that it's crucial to understand how multiple forms of discrimination intersect for transgender people, making access to healthcare even more challenging.* [Participant 6]

> *Political and social discrimination compounds the barriers transgender individuals face in accessing care, highlighting the need to address overlapping layers of bias in healthcare systems.* [Participant 11]

Limited Research on TNBI Health Issues

Participants said that they rely on evidence-based practices to guide their clinical decision-making and provide quality care to patients. The limited research on TNBI health issues means that there may be a lack of robust evidence to inform best practices in the diagnosis, treatment, and management of health conditions:

> *Without sufficient research, we are not only constrained in delivering optimal care, but it is also difficult to fully trust the treatments we prescribe to our patients.* [Participant 9]

> *Lack of research not only hampers our ability to deliver effective care but also undermines our confidence in the treatments and interventions we provide [...] As a result, we often find ourselves relying on anecdotal evidence, expert opinions, and extrapolations from related fields to inform our clinical decisions.* [Participant 4]

They also reported the lack of data on prevalence, risk factors, and outcomes of health conditions within these populations as a hindrance to their ability to assess and address their health care needs accurately. Without this information, it is challenging to determine the scope and magnitude of health disparities and to allocate resources effectively to address them:

> *Without accurate data, it's challenging to develop informed strategies for promoting the health and well-being of transgender and non-binary communities.* [Participant 3]

> *The absence of prevalence data makes it difficult to gauge the extent of health disparities within transgender and non-binary populations.* [Participant 13]

In addition, the lack of data on risk factors means that health care professionals may struggle to identify and mitigate factors that contribute to adverse health outcomes among TNBI individuals. Furthermore, the absence of data on outcomes of health conditions within TNBI populations hampers efforts to evaluate the effectiveness of interventions and treatments. Without data on treatment outcomes, health care professionals may be limited in their ability to tailor interventions to the unique needs of TNBI individuals and to optimize their health outcomes:

> *We struggle to allocate resources effectively and prioritize interventions without a clear understanding of the prevalence and severity of health conditions among transgender and non-binary individuals.* [Participant 5]

Participants also recognized the importance of investing in research initiatives focused on TNBI health to fill existing knowledge gaps and improve health care delivery:

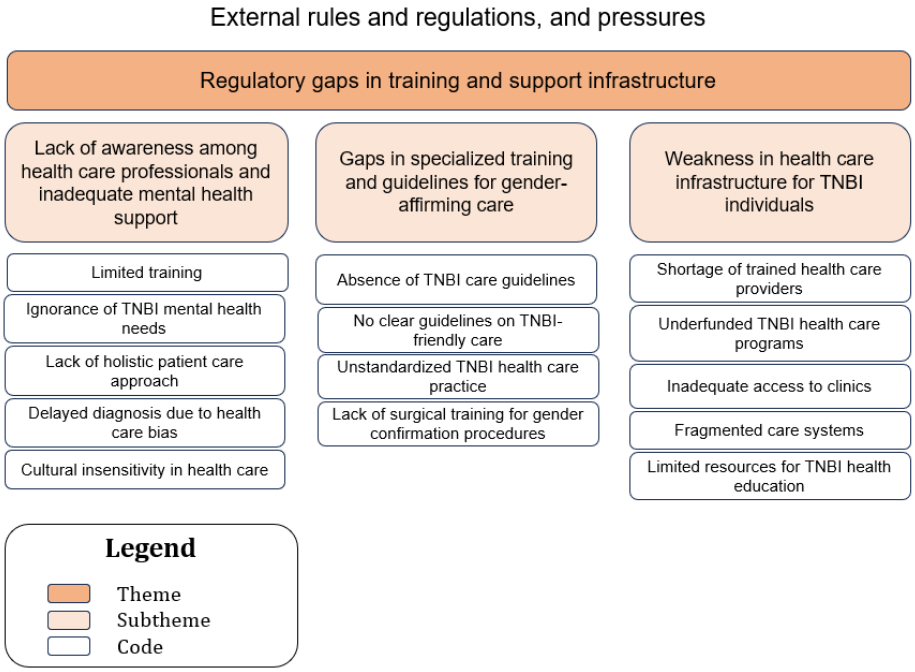
> *It's necessary that we prioritize funding and resources for research aimed at filling existing knowledge gaps, addressing disparities, and promoting health equity among transgender and non-binary populations in transgender and non-binary healthcare.* [Participant 4]

XSL•FO
RenderX

Sociotechnical Aspect: External Rules and Regulations and Pressures (Theme 4: Regulatory Gaps in Training and Support Infrastructure)

A hierarchical structure with the themes at the top, subthemes in the middle, and the corresponding codes at the base is illustrated in Figure 5.

Figure 5. The relationship between the codes and subthemes for theme 4—regulatory gaps in training and support infrastructure. TNBI: transgender, nonbinary, and intersex.



Lack of Awareness Among Health Care Professionals and Inadequate Mental Health Support

Significant gaps discussed by the participants include a lack of awareness and understanding of TNBI health issues. They also expressed feelings of being frustrated by the systemic gaps in education and training on TNBI health issues within their profession. They were also concerned about the limited availability of culturally competent and affirming mental health services for TNBI patients, recognizing the detrimental impact of untreated mental health conditions on their overall well-being:

> *I feel ill-equipped to address their unique healthcare needs, leading to challenges in providing culturally competent and affirming care.* [Participant 3]

Gaps in Specialized Training and Guidelines for Gender-Affirming Care

The participants' identification of gaps in specialized training and guidelines for gender-affirming care highlights a critical challenge within health care systems. These gaps stem from limited education and training on transgender health issues and gender-affirming care during both formal education and professional training programs. They also expressed concerns about the lack of comprehensive education and training on transgender health topics throughout their academic and professional journeys:

> *I feel the lack of preparation and education on transgender health issues is a systemic issue that requires immediate attention and action from healthcare institutions and educational programs.* [Participant 3]

Many reported minimal exposure to transgender health issues and gender-affirming care protocols during their formal education, which left them feeling ill-prepared to provide culturally competent and affirming care to transgender patients.

In contrast, they expressed that they are encountering inconsistencies in clinical protocols, treatment approaches, and referral criteria for TNBI patients, resulting in variations in care quality and patient experiences due to the lack of standardized guidelines within health care institutions and disciplines.

Even though some of the participants were interested in specializing in transgender health, they said that they were facing challenges in accessing advanced training opportunities to develop expertise in this area, leading to a scarcity of trained specialists within the health care workforce:

> *Even with a strong interest in transgender healthcare, I'm struggling to access the specialized training necessary to develop my skills in this field.* [Participant 12]
>
> *Despite my interest in specializing in transgender health, I am encountering challenges in accessing advanced training opportunities.* [Participant 13]

Weaknesses in Health Care Infrastructure for TNBI Individuals

The participants also stated that they observed a lack of specialized health care facilities equipped to provide gender-affirming care for TNBI individuals. In addition, they also reported the absence of dedicated clinics or centers specializing in transgender health, which limits access to competent care and contributes to disparities in health outcomes:

XSL•FO
RenderX

> *There's a noticeable absence of dedicated clinics specializing in transgender health within our healthcare system, due to which patients are facing difficulties in receiving the affirming care they need.* [Participant 6]

> *The scarcity of transgender-focused healthcare facilities means many patients struggle to find affirming care, which compromises their overall well-being.* [Participant 4]

They have also mentioned the challenges they were facing in accessing gender-affirming treatments, such as hormone therapy and gender-affirming surgeries, for TNBI individuals, including a shortage of trained health care providers and long waiting times for appointments. Apart from the aforementioned challenges, they also recognized the need for expanded access to mental health support services for TNBI individuals:

> *The limited availability of trained providers is impeding my ability to provide timely and comprehensive gender-affirming care.* [Participant 4]

> *The shortage of providers skilled in transgender care significantly delays access to the affirming treatments many patients require.* [Participant 8]

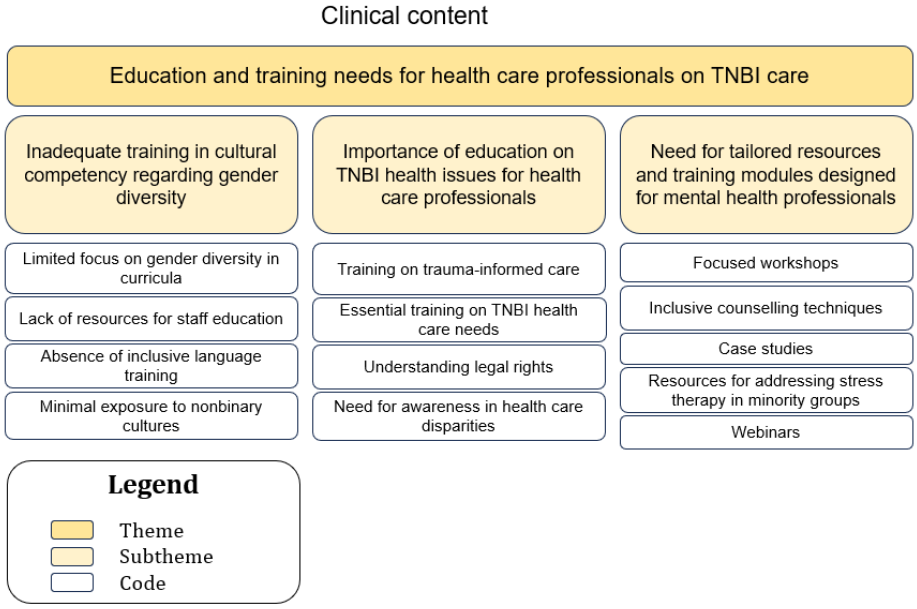
> *For transgender individuals, mental health support is essential for comprehensive care, yet it remains frustratingly inadequate in many healthcare settings.* [Participant 9]

> *Access to mental health support services is crucial for the holistic well-being of transgender individuals, yet it remains limited in many healthcare settings.* [Participant 6]

Sociotechnical Aspect: Clinical Content (Theme 5: Education and Training Needs for Health Care Professionals on TNBI Care)

A hierarchical structure with the themes at the top, subthemes in the middle, and the corresponding codes at the base is illustrated in Figure 6.

Figure 6. The relationship between the codes and subthemes for theme 5—education and training needs for health care professionals on transgender, nonbinary, and intersex (TNBI) care.



Inadequate Training in Cultural Competency Regarding Gender Diversity

Participants reported a lack of comprehensive instructions on gender diversity and cultural competency during their formal education and professional training. Many indicated that TNBI health topics were either inadequately covered or entirely omitted from their curriculum. Without proper education and training in this area, health care professionals find themselves inadequately prepared to navigate the complex landscape of gender identity and expression:

> *Without adequate training in transgender health, many of us enter practice unsure of how to approach the nuanced aspects of gender identity and provide affirming care to all patients.* [Participant 9]

> *The limited coverage of transgender health topics in our curriculum left us feeling ill-equipped to navigate the nuances of gender identity and expression in clinical practice.* [Participant 4]

> *Our training offered minimal focus on transgender health, leaving us underprepared to address the complexities of gender identity in practice.* [Participant 5]

Participants also expressed their struggles to provide patient-centered care that respects and affirms the diverse identities and needs of transgender patients, resulting in disparities in health care access, quality, and satisfaction:

> *I can say that I have experienced difficulties in tailoring care approaches to align with the diverse identities and needs of transgender patients. My*

XSL•FO
RenderX

> *attempts to provide affirming care to transgender patients have revealed gaps in my understanding and implementation of patient-centered principles.* [Participant 6]

Importance of Education on TNBI Health Issues for Health Care Professionals

Participants felt that education on TNBI health fosters the creation of inclusive health care environments where all patients feel respected, affirmed, and understood. Health care professionals who received training on transgender health issues are better equipped to provide culturally competent care, use affirming language, and create safe spaces for transgender patients to access health care without fear of discrimination or mistreatment:

> *I strongly believe education on transgender health is a crucial step towards building a healthcare system that is truly inclusive and affirming of all gender identities.* [Participant 5]

> *I assume by investing in education on transgender health, healthcare institutions can promote inclusivity and reduce disparities in healthcare access and outcomes for transgender individuals.* [Participant 7]

> *Investing in education about transgender health is a critical step toward fostering inclusivity and addressing disparities in care outcomes.* [Participant 9]

They also felt that education regarding TNBI health issues empowers them to recognize and address barriers, including delayed diagnosis and inappropriate treatment, ensuring that TNBI individuals receive timely, appropriate, and affirming health care services that meet their unique needs:

> *By educating ourselves on transgender health issues, we can break down barriers and create more inclusive healthcare environments and we can also work towards reducing disparities in healthcare access and outcomes for transgender individuals.* [Participant 10]

> *Through education on transgender health issues, we can eliminate obstacles to care and strive to reduce healthcare disparities for transgender individuals, ensuring better outcomes for all.* [Participant 12]

Participants were also interested in learning about the specific health care needs of TNBI patients, including hormone therapy, gender-affirming surgeries, preventive care, and mental health support, enabling them to deliver comprehensive and evidence-based care:

> *At least, I am eager to gain insights into the unique healthcare needs of transgender patients, including understanding nuances of hormone therapy and this interest in learning about the healthcare needs of transgender patients reflects our commitment to providing inclusive and affirming care within our practice.* [Participant 4]

Need for Tailored Resources and Training Modules Specifically Designed for Mental Health Professionals

The need for specialized training for mental health professionals to understand the unique mental health challenges faced by TNBI individuals has been observed by some participants, as these individuals are at increased risk of mental health disorders such as depression, anxiety, suicidality, and gender dysphoria due to societal stigma, discrimination, and identity-related stressors.

One of the participants, who was a mental health professional, was also expecting training in gender-affirming care principles to provide affirming and culturally competent mental health services to transgender and gender-diverse clients. The participant feels that tailored resources and training modules equip mental health professionals with strategies for creating affirming therapeutic environments and implementing gender-affirming interventions:

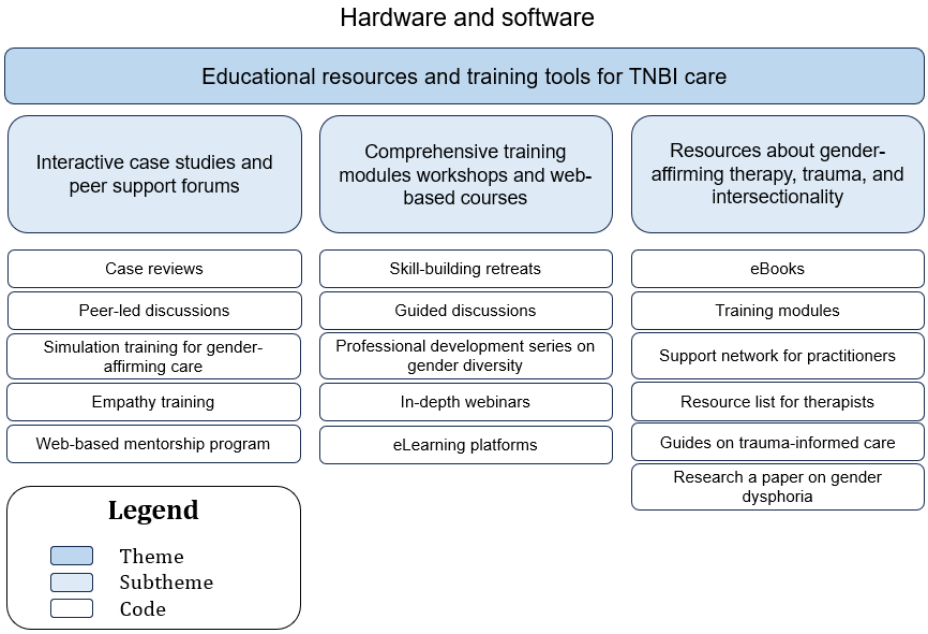
> *As per my observation, through education and training, mental health professionals can gain the knowledge and skills needed to provide competent and compassionate care to transgender and gender-diverse clients.* [Participant 13]

Sociotechnical Aspect: Hardware and Software (Theme 6: Educational Resources and Training Tools for TNBI Care)

A hierarchical structure with the themes at the top, subthemes in the middle, and the corresponding codes at the base is illustrated in Figure 7.

XSL•FO
RenderX

Figure 7. The relationship between the codes and subthemes for theme 6—educational resources and training tools for transgender, nonbinary, and intersex (TNBI) care.



Interactive Case Studies and Peer Support Forums

The participants generally perceived interactive case studies and peer support forums as valuable educational resources that promote active learning, cultural competency, professional networking, and skill development in transgender health.

Some of them felt that interactive case studies allow them to engage actively in the learning process by analyzing real-life scenarios, making decisions, and receiving immediate feedback:

> *My personal preference is to have interactive case studies, as they enable active participation and engagement in the learning process.* [Participant 7]

Others were interested in peer support forums that provide opportunities to learn from each other's experiences, perspectives, and insights, fostering a collaborative learning environment. They strongly believed that peer support forums facilitate knowledge sharing, discussion of clinical challenges, and exchange of best practices among peers, leading to enhanced learning outcomes and professional development.

They also felt that they could access these resources at their convenience to refresh their knowledge, stay updated on emerging practices, and enhance their clinical competencies in transgender health:

> *I agree that these resources encourage reflection, critical thinking, and dialogue among health care professionals, promoting continuous improvement in transgender healthcare delivery.* [Participant 3]

Comprehensive Training Modules, Workshops, and Web-Based Courses Focusing on Gender Diversity

To have an in-depth understanding of gender diversity, including the spectrum of gender identities and expressions, health care professionals were in favor of comprehensive training modules, workshops, and web-based courses.

They expect the training modules, workshops, and web-based courses to facilitate interdisciplinary collaboration among themselves from different specialties and disciplines, leading them to work collaboratively as part of multidisciplinary care teams to address the complex health care needs of transgender and gender-diverse patients, promoting coordinated and holistic care approaches:

> *The opportunity to exchange knowledge and best practices among health care professionals is vital for advancing our shared understanding of transgender health issues.* [Participant 11]

Some participants felt that training modules, workshops, and web-based courses focusing on gender diversity contribute to advancing health equity and social justice for transgender and nonbinary gender-diverse individuals. On the contrary, one of the participants expressed concern regarding time management:

> *In my opinion, it's essential to tackle issues regarding time management and resource allocation to ensure health care professionals can actively participate in gender diversity training and play their part in promoting health equity and social justice.* [Participant 2]

Resources About Gender-Affirming Therapy, Trauma, and Intersectionality

Health care professionals are expected to have resources that cover principles and practices of gender-affirming care, including approaches to hormone therapy, gender-affirming surgeries, and psychotherapy, enabling them to support transgender patients in aligning their bodies with their gender identities:

> *I expect to have access to resources on gender-affirming care will equip us with the knowledge and skills needed to navigate complex healthcare decisions and provide holistic support to*

XSL•FO
RenderX

transgender patients throughout their transition journey. [Participant 3]

In addition, they were willing to have and learn about trauma-sensitive approaches to care delivery, recognizing the impact of past traumas on patients' mental health and well-being and creating safe and supportive environments for survivors of trauma to access health care services:

I am open to learning and implementing trauma-sensitive approaches into our practice, as I believe they are essential for providing patient-centered and holistic care to individuals affected by trauma. [Participant 13]

As health care professionals, we recognize the importance of ongoing education and training in trauma-sensitive care to ensure that we can effectively meet the needs of patients who have experienced trauma. [Participant 10]

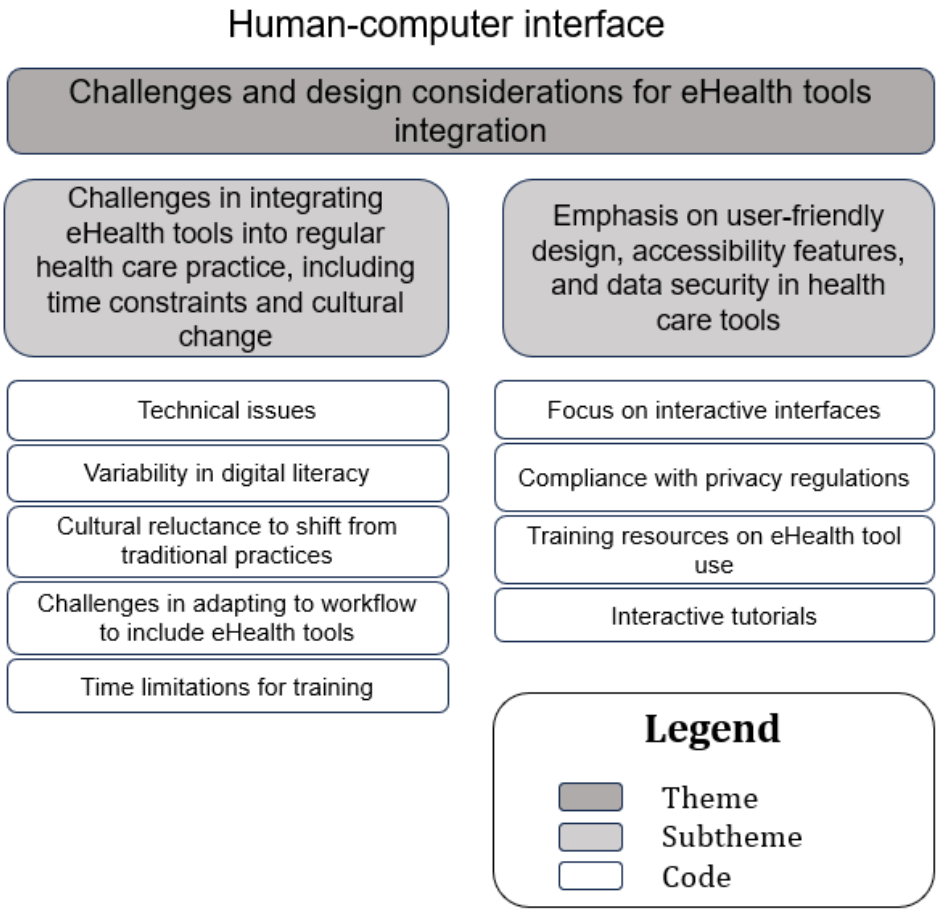
Health care professionals were interested in gaining an understanding of how multiple intersecting identities influence individuals' experiences of discrimination, access to health care, and health outcomes, enabling them to provide more nuanced and inclusive care to diverse transgender communities:

My priority is having education about intersectionality to work towards creating more equitable and inclusive healthcare systems that address the complex needs of all patients. [Participant 13]

Sociotechnical Aspect: Human-Computer Interface (Theme 7: Challenges and Design Considerations for eHealth Tools Integration)

A hierarchical structure with the themes at the top, subthemes in the middle, and the corresponding codes at the base is illustrated in Figure 8.

Figure 8. The relationship between the codes and subthemes for theme 7—challenges and design considerations for eHealth tools integration.



Challenges in Integrating eHealth Tools Into Regular Health Care Practice, Including Time Constraints and Cultural Change

Several challenges were recognized in integrating eHealth tools into regular health care practice, including time constraints and cultural change. Participants said that they often face time constraints due to busy schedules, heavy workloads, and competing priorities in clinical practice.

They recognized that integrating eHealth tools could require additional time for training, learning new technologies, documentation, and troubleshooting technical issues. This demand may strain already limited time resources and disrupt workflow efficiency. Participants acknowledged that integrating eHealth tools requires a cultural shift in attitudes, behaviors, and practices within health care organizations and among health care professionals:

XSL•FO
RenderX

> *With our heavy workloads, it can be challenging to dedicate time to training and troubleshooting the technical issues that arise with eHealth tools.* [Participant 14]

> *Require additional time for documentation and adapting to new workflows, which can strain our already limited time resources.* [Participant 7]

They recognized that resistance to change, skepticism about technology, and concerns about the impact on traditional care delivery models may impede the adoption and acceptance of eHealth tools, necessitating cultural change initiatives, leadership support, and stakeholder engagement to foster a culture of innovation and digital transformation:

> *In the place I work, to overcome resistance to change, proactive efforts are required to educate, train, and support healthcare staff in adopting new technologies and workflows.* [Participant 7]

One participant anticipated that integrating eHealth tools into regular health care practice may disrupt existing workflows and processes, leading to initial challenges in adaptation and implementation. They expressed concerns about potential workflow inefficiencies, disruptions in patient flow, and coordination issues among health care team members, particularly during the transition phase, when adapting to new technologies and integrating them into clinical routines:

> *It would be difficult for most of us during the transition phase, as there may be a need for additional training, support, and resources to help us navigate the changes and overcome implementation barriers.* [Participant 13]

Participants expressed that they even encounter technical and logistical barriers, such as inadequate infrastructure, limited access to technology, interoperability challenges, and data security concerns, which hinder the seamless integration of eHealth tools into regular health care practice. They recognized the need for investment in IT infrastructure, resources for training and support, and adherence to regulatory requirements and standards to address these barriers effectively and ensure the successful implementation and use of eHealth tools:

> *Inadequate infrastructure and limited access to technology might hinder the seamless integration of eHealth tools into our practice.* [Participant 3]

They recognized that patients may have varying levels of comfort, literacy, and access to technology, which can influence their willingness and ability to engage with eHealth tools. Health care professionals emphasized the need for patient education, support, and empowerment to promote successful integration and use of eHealth tools in regular health care practice:

> *As health care professionals, we play a critical role in facilitating patient engagement and empowerment in the use of eHealth tools by providing guidance, encouragement, and ongoing support.* [Participant 15]

Emphasis on User-Friendly Design, Accessibility Features, and Data Security in eHealth Tools

Health care professionals prioritized user-friendly design in eHealth tools to ensure ease of use, intuitive navigation, and efficient workflow integration. They recognize that user-friendly interfaces enhance usability, minimize user errors, and promote acceptance and adoption among health care professionals, ultimately improving efficiency and productivity in clinical practice.

Health care professionals prioritized data security and privacy in eHealth tools to protect sensitive patient information, maintain confidentiality, and comply with regulatory requirements:

> *I prefer the user-friendly design of eHealth tools to ensure that we can easily navigate and utilize these technologies in our daily practice.* [Participant 4]

Health care professionals recognized that user-friendly design and accessibility features in eHealth tools contribute to patient engagement and empowerment. They believed that tools that are easy to use, accessible across devices, and customizable to individual preferences encourage active participation, self-management, and shared decision-making among patients, leading to improved health outcomes and patient satisfaction. Health care professionals' confidence and satisfaction with eHealth tools are influenced by the design, usability, and security features of these tools.

They expressed a preference for tools that are intuitive, reliable, and secure, enabling them to focus on patient care rather than technical frustrations or concerns about data integrity, thereby enhancing their overall professional experience and job satisfaction:

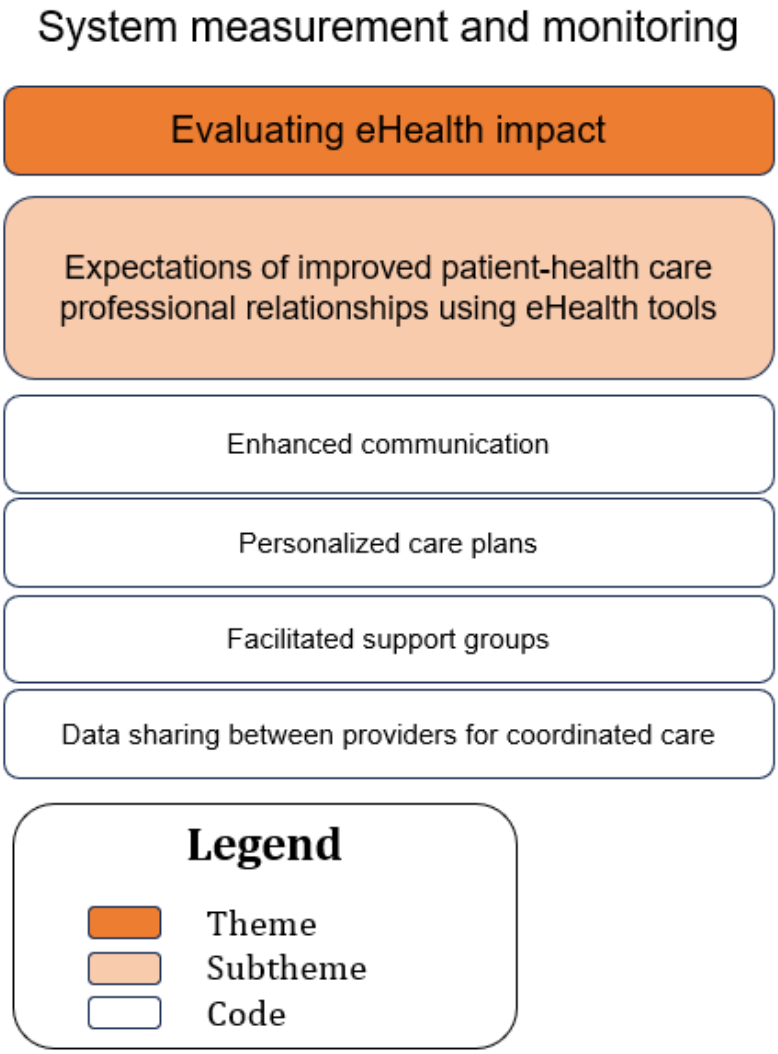
> *I believe that patient-centric design principles, such as intuitive interfaces and clear navigation, are essential for promoting patient autonomy and engagement in their healthcare.* [Participant 2]

Sociotechnical Aspect: System Measurement and Monitoring (Theme 8: Evaluating eHealth Impact, Subtheme: Expectations of Improved Patient–Health Care Professional Relationships Using eHealth Tools)

A hierarchical structure with the themes at the top, subthemes in the middle, and the corresponding codes at the base is illustrated in Figure 9.

XSL•FO
RenderX

Figure 9. The relationship between the codes and subthemes for theme 8—evaluating eHealth impact.



Health care professionals anticipated that eHealth tools would enhance communication, enabling more frequent and accessible contact with patients. This, in turn, could strengthen the patient–health care professional relationship by fostering trust, engagement, and continuity of care:

> *I still work in a healthcare setting, where there is no eHealth tools included in the regular practice. Through the integration of eHealth tools into our practice, we aim to create a healthcare environment that values transparency, accessibility, and patient engagement.* [Participant 7]

Discussion

Principal Findings and Comparison With Prior Work

Overview

The overall theme of this study is enhancing health care professionals' education and awareness of inclusive TNBI care. The analysis of health care professionals' interviews revealed significant insights into the challenges they face in providing care to TNBI individuals and the gaps in health care systems regarding these individuals' health. The analysis also highlighted the pivotal role of eHealth educational tools in addressing various challenges faced by health care professionals in providing inclusive and affirming care to transgender and gender-diverse individuals. Through its features and functionalities, an educational tool aims to bridge gaps in knowledge, communication, and access to resources, ultimately enhancing health care delivery and patient outcomes. In total, 8 themes were identified, highlighting the multifaceted nature of TNBI health care and the complex interplay among societal, cultural, regulations, institutional, technical, and individual factors.

Societal stigma and structural challenges emerged as a pervasive issue affecting the health and well-being of TNBI individuals. Participants highlighted the detrimental effects of discrimination, prejudice, and social rejection on mental health, emphasizing the need for culturally competent and affirming care to mitigate these challenges. In countries such as the United Kingdom and the United States, social acceptance and legal protections for TNBI individuals have progressed significantly, with antidiscrimination laws and health care rights actively supporting gender diversity. However, unlike in India, where acceptance remains complex due to strong cultural and religious norms, participants from these Western countries still reported that many patients experience fear, shame, and hesitation when disclosing their gender identity—whether to health care

XSL•FO
RenderX

professionals, in the workplace, or even within their families. The fear of being misgendered or invalidated, coupled with past traumas and societal pressures, creates significant barriers to accessing health care services, underscoring the importance of creating safe and supportive environments within health care settings.

Regarding external rules and regulatory gaps, although the United Kingdom and the United States are generally more open in terms of social and cultural norms, the recent sociopolitical shift in both countries has negatively affected TNBI care. For example, the United Kingdom has recently removed gender-affirming care with puberty blockers for those aged <18 years [50]. In the United States, >20 states have aimed to limit or ban access to gender-affirming care, especially for minors. These measures also included restrictions on access to mental health support and educational resources for TNBI individuals. Such policies reflect a growing sociopolitical climate that not only affects the patients directly but also impacts the engagement of health care professionals in providing gender-affirming care. These changes can lead to decreased accessibility to necessary treatments and negatively affect health outcomes for TNBI individuals [51].

Gender diversity awareness, inclusive communication, and understanding the needs of TNBI individuals have revealed a lack of awareness and understanding of TNBI health issues among health care professionals. Communication barriers, limited education on transgender and nonbinary terminology, and inadequate training in cultural competency contribute to misunderstandings and discomfort for these individuals. In addition, the limited understanding of gender diversity and institutional barriers further hinder effective care delivery, highlighting the need for comprehensive education and training on transgender and nonbinary health topics. Participants in this study noted that despite social acceptance, a lack of clear guidance and limited resources leaves them struggling with issues around gender identity. They expressed a fear of misgendering individuals and often avoided discussions on gender identity. Participants mentioned that their training has emphasized focusing on biological sex in treatment, often overlooking the psychological aspects of gender identity.

Various systemic issues, including the lack of awareness among health care professionals, limited mental health support, and gaps in research and specialized training for gender-affirming care, were identified as substantial barriers to providing comprehensive care for TNBI individuals. All participants emphasized the need for increased awareness, research, and resources to address these gaps and promote health equity and social justice for TNBI individuals.

Education and training needs for health care professionals on TNBI care highlighted the importance of comprehensive education on gender diversity and transgender and nonbinary health issues for health care professionals. Participants from all 4 countries expressed a desire for tailored resources and training modules specifically designed to enhance cultural competency and provide gender-affirming care. Moreover, the lack of specialized training and guidelines for gender-affirming care underscores the need for systemic changes within health care institutions to support the professional development of health care professionals in this area.

Educational resources and training tools for TNBI care identified interactive case studies, peer support forums, and comprehensive training modules as valuable tools for promoting active learning and skill development in transgender and nonbinary health. These resources facilitate knowledge sharing, collaboration, and reflection among health care professionals, ultimately enhancing the quality of care for these individuals.

Challenges and design considerations for eHealth tools integration and evaluating eHealth impact presented both opportunities and challenges in health care delivery. While eHealth tools have the potential to streamline communication and improve patient–health care professional relationships, participants identified various barriers, including time constraints, cultural change, and technical issues. Overcoming these barriers requires proactive efforts to address resistance to change, invest in IT infrastructure, and prioritize user-friendly design and data security.

Subanalysis of Participants With Experience in Providing Health Care to TNBI Individuals and Those Without Such Experience

Most of the participants without experience in working with TNBI individuals felt ill-equipped to address their unique health care needs. They struggled to build trusting relationships due to a limited understanding of cultural identities and health care requirements, citing that their education focused solely on binary male and female individuals' perspectives. Participants from Sweden noted that, despite living in an open society, gaps in transgender and nonbinary-specific training persist.

Both groups emphasized that delivering trustworthy care requires sufficient research and evidence-based practice to guide clinical decisions. There was no perceived difference in mental health support for TNBI individuals across Sweden, India, the United States, and the United Kingdom. All participants—regardless of prior experience—reported a lack of comprehensive training and instructions on TNBI health, with a shared belief that education is crucial for fostering an inclusive and affirming health care system. Participants across countries agreed that exchanging knowledge, ongoing training, and access to gender-affirming resources are essential for equipping health care professionals. While many were optimistic about eHealth tools improving communication with patients and continued education, they noted challenges such as additional training requirements, adapting to new workflows, and the importance of user-friendly designs.

The findings of this study align with and extend upon existing research in the field of health care education, particularly regarding the communication needs of health care professionals when interacting with TNBI individuals.

Hughto et al [33] highlighted the importance of addressing societal stigma and vulnerability in health care settings, particularly for TNBI individuals. Similarly, our research underscores the detrimental effects of societal stigma and cultural norms on TNBI mental health and well-being, as well

XSL•FO
RenderX

as the barriers to accessing health care services due to fear of discrimination and misgendering.

Furthermore, our study supports previous research conducted by Grant et al [22] on the challenges faced by health care professionals in delivering inclusive and affirming care to transgender patients. Consistent with these findings, our participants expressed concerns about communication barriers, limited understanding of gender diversity, and gaps in specialized training and guidelines for gender-affirming care. These challenges highlight the need for comprehensive education and training initiatives to enhance health care professionals' cultural competency and sensitivity to transgender and nonbinary health issues.

A scoping review of transgender health training in internal medicine and subspecialty residency programs identified significant gaps in medical education, emphasizing the need for clearly defined objectives to prepare health care professionals for competent and affirming transgender care [52]. Similarly, a systematic review of educational interventions for medical students and residents working with sexual and gender-minority patients demonstrated the effectiveness of structured programs in improving knowledge, attitudes, confidence, and skills, highlighting the importance of implementing comprehensive training to bridge these gaps [53]. Davidge-Pitts et al [54] studied the importance of comprehensive training and educational resources to address the gaps in health care professional's knowledge and skills related to TNBI health issues. Similarly, our study underscores the significance of tailored educational tools in enhancing health care professional's understanding and competency in providing gender-affirming care. In addition, the Transgender Education for Affirmative and Competent HIV and Healthcare Program further highlights the impact of structured educational initiatives in fostering gender-affirming knowledge, perceived competency, and inclusive practice behaviors among health care providers [55]. These findings align with our study's recommendation for interactive, solution-oriented tools to promote ongoing skill development and professional collaboration. In addition, research on barriers to transgender and gender-diverse care highlights several challenges, including the absence of clear guidelines; extended waiting times; a shortage of specialist centers; insufficient training in transgender and gender-diverse health; and technical, cultural, and social obstacles [26]. These findings align with and reinforce the results of our study.

Our study supports the notion that interactive case studies and peer support forums, integrated within an educational tool, promote active learning, cultural competency, and professional networking among health care professionals. These features facilitate ongoing learning and skill development, fostering a collaborative environment conducive to improving TNBI health care delivery. Moreover, the challenges identified in the integration and use of eHealth tools, as discussed in our study, echo the findings of previous research by Light et al [56]. Time constraints, cultural barriers, and technical issues have consistently been recognized as barriers to the adoption of digital health technologies in clinical practice by the participants.

In contrast to earlier studies conducted by Mansh et al [57], which primarily focused on identifying gaps and challenges in TNBI health care education, our study advances the field by proposing a solution-oriented approach. This study provides actionable recommendations for improving TNBI health care education and training initiatives. This shift toward solution-focused research aligns with the broader goal of enhancing health care delivery and patient outcomes through innovative educational interventions. The findings of this study contribute to the growing body of literature on educational interventions for health care professionals and their implications for TNBI health.

Strengths and Limitations of the Study

The main strength of this study is the use of a qualitative research approach, which allowed for an in-depth exploration of the challenges faced by health care professionals when communicating with TNBI individuals, facilitating a comprehensive understanding of the nuances and complexities surrounding this topic.

In addition, the study included participants from various health care backgrounds and geographic locations, enhancing the richness and diversity of perspectives gathered. This diversity contributed to a more comprehensive analysis of the educational eHealth tools' requirements and potential impact across different health care contexts. The diverse backgrounds of the participants offered a wide range of experiences and provided a more comprehensive understanding of the issues. While our study provides a broad overview of the needs across different medical professions, we recognize the importance of conducting more in-depth studies for each profession. Future research should focus on the specific needs and challenges faced by each medical specialty when working with TNBI individuals. As qualitative studies do not guarantee the generalizability of the results, we argue that the results of our study are transferable to other contexts.

We believe that the number of interviews conducted with health care professionals was sufficient to achieve data saturation across the overall sample. This indicates that we obtained a sufficient breadth and depth of information to address the research objectives. This ensured that a thorough exploration of the topic was achieved without the need for additional participants [37]. It is also important to note that due to the diversity of countries and health care professions in the sample, combined with the relatively low number of participants from each profession and country, data saturation specific to each subgroup may not have been fully achieved. This limitation could affect the transferability of our findings, particularly for specific roles or cultural contexts. Future research could benefit from a more focused examination of these subgroups to deepen the understanding of context-specific nuances. In addition, conducting additional interviews with health care professionals from settings not represented in this study could enrich findings within the subject.

Another limitation of the study is that the interview guide was not pilot-tested. However, the authors carefully designed the guide and conducted multiple meetings to review and refine it. In addition, during the interviews, we maintained a flexible and

XSL•FO
RenderX

adaptive approach, making minor adjustments needed to better capture health care professionals' experiences.

Conclusions

The study aimed to explore health care professionals' perspectives on education, awareness, and preferences for digital educational resources to support TNBI care. The results provided valuable insights into the barriers health care professionals encounter when providing care to TNBI. The study identified key gaps in health care professionals' understanding of gender diversity, cultural competency, and the need for inclusive communication. In addition, the study emphasized the importance of specialized training and the integration of user-friendly eHealth tools to improve the relationships between health care professionals with TNBI individuals.

eHealth tools play a significant role in enhancing patient–health care professional relationships, improving access to care, and promoting patient engagement in health care. Despite the challenges associated with their integration, health care professionals acknowledged their potential to facilitate more efficient, patient-centered care delivery.

Addressing the identified barriers and challenges through targeted interventions, such as providing training and support for health care professionals, investing in user-friendly design and data security, and promoting cultural competence in providing health care for TNBI individuals, is essential.

In conclusion, this study contributes to the growing literature on eHealth interventions in TNBI health care and sets the stage for future research and practice initiatives aimed at leveraging technology to improve health outcomes and reduce health disparities for these individuals.

Acknowledgments

The authors would like to thank all the study participants for their time and valuable contributions to this study. The authors received no specific funding for this work.

Generative artificial intelligence technologies, including Grammarly [58] and OpenAI's ChatGPT 3.5 [59], were used during manuscript preparation to improve grammatical accuracy and enhance clarity in the text. These tools were not used to refine or alter the original qualitative data or inform the thematic analysis. The qualitative data were analyzed in their original form, ensuring accuracy and consistency with participants' responses. The authors reviewed and edited all aspects related to grammatical accuracy and textual clarity to ensure alignment with the study's objectives and maintain content integrity, taking full responsibility for the content of the publication.

Authors' Contributions

Both SK and ND were involved in the study design. SK conducted the data collection and performed the initial data analysis, which was subsequently reviewed and refined by ND. Both the authors were actively involved in writing and reviewing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[PDF File (Adobe PDF File), 72 KB - mededu_v11i1e67993_app1.pdf]

References

1. Gooren L. The biology of human psychosexual differentiation. Horm Behav 2006 Nov;50(4):589-601. [doi: 10.1016/j.yhbeh.2006.06.011] [Medline: 16870186]
2. Understanding transgender people, gender identity and gender expression. American Psychological Association. URL: https://www.apa.org/topics/lgbtq/transgender-people-gender-identity-gender-expression [accessed 2024-04-29]
3. Fiani CN, Han HJ. Navigating identity: experiences of binary and non-binary transgender and gender non-conforming (TGNC) adults. Int J Transgend 2019;20(2-3):181-194 [FREE Full text] [doi: 10.1080/15532739.2018.1426074] [Medline: 32999605]
4. Sumerau JE, Mathers LA. America through Transgender Eyes. Lanham, MD: Rowman & Littlefield; 2019.
5. Gates GJ, Distinguished W. How many people are lesbian, gay, bisexual, and transgender? Williams Institute. 2011. URL: https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-People-LGBT-Apr-2011.pdf [accessed 2024-04-29]
6. Conron KJ, Scott G, Stowell GS, Landers SJ. Transgender health in Massachusetts: results from a household probability sample of adults. Am J Public Health 2012 Jan;102(1):118-122. [doi: 10.2105/ajph.2011.300315]
7. Spizzirri G, Eufrásio R, Lima MC, de Carvalho Nunes HR, Kreukels BP, Steensma TD, et al. Proportion of people identified as transgender and non-binary gender in Brazil. Sci Rep 2021 Jan 26;11(1):2240 [FREE Full text] [doi: 10.1038/s41598-021-81411-4] [Medline: 33500432]

XSL•FO
RenderX

8. The struggle of trans and gender-diverse persons. Office of the United Nations High Commissioner for Human Rights (OHCHR). URL: https://www.ohchr.org/en/special-procedures/ie-sexual-orientation-and-gender-identity/struggle-trans-and-gender-diverse-persons [accessed 2024-05-12]
9. Kearns S, Hardie P, O'Shea D, Neff K. Instruments used to assess gender-affirming healthcare access: a scoping review. PLoS One 2024 Jun 3;19(6):e0298821 [FREE Full text] [doi: 10.1371/journal.pone.0298821] [Medline: 38829881]
10. Kearns S, Kroll T, O'Shea D, Neff K. Experiences of transgender and non-binary youth accessing gender-affirming care: a systematic review and meta-ethnography. PLoS One 2021 Sep 10;16(9):e0257194 [FREE Full text] [doi: 10.1371/journal.pone.0257194] [Medline: 34506559]
11. Jessani A, Berry-Moreau T, Parmar R, Athanasakos A, Prodger JL, Mujugira A. Healthcare access and barriers to utilization among transgender and gender diverse people in Africa: a systematic review. BMC Glob Public Health 2024 Jun 27;2(1):44 [FREE Full text] [doi: 10.1186/s44263-024-00073-2] [Medline: 38948028]
12. Renner J, Blaszcyk W, Täuber L, Dekker A, Briken P, Nieder TO. Barriers to accessing health care in rural regions by transgender, non-binary, and gender diverse people: a case-based scoping review. Front Endocrinol (Lausanne) 2021 Nov 18;12:717821 [FREE Full text] [doi: 10.3389/fendo.2021.717821] [Medline: 34867775]
13. Ahuja TK, Goel AD, Gupta MK, Joshi N, Choudhary A, Suman S, et al. Health care needs and barriers to care among the transgender population: a study from western Rajasthan. BMC Health Serv Res 2024 Aug 26;24(1):989 [FREE Full text] [doi: 10.1186/s12913-024-11010-2] [Medline: 39187822]
14. Gillespie C. 7 major health disparities affecting the LGBTQ+ community. Health. URL: https://www.health.com/mind-body/lgbtq-health-disparities [accessed 2025-01-08]
15. Health disparities and equitable access to health care persist with transgender adults. The American Heart Association. URL: https://newsroom.heart.org/news/health-disparities-and-equitable-access-to-health-care-persist-with-transgender-adults [accessed 2024-12-20]
16. Min L. Health disparities in the LGBTQ community. Int J Arts Humanit Soc Sci Stud 2023:1-10 [FREE Full text]
17. Hughes C. A review of: “Transgender Health and HIV Prevention: Needs Assessment Studies from Transgender Communities across the United States. Edited by Walter Bockting and Eric Avery". J HIV AIDS Soc Serv 2008 Oct 12;7(3):305-307. [doi: 10.1080/15381500802309712]
18. Melendez RM, Pinto R. 'It's really a hard life': love, gender and HIV risk among male-to-female transgender persons. Cult Health Sex 2007;9(3):233-245 [FREE Full text] [doi: 10.1080/13691050601065909] [Medline: 17457728]
19. Nemoto T, Sausa LA, Operario D, Keatley J. Need for HIV/AIDS education and intervention for MTF transgenders: responding to the challenge. J Homosex 2006;51(1):183-202. [doi: 10.1300/J082v51n01_09] [Medline: 16893831]
20. Lombardi EL, Wilchins RA, Priesing D, Malouf D. Gender violence: transgender experiences with violence and discrimination. J Homosex 2002 Mar 26;42(1):89-101. [doi: 10.1300/J082v42n01_05]
21. Bockting WO, Robinson BE, Forberg J, Scheltema K. Evaluation of a sexual health approach to reducing HIV/STD risk in the transgender community. AIDS Care 2005 Apr 27;17(3):289-303. [doi: 10.1080/09540120412331299825] [Medline: 15832877]
22. Grant JM, Mottet LA, Tanis J, Harrison J, Herman JL, Keisling M. Injustice at every turn: a report of the national transgender discrimination survey. National Center for Transgender Equality and National Gay and Lesbian Task Force. 2011. URL: https://transequality.org/sites/default/files/docs/resources/NTDS_Report.pdf [accessed 2024-04-29]
23. Soled KR, Dimant OE, Tanguay J, Mukerjee R, Poteat T. Interdisciplinary clinicians' attitudes, challenges, and success strategies in providing care to transgender people: a qualitative descriptive study. BMC Health Serv Res 2022 Sep 08;22(1):1134 [FREE Full text] [doi: 10.1186/s12913-022-08517-x] [Medline: 36076288]
24. Reisner SL, Bailey Z, Sevelius J. Racial/ethnic disparities in history of incarceration, experiences of victimization, and associated health indicators among transgender women in the U.S. Women Health 2014;54(8):750-767 [FREE Full text] [doi: 10.1080/03630242.2014.932891] [Medline: 25190135]
25. Canvin L, Twist J, Solomons W. How do mental health professionals describe their experiences of providing care for gender diverse adults? A systematic literature review. Psychol Sex 2021 Apr 23;13(3):717-741 [FREE Full text] [doi: 10.1080/19419899.2021.1916987]
26. Mikulak M, Ryan S, Ma R, Martin S, Stewart J, Davidson S, et al. Health professionals’ identified barriers to trans health care: a qualitative interview study. Br J Gen Pract 2021 Jun 15;71(713):e941-e947. [doi: 10.3399/bjgp.2021.0179]
27. Agapoff J. Supporting and understanding non-binary and gender diverse youth: a physician’s view. Child Adolesc Psychiatry Ment Health 2024 Aug 24;18(1):105 [FREE Full text] [doi: 10.1186/s13034-024-00798-w] [Medline: 39182103]
28. Ilango TS, Karthikeyan S, Sumithra Devi S, Arumuganathan S, Usaid S, Sethumadhavan V. An online survey of education, knowledge and attitude toward homosexuality in adults. Indian J Soc Psychiatry 2020;36(4):344. [doi: 10.4103/ijsp.ijsp_43_20]
29. Dubin SN, Nolan IT, Streed Jr CG, Greene RE, Radix AE, Morrison SD. Transgender health care: improving medical students' and residents' training and awareness. Adv Med Educ Pract 2018;9:377-391 [FREE Full text] [doi: 10.2147/AMEP.S147183] [Medline: 29849472]

XSL•FO
RenderX

30. Tollemache N, Shrewsbury D, Llewellyn C. Que(e) rying undergraduate medical curricula: a cross-sectional online survey of lesbian, gay, bisexual, transgender, and queer content inclusion in UK undergraduate medical education. BMC Med Educ 2021 Feb 12;21(1):100 [FREE Full text] [doi: 10.1186/s12909-021-02532-y] [Medline: 33579262]
31. Persson Tholin J, Broström L. Transgender and gender diverse people's experience of non-transition-related health care in Sweden. Int J Transgend 2018 May 23;19(4):424-435 [FREE Full text] [doi: 10.1080/15532739.2018.1465876]
32. Nahata L, Quinn GP, Caltabellotta NM, Tishelman AC. Mental health concerns and insurance denials among transgender adolescents. LGBT Health 2017 Jun;4(3):188-193. [doi: 10.1089/lgbt.2016.0151] [Medline: 28402749]
33. Hughto JM, Reisner SL, Pachankis JE. Transgender stigma and health: a critical review of stigma determinants, mechanisms, and interventions. Soc Sci Med 2015 Dec;147:222-231 [FREE Full text] [doi: 10.1016/j.socscimed.2015.11.010] [Medline: 26599625]
34. Bradford J, Reisner SL, Honnold JA, Xavier J. Experiences of transgender-related discrimination and implications for health: results from the Virginia Transgender Health Initiative Study. Am J Public Health 2013 Oct;103(10):1820-1829. [doi: 10.2105/AJPH.2012.300796] [Medline: 23153142]
35. Do TT, Nguyen AT. 'They know better than we doctors do': providers' preparedness for transgender healthcare in Vietnam. Health Sociol Rev 2020 Mar 29;29(1):92-107. [doi: 10.1080/14461242.2020.1715814] [Medline: 33411663]
36. Allen LR, Adams N, Ashley F, Dodd C, Ehrensaft D, Fraser L, et al. Principlism and contemporary ethical considerations for providers of transgender health care. Int J Transgend Health 2024 Jan 19:1-19 [FREE Full text] [doi: 10.1080/26895269.2024.2303462]
37. Tamim RM, Bernard RM, Borokhovski E, Abrami PC, Schmid RF. What forty years of research says about the impact of technology on learning. Rev Educ Res 2011 Mar 01;81(1):4-28. [doi: 10.3102/0034654310393361]
38. Davies RS, West RE. Technology integration in schools. In: Spector JM, Merrill MD, Elen J, Bishop MJ, editors. Handbook of Research on Educational Communications and Technology. Cham, Switzerland: Springer; 2014:841-853.
39. Proenca L, Mendes JJ, Botelho J, Machado V. E-learning and Digital Training in Healthcare Education: Current Trends and New Challenges. New York, NY: MDPI; 2023.
40. eHealth. World Health Organization. URL: http://www.emro.who.int/health-topics/ehealth/ [accessed 2024-04-29]
41. Young J, Gregory J, Rojas M, Justin G, Kalir T. Transgender healthcare: development of an illustrated elearning tool for medical education. MedEdPublish (2016) 2021;10(1):159 [FREE Full text] [doi: 10.15694/mep.2021.000159.1] [Medline: 38486569]
42. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19 Suppl 3(Suppl 3):i68-i74 [FREE Full text] [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
43. Flick U, von Kardoff E, Steinke I. A Companion to Qualitative Research. Thousand Oaks, CA: Sage Publications; 2004.
44. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018 Mar;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
45. Braun V, Clarke VZ. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
46. Saladana J. The Coding Manual for Qualitative Researchers. 2nd edition. Thousand Oaks, CA: Sage Publications; 2013.
47. The common rule. U.S. Department of Health & Human Services. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html [accessed 2024-04-29]
48. UK policy framework for health and social care research. Health Research Authority. URL: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/uk-policy-framework-health-social-care-research/ [accessed 2024-04-29]
49. National ethical guidelines for biomedical and health research involving human participants. Indian Council of Medical Research. URL: https://ethics.ncdirindia.org/asset/pdf/ICMR_National_Ethical_Guidelines.pdf [accessed 2024-04-29]
50. The NHS ends the "gender-affirmative care model" for youth in England. Society for Evidence-Based Gender Medicine. URL: https://segm.org/England-ends-gender-affirming-care [accessed 2024-04-29]
51. Bans on gender-affirming care for transgender youth: a 2023 legislative update. Williams Institute. URL: https://williamsinstitute.law.ucla.edu/wp-content/uploads/Trans-Youth-Health-Bans-Mar-2023.pdf [accessed 2024-04-29]
52. Ahmad T, Robinson L, Uleryk E, Yu C. Trans health training objectives: a scoping review. Clin Teach 2024 Feb 08;21(1):e13673. [doi: 10.1111/tct.13673] [Medline: 37806669]
53. Cooper RL, Ramesh A, Radix AE, Reuben JS, Juarez PD, Holder CL, et al. Affirming and inclusive care training for medical students and residents to reduce health disparities experienced by sexual and gender minorities: a systematic review. Transgend Health 2023 Aug 01;8(4):307-327 [FREE Full text] [doi: 10.1089/trgh.2021.0148] [Medline: 37525832]
54. Davidge-Pitts CJ, Nippoldt TB, Natt N. Endocrinology fellows' perception of their confidence and skill level in providing transgender healthcare. Endocr Pract 2018 Dec;24(12):1038-1042. [doi: 10.4158/ep-2018-0307]
55. Lacombe-Duncan A, Logie CH, Persad Y, Leblanc G, Nation K, Kia H, et al. Implementation and evaluation of the 'Transgender Education for Affirmative and Competent HIV and Healthcare (TEACHH)' provider education pilot. BMC Med Educ 2021 Nov 04;21(1):561 [FREE Full text] [doi: 10.1186/s12909-021-02991-3] [Medline: 34732178]

XSL•FO
RenderX

56. Light AD, Obedin-Maliver J, Sevelius JM, Kerns JL. Transgender men who experienced pregnancy after female-to-male gender transitioning. Obstet Gynecol 2014 Dec;124(6):1120-1127 [FREE Full text] [doi: 10.1097/AOG.0000000000000540] [Medline: 25415163]
57. Mansh M, White W, Gee-Tong L, Lunn MR, Obedin-Maliver J, Stewart L, et al. Sexual and gender minority identity disclosure during undergraduate medical education: "in the closet" in medical school. Acad Med 2015 May;90(5):634-644. [doi: 10.1097/ACM.0000000000000657] [Medline: 25692563]
58. Grammarly. URL: https://www.grammarly.com/ [accessed 2025-03-04]
59. ChatGPT 3.5. OpenAI. URL: https://chatgpt.com/g/g-F00faAwkE-open-a-i-gpt-3-5 [accessed 2025-03-04]

Abbreviations

ICT: information and communication technology
TNBI: transgender, nonbinary, and intersex

Edited by B Lesselroth; submitted 25.10.24; peer-reviewed by H Pilabré, J Wells; comments to author 03.12.24; revised version received 14.01.25; accepted 16.02.25; published 06.03.25.

Please cite as:
Katta S, Davoody N
Exploring Health Care Professionals' Perspectives on Education, Awareness, and Preferences for Digital Educational Resources to Support Transgender, Nonbinary, and Intersex Care: Interview Study
JMIR Med Educ 2025;11:e67993
URL: https://mededu.jmir.org/2025/1/e67993
doi:10.2196/67993
PMID:40053815

©Sravya Katta, Nadia Davoody. Originally published in JMIR Medical Education (https://mededu.jmir.org), 06.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluation of an Interdisciplinary Educational Program to Foster Learning Health Systems: Education Evaluation

Sathana Dushyanthen[1], PhD; Nadia Izzati Zamri[2], MPharm; Wendy Chapman[1], PhD; Daniel Capurro[1,3], MD, PhD; Kayley Lyons[1], PharmD, PhD

[1]Centre for Digital Transformation of Health, University of Melbourne, Carlton, Australia
[2]Faculty of Pharmacy and Pharmaceutical Sciences, Monash University, Parkville, Australia
[3]School of Computing and Information Systems, University of Melbourne, Melbourne, Australia

Corresponding Author:
Kayley Lyons, PharmD, PhD
Centre for Digital Transformation of Health, University of Melbourne, Carlton, Australia

Abstract

Background: Learning health systems (LHS) have the potential to use health data in real time through rapid and continuous cycles of data interrogation, implementing insights to practice, feedback, and practice change. However, there is a lack of an appropriately skilled interprofessional informatics workforce that can leverage knowledge to design innovative solutions. Therefore, there is a need to develop tailored professional development training in digital health, to foster skilled interprofessional learning communities in the health care workforce in Australia.

Objective: This study aimed to explore participants' experiences and perspectives of participating in an interprofessional education program over 13 weeks. The evaluation also aimed to assess the benefits, barriers, and opportunities for improvements and identify future applications of the course materials.

Methods: We developed a wholly online short course open to interdisciplinary professionals working in digital health in the health care sector. In a flipped classroom model, participants (n=400) undertook 2 hours of preclass learning online and then attended 2.5 hours of live synchronous learning in interactive weekly Zoom workshops for 13 weeks. Throughout the course, they collaborated in small, simulated learning communities (n=5 to 8), engaging in various activities and problem-solving exercises, contributing their unique perspectives and diverse expertise. The course covered a number of topics including background on LHS, establishing learning communities, the design thinking process, data preparation and machine learning analysis, process modeling, clinical decision support, remote patient monitoring, evaluation, implementation, and digital transformation. To evaluate the purpose of the program, we undertook a mixed methods evaluation consisting of pre- and postsurveys rating scales for usefulness, engagement, value, and applicability for various aspects of the course. Participants also completed identical measures of self-efficacy before and after (n=200), with scales mapped to specific skills and tasks that should have been achievable following each of the topics covered. Further, they undertook voluntary weekly surveys to provide feedback on which aspects to continue and recommendations for improvements, via free-text responses.

Results: From the evaluation, it was evident that participants found the teaching model engaging, useful, valuable, and applicable to their work. In the self-efficacy component, we observed a significant increase ($P<.001$) in perceived confidence for all topics, when comparing pre- and postcourse ratings. Overall, it was evident that the program gave participants a framework to organize their knowledge and a common understanding and shared language to converse with other disciplines, changed the way they perceived their role and the possibilities of data and technologies, and provided a toolkit through the LHS framework that they could apply in their workplaces.

Conclusions: We present a program to educate the health workforce on integrating the LHS model into standard practice. Interprofessional collaborative learning was a major component of the value of the program. This evaluation shed light on the multifaceted challenges and expectations of individuals embarking on a digital health program. Understanding the barriers and facilitators of the audience is crucial for creating an inclusive and supportive learning environment. Addressing these challenges will not only enhance participant engagement but also contribute to the overall success of the program and, by extension, the broader integration of digital health solutions into health care practice and, ultimately, patient outcomes.

(JMIR Med Educ 2025;11:e54152) doi:10.2196/54152

XSL•FO
RenderX

KEYWORDS

continuing professional development; learning health system; flipped classroom; digital health informatics; data science; health professions education; interdisciplinary education; foster; foster learning; health data; design; innovative; innovative solution; health care workforce; Australia; real time; teaching model

Introduction

As health care delivery evolves in complexity and scope, the need for systems that promote continuous learning and adaptation is paramount. The learning health systems (LHS) concept has emerged as a transformative framework that bridges clinical practice with ongoing research, ensuring that health care institutions remain at the forefront of scientific and patient-centered care advancements [1,2]. Central to the LHS paradigm is the notion that data contribute to a broader system of knowledge and is used to refine care practices in real time [1,3]. Achieving this idea requires an interdisciplinary workforce adept in information systems, informatics, data interrogation, quality improvement and implementation methods, and system-based practice, to be able to use the existing data to inform future care [3]. Moreover, health care transformation such as this requires the skills of various professions working together towards solving these complex problems [4,5].

While there are previous studies that have described their LHS-focused programs, few have robustly evaluated the purpose of their implementations. Furthermore, other programs have focused on specific cohorts of participants such as PhD students [6], postdoctoral students [7,8], and clinical fellows [9-11] in the United States [6,9-11] and Canada [7]. Our study adds new insights to the literature given the interprofessional nature of the program, as well as its design (flipped classroom, working groups) and delivery (wholly online). To our knowledge, few programs have involved teaching a structured curriculum [8,12], while other programs have involved mainly project-based work and on-the job learning [7,10,13,14].

For such a dynamic and integrated approach to take root, educating the next generation of health care professionals about LHS principles is crucial. While the theoretical foundation of LHS has been well established, there has been a paucity of research evaluating the efficacy and impact of educational interventions centered on LHS. We developed a 13-week short course called Applied Learning Health Systems, which commenced in September 2021 and has now been running for 2 years [15]. The program is open to all professionals working in the health care setting—clinical and nonclinical—and focuses on interdisciplinary work; the LHS concept can be taught to both digital health and informatics generalists and specialists, clinicians and nonclinicians, front-line workers, and upper management [15].

As institutions increasingly incorporate LHS into their curricula, understanding the nuances of its educational translation becomes vital. This research aims to evaluate the motivations, experiences, and perceptions of participants learning in a collaborative learning environment, as well as the effectiveness, confidence, applicability, challenges, and outcomes of LHS education, providing insights that will shape pedagogical strategies and potentially influence the future of health care education.

The purpose of this paper is to explore participants' experiences and perspectives of participating in a wholly online interprofessional education program. This evaluation also aimed to assess the benefits, barriers, and opportunities for improvements, and identify future applications of the course materials to the participants' workplace endeavors. We will also discuss the implementation, feasibility, and outcomes of the program which aimed to foster LHS skills in the Australian health care workforce through didactic coursework, interactive workshops, and collaborative learning. By describing our program and its 2-year evaluation, we believe that current and future educators can learn from our experience when building their own programs. Additionally, our paper will contribute to the emerging education literature on how to foster LHS through workforce development and education. Compared to previous publications on LHS education programs, we are contributing novel insights to this literature through new perspectives based on our location (ie, Australia), the health system data infrastructure (ie, recent electronic medical record [EMR] implementations and digital immaturity), and our participants (ie, diverse interprofessionals). While we have had early successes, we also wish to highlight the obstacles we encountered and how we refined our approach in response. Our results will be valuable to other educators as they consider similar endeavors.

Methods

Study Design and Recruitment

We undertook a mixed methods study consisting of both quantitative and qualitative data collection methods. Surveys were conducted precourse, throughout teaching, and postcourse. The surveys consisted of metric scales, qualitative scales, and open free-text boxes. Participation in the research project was via opt-out. Therefore, all enrolled participants were eligible to participate in the project voluntarily, unless they chose not to. There were several modes of recruitment for the course itself. These included reaching out to existing precinct partners who undertook internal expression of interest recruitment processes to sponsor a number of places, social media advertising on X and LinkedIn, Google search search engine optimization, and university students undertaking electives or formal university-accredited certificates.

Ethical Considerations

This study was approved by the University of Melbourne ethics committee (project ID 22641). In certain parts of the study, participants had the option to opt out (eg, surveys) or provide consent to participate (eg, interviews). In terms of informed consent, participants were provided with a plain language statement describing the purpose and design of the study.

XSL•FO
RenderX

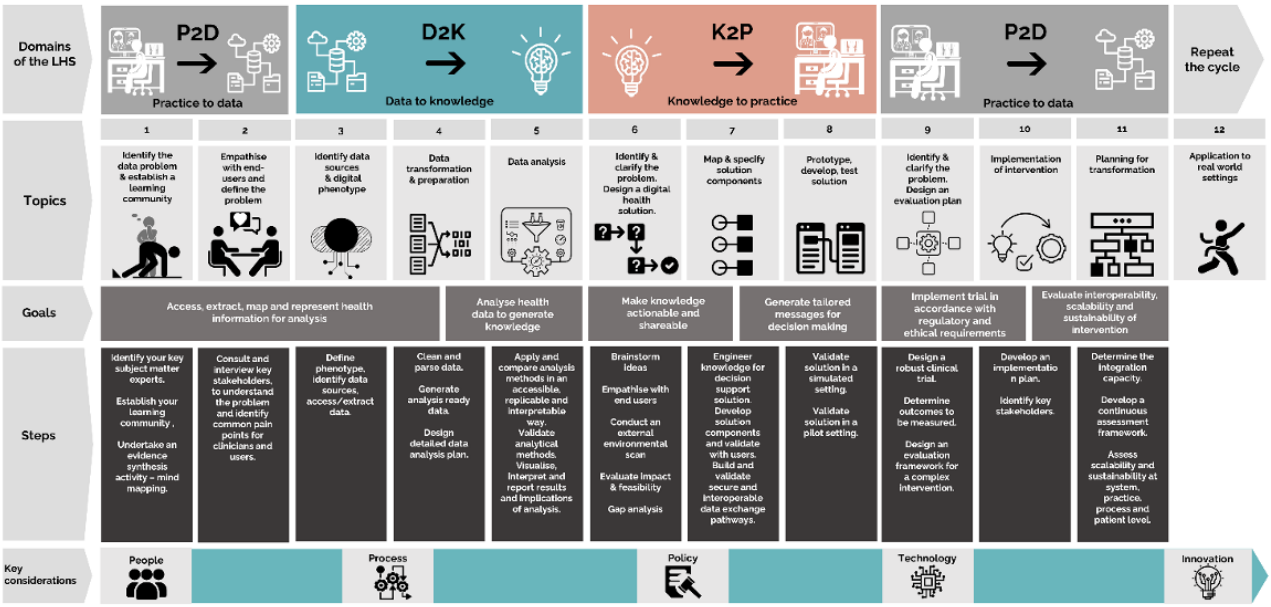
Participants were notified that participation was voluntary and were given the option to opt out. For privacy and confidentiality, data were completely deidentified and only aggregate data were analyzed and presented. Data were housed on secure University of Melbourne single sign-on Qualtrics servers and restricted access to OneDrive servers. As participation was completely voluntary, no compensation was provided to participants; however, participants in the pilot version of the course were given free scholarship admission in return for their feedback.

The Program

The LHS short course was created by the University of Melbourne Centre for Digital Transformation of Health, a high-research academic institution with existing partnerships with local and regional hospitals and primary care networks. The course has been delivered 5 times to 400 participants. Each iteration of the short course involved a 13-week online course revolving around LHS and was delivered wholly online, by diverse instructors, in a flipped classroom learning format. Participants were from a range of backgrounds, including working professionals in health care, PhD research students, masters-level university students, and consumers. The course structure involves 3 hours of weekly individual asynchronous prereadings, followed by 2.5 hours of weekly workshops. Each week participants work through activities associated with a threaded diabetes case scenario, in their assigned interprofessional working group [15].

We mapped the stages of the LHS system onto a swim lane diagram and created specific learning objectives for skills and knowledge at each stage, which were then operationalized into the diabetes scenario. Filling in this swim lane and competency map required knowledge from many disciplines, including data science and biostatistics, standards, user-centered design, change management, workflow mapping, app development, implementation science, and evaluation as well as expertise in the clinical domain and in how the Australian health system works. No single person could effectively design the course we developed, which posed challenges and opportunities for curriculum development. Using the LHS cycle enabled curriculum designers to join the varied subject matter expertise, by mapping it to an agreed framework. Details of the full course design, development, and curriculum outline are published elsewhere (Figure 1) [16].

Figure 1. The Applied Learning Health Systems short course curriculum [16].



Evaluation Framework

We used the Kirkpatrick model of evaluation [17] to map out our measurements (Table 1). This model is a widely used evaluation framework in education and is used to shift researchers away from simply measuring perceptions and satisfaction. We examined whether participants' attitudes, knowledge, behavior, and professional practice changed as a result. Additionally, we applied a mixed methods approach that included pre- and postsurveys, weekly surveys, and postinterviews.

XSL•FO
RenderX

Table . Application of the Kirkpatrick model of evaluation, adapted from Barr et al [18], to this project.

Level	Details	Evaluation measures and data sources in this project
1	Perception of training among subjects	Pre-, weekly, and postsurveys; postcourse participant interviews
2a	Change in the attitudes of subjects	Pre- and postchange in digital health interest and identity
2b	Change in the knowledge and/or skills of subjects	Pre- and post-self-efficacy changes in specific LHS[a] concepts (skills); pre- and postconcept maps (knowledge; out of scope for this paper)
3	Changes in the behavior of subjects	Postcourse participant interviews (will follow up in 1 year with participant interviews)
4a	Change in professional practice	Postcourse participant interviews (will follow up in 1 year with participant interviews)
4b	Changes in patients' condition	Not applicable

[a]LHS: learning health system.

Pre- and Postcourse Surveys

The pre- and postcourse surveys were developed by using a combination of psychological scales and open-ended questions. The pre- and postcourse surveys included the same self-efficacy scale (100 points; cannot do at all to highly certain can do) [19] which has significant evidence of reliability and validity. We choose to evaluate self-efficacy as it is one of the strongest proxy measures in education to predict actual and future performance, which are more difficult and take longer to measure [20]. The 10 items on the self-efficacy scale were adapted from the material taught in the LHS course and language from the LHS literature (eg, use machine learning algorithms to create a model for predicting a health outcome) [21,22]. The open-ended questions included demographic questions (eg, job title) and questions related to digital health identity development, course benefits, course barriers, what to keep, what to improve, and other suggestions or comments.

Surveys were designed and distributed via Qualtrics. Participants were invited to complete the surveys through emails and the learning management system. Responses to open-ended survey questions were also analyzed through qualitative content analysis. Two coders independently coded the text responses using NVivo (Lumivero) software. Coders met to resolve discrepancies and solidify themes and categories under each research question. The self-efficacy scales were analyzed using a 2-tailed, unpaired *t* test in GraphPad Prism to determine whether there was an improvement in self-efficacy across the 13 LHS concepts.

Weekly Surveys

Over the 12 weeks, participants had the opportunity to provide feedback on the level of engagement, usefulness, value, satisfaction, and areas for improvement in the course content, through participation in weekly surveys. These surveys contained scales (strongly disagree to strongly agree) and ask questions such as "how useful did you find this topic" and "how engaged did you feel" and open boxes for free-text responses. Descriptive statistics such as frequency, mean, and standard deviation will be used to summarize the data from these questions. Completion of these weekly surveys ranged from 2530 participants each week.

Qualitative Coding of Free-Text Responses

To analyze the text response according to our research questions, we first deidentify the transcripts for participant and institution names. The transcripts will be uploaded to NVivo software for qualitative content analysis [23]. A codebook was developed deductively from the literature and inductively from the research data. Two coders independently analyzed the transcripts according to the codebook. The 2 coders met to calculate an interrater agreement rate and resolve any discrepancies. The final codes were synthesized by creating summaries, narratives, and matrices. The final results included coding frequencies, themes, and categories according to the research questions.

Quantitative Statistical Analysis

For descriptive statistics, number of participants and proportion of participants are shown. For rating scales, frequency and proportion are shown. Pre- and postcourse self-efficacy comparisons were undertaken using a 2-tailed, unpaired *t* test. Incomplete or missing data were excluded from the analysis.

Results

Demographics

Thus far, the Applied Learning Health Systems program has had approximately 400 participants from various organizations (health care, government, research or university, industry) and job roles (clinician, researcher, data or information technology [IT], health services management, allied health, EMR implementation, health administration, consumer advocacy) (Table 2). Of the 400 participants, 343 (85.8%) completed the presurvey (week 0) and 200 (50%) completed the postsurvey (week 12). A few participants were lost to follow-up during the final week because they were ill, dropped out due to overcommitment, or did not respond to requests.

XSL•FO
RenderX

Table . Demographics shared by participants in the Applied Learning Health Systems program.

Characteristic		Participants, n (%)
Professional background (n=399)		
	Primary health care	44 (11)
	Tertiary health care	141 (35.3)
	Health services management	29 (7.3)
	Allied health	48 (12)
	Government	10 (2.5)
	Academia or research	73 (18.3)
	Business, IT[a], tech or data analytics	47 (11.8)
	Other	7 (1.8)
Role type (n=343)		
	Clinician (medical)	67 (17)
	Clinician (nursing)	25 (6.4)
	Clinical informatician	22 (5.6)
	Researcher (health services research or public health)	68 (17.3)
	Data analyst	28 (7.1)
	Allied health professional	58 (14.8)
	Health services manager	36 (9.2)
	Quality improvement lead	24 (6.1)
	Consultant or IT professional	19 (4.8)
	EMR[b] implementation team	18 (4.6)
	Health administration	8 (2)
	Consumer advocate	20 (5.1)

[a]IT: information technology.

[b]EMR: electronic medical record.

What Were Participants' Previous Encounters With the LHS Framework?

At the beginning of the course, participants were asked if they had any previous exposure to the LHS framework. Almost one-third of the participants had no previous experience with the LHS concept or any digital health concepts (121/343, 35.3%). Some participants stated that they had previous exposure to digital health and informatics concepts (50/343, 14.6%) through other courses and certifications (27/343, 7.8%), as well as through work-based activities, for example, EMR implementation and optimization (47/343, 13.1%), quality improvement, data interrogation (56/343, 16.3%), and various other health services projects (45/343, 13.1%). Others stated that they had no previous exposure to digital health or LHS concepts (49/343, 14.2%).

What Type of Teaching Approaches Did Participants Perceive as Effective?

Participants were asked to rate the usefulness and engagement of the topic's preclass learning and in-class sessions. In terms of usefulness, the majority found the preclass materials useful (880/956, 92.1%—"the preclass material was excellent and really helped to clarify many of the terms that I had heard people say but not truly understood") and in-class sessions useful (902/955, 94.5%—"analyzing the data during the class was useful and to see it connect with prelearning materials was good"). When asked to rate engagement, the majority found the preclass (881/954, 92.3%) and in-class activities engaging (881/955, 92.3%) (Tables 3-6).

XSL•FO
RenderX

Table . Ratings of usefulness and engagement with preclass learning materials and in-class Zoom sessions. Participants were asked to rate the agreement for usefulness (extremely useless to extremely useful) and engagement (extremely unengaged to engaged), weekly for each topic (1-13).

Questions	Rating							Total, n
	Extremely useless, n (%)	Moderately useless, n (%)	Slightly useless, n (%)	Neither useful nor useless, n (%)	Slightly useful, n (%)	Moderately useful, n (%)	Extremely useful, n (%)	
I found this topic's preclass learning useful (13 topics)	5 (0.5)	39 (4.1)	13 (1.4)	20 (2.1)	210 (22.0)	265 (27.7)	404 (42.3)	956
I found this topic's in-class session useful (13 topics)	2 (0.2)	13 (1.4)	9 (0.9)	28 (2.9)	103 (10.8)	360 (37.7)	440 (46.0)	955
I felt engaged when completing the preclass learning for this topic (13 topics)	5 (0.5)	9 (0.9)	29 (3.0)	30 (3.1)	127 (13.3)	428 (44.9)	326 (34.2)	954
I felt engaged when participating in the topic's in-class session (13 topics)	10 (1.0)	13 (1.4)	17 (1.8)	33 (3.5)	104 (10.9)	349 (36.5)	429 (44.9)	955

Table . Participants' ratings of value pertaining to overall value to personal career development for all topics.

Question	Rating					Total, n
	Highly unvaluable, n (%)	Unvaluable, n (%)	Neutral, n (%)	Valuable, n (%)	Highly valuable, n (%)	
Valuable to your personal career development (13 topics, n=189)	12 (0.6)	13 (0.6)	251 (12.3)	989 (48.5)	776 (38.0)	2041

Table . Participants' ratings of value pertaining to applicability to current workplace role for all topics.

Question	Rating					Total, n
	Highly not applicable, n (%)	Not applicable, n (%)	Neutral, n (%)	Applicable, n (%)	Highly applicable, n (%)	
Applicability to your current workplace role (13 topics, n=189)	64 (3.1)	154 (7.5)	325 (15.9)	837 (41.0)	661 (32.4)	2041

XSL•FO
RenderX

Table . Participants' ratings of value pertaining to overall satisfaction with the quality of the course, recommendation, instructors, and choice to revisit, as well as the value of educational activities (instructors, Zoom workshops, Canvas preclass activities, collaborative learning, the diabetes case scenario, Jupyter Notebooks, and discussion boards).

Questions	Rating							Total, n
	Extremely valueless, n (%)	Moderately valueless, n (%)	Slightly valueless, n (%)	Neither valuable nor valueless, n (%)	Slightly valuable, n (%)	Moderately valuable, n (%)	Extremely valuable, n (%)	
Collaborative learning in the working groups	1 (0.5)	6 (3.3)	3 (1.6)	5 (2.7)	29 (15.9)	63 (34.6)	75 (41.2)	182
Preclass learning activities on Canvas	0 (0)	3 (1.6)	2 (1.1)	2 (1.1)	21 (11.5)	77 (42.3)	77 (42.3)	182
In-class learning (Zoom) sessions	1 (0.5)	2 (1.1)	1 (0.5)	6 (3.3)	17 (9.3)	74 (40.7)	81 (44.5)	182
The diabetes case scenario	1 (0.5)	4 (2.2)	3 (1.6)	10 (5.5)	30 (16.5)	70 (38.5)	64 (35.2)	182
Jupyter Notebooks	0 (0)	11 (6.0)	11 (6.0)	20 (11.0)	53 (29.1)	55 (30.2)	32 (17.6)	182
Canvas learning management system	0 (0)	1 (0.5)	3 (1.6)	12 (6.6)	30 (16.5)	83 (45.6)	53 (29.1)	182
Discussion boards	5 (2.7)	10 (5.5)	10 (5.5)	43 (23.6)	62 (34.1)	39 (21.4)	13 (7.1)	182
The instructors	0 (0)	1 (0.5)	0 (0)	4 (2.2)	8 (4.4)	49 (26.9)	120 (65.9)	182

Responses to the question of satisfaction also yielded highly positive results. For the overall quality of the short course, most agreed that it was of a high standard (178/182, 97.8%), including the instructor quality (175 /182, 96.2%). When asked if they would recommend the short course to a colleague, 89.5% (163/182) said they would. In terms of revisiting the decision to complete it again, 85.1% (154/182) still said they would choose to take the course. When rating the value of the course to their personal career development, a majority found the course valuable (173/200, 86.5%). Participants were also asked to rate the applicability of the course to their day-to-day work, where 73.4% (134/182) found it applicable.

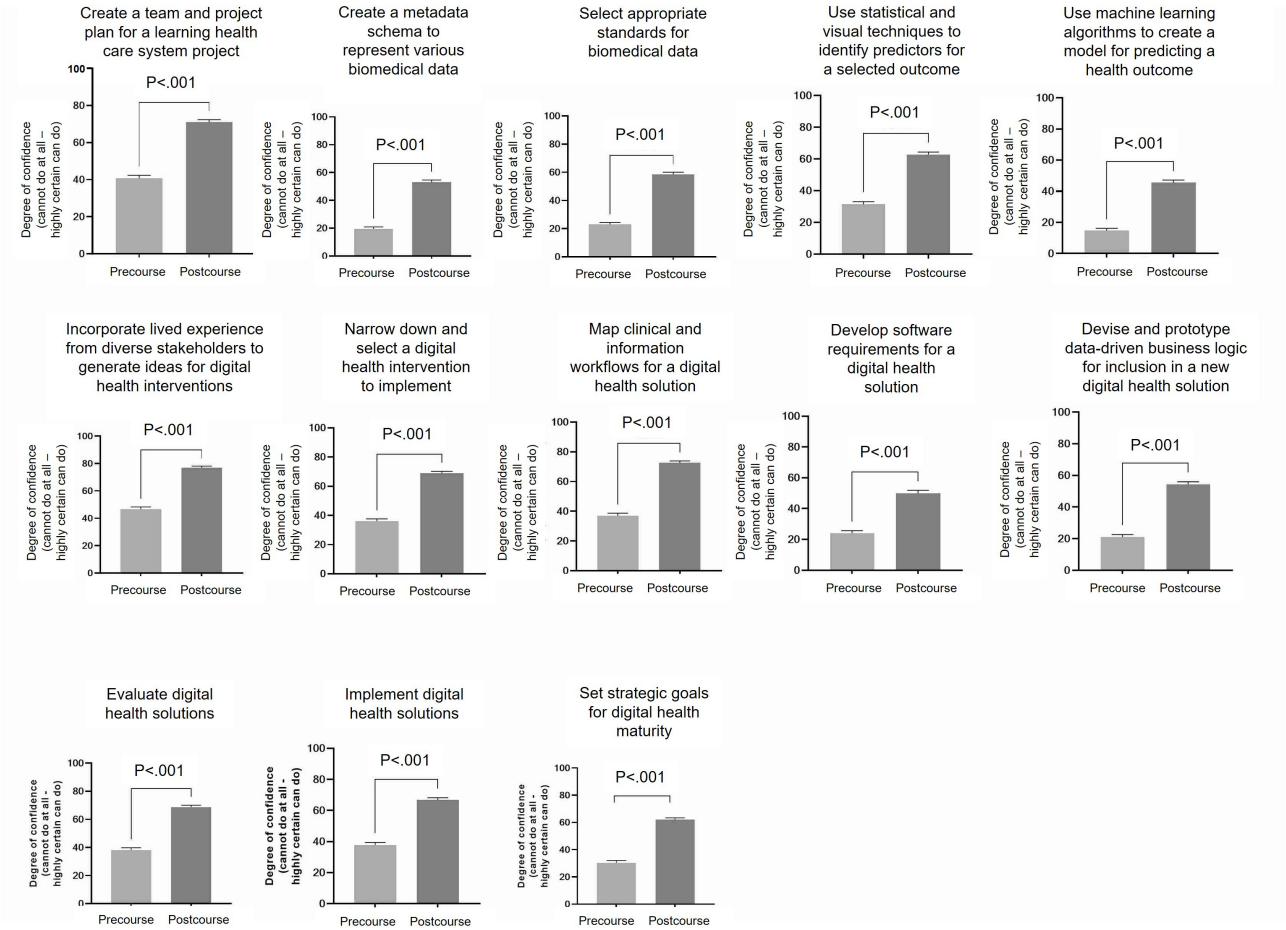
Given the number of facets implemented in the course, we asked participants to rate the value of these various elements. The most highly rated was the instructors: "the speakers were great, and the structure of having a short lecture and then doing an activity then coming back and having another lecture was good," with 92.8% (169/182) finding them moderately or extremely valuable. Next, in-class learning (155/182, 85.2%), preclass learning (154/182, 84.6%), collaborative learning (138/182, 75.8%), the diabetes case scenario (134/182, 73.7%), and the Canvas learning management system platform (136/182, 74.7%) rated similarly. The use of Jupyter Notebooks (87/182, 47.8%), and the discussion boards (52/182, 28.6%) rated lower (Table 6).

How Did Participants' Self-Efficacy for Digital Health Topics Change After the Course?

To explore the change in self-confidence levels pre- and postcourse, participants were surveyed on the key competencies for the 13 topics. Participants completed the same set of ratings at the beginning and at the end of the course, following completion of all the material. For all 13 learning outcomes, there was a statistically significant increase in self-efficacy (n=200, $P<.001$) (Figure 2).

XSL•FO
RenderX

Figure 2. Pre- and postcourse self-efficacy in LHS concepts. Participants rated confidence on a scale of 0 - 100 (0=cannot do at all to 100=highly certain can do). Two-tailed, unpaired *t* test was undertaken (n=296 precourse, n=200 postcourse). Changes from baseline to postcourse confidence are shown for each LHS concept. LHS: learning health system.



How Did Participants' Self-Perceived Role in Digital Health Change?

In the pre- and postsurvey, participants were asked to respond to the open-ended question of "What do you see as your role in digital health?" There were several types of roles that participants perceived themselves embodying postcourse. These included users of digital health or learners; champions, advocates, or change agents; researchers, innovators, or entrepreneurs; leaders, managers, strategic planners, or decision makers; educators or mentors; specialists or implementers; community builders, connectors, facilitators, collaborators, or translators (Table 7).

After the course, there was an increase in participants who viewed their role as an end user or learner and a community builder or facilitator, whereas there was a decrease in those who viewed their role as a champion or advocate and leader in digital health.

Table . Participants' perceived roles in digital health pre- and postcourse (qualitative themes).

	Precourse responses (n=274), n (%)	Postcourse responses (n=228), n (%)
End user of digital health or learner	41 (15.0)	54 (23.7)
Champion, advocate, or change agent	62 (22.6)	37 (16.2)
Researcher, innovator, or entrepreneur	32 (11.7)	30 (13.2)
Leader, manager, strategic planner, or decision maker	38 (13.9)	22 (9.6)
Educators or mentors	13 (4.7)	6 (2.6)
Specialist or implementer	57 (20.8)	41 (18.0)
Community builder, connector, facilitator, collaborator, or translator	31 (11.3)	38 (16.7)

XSL•FO
RenderX

What Did Participants Perceive as the Applications of the Learning in Their Workplace?

There were five main themes that arose for the types of applications that participants foresaw themselves using the course learnings: (1) learning and professional development: "upskilling in the current role, more understanding of the roles of my team members"; (2) using data and undertaking data analysis more effectively: "data mining and improving processes at work"; (3) implementing the LHS framework for digital health interventions: "we are embarking on establishing a data and analytics 3-year plan and we intend to incorporate LHS principals into this strategy"; (4) for undertaking research and quality improvement activities: "I now play a role in learning health networks for Safer Care Victoria, where I believe I could encourage digital health projects focused on quality improvement and patient safety"; and (5) collaborating and sharing knowledge and learnings with colleagues: "I intend to instill the LHS framework into my role, the work that I do and share it with my team" (Table 8).

Table . Participants' anticipated applications of learning in the workplace (qualitative themes).

	Precourse responses (n=338), n (%)	Postcourse responses (n=231), n (%)
Learning and professional development	74 (21.9)	47 (20.4)
Using data and undertaking data analysis	54 (16.0)	43 (18.6)
Implementing digital health solutions with the LHS[a] framework	63 (18.6)	52 (22.5)
Researching and quality improvement	82 (24.3)	62 (26.8)
Collaborating and knowledge sharing	65 (19.2)	27 (11.7)

[a]LHS: learning health system.

What Were the Perceived Benefits of the Program?

Participants were asked to state the benefits of the program. The major themes that arose were learning and knowledge acquisition: "the course material was presented well on Canvas and had a good mix of different learning resources to use," value of collaboration: "the course has been extremely eye-opening and has led me to begin collaborations on digital health projects through contacts made through the course," participant diversity and group work : "being in a group of people with all different work backgrounds and skills coming together with a common interest was really good for tackling the problems to solve in the class," beneficial course structure and content delivery (preclass: "the course material was presented well on Canvas and had a good mix of different learning resources to use" and in-class: "beneficial to be in a diverse group of other health care professionals - I learnt a lot from the robust and engaging discussions on Zoom)," and learning tools, importance of real-world applications: case study and personal work: "applying course concepts to this real-world scenario was instrumental in reinforcing their understanding," appreciation for instructors' diversity, expertise, engagement, and quality: "the instructors were very engaged and passionate about their topics," consumer focus, and focus on data analytics (Table 9).

Table . Beneficial elements of the course.

Theme	Responses (n=295), n (%)
Collaborative group work, diversity, or multidisciplinary approach	63 (21.4)
Course structure and content delivery or pre- and in-class material	54 (18.3)
Learning and knowledge acquisition	53 (18.0)
Real-world scenarios or real-world applicability	44 (14.9)
Exposure to tools and techniques	28 (9.5)
Appreciation for instructors	25 (8.5)
Exposure to complexity and challenges	14 (4.7)
Focus on consumers	14 (4.7)

What Were Participants' Barriers to Engaging With the Program?

When asked regarding barriers to participating in the course, participants' responses formed the following major categories: time constraints due to work, family, and other social commitments: "time constraints, balancing clinical work, other non-clinical work and home life," a lack of knowledge, terminology, and experience: "limited coal-face/frontline exposure and visibility of emerging frontline issues. I work at a more systems-based level and am not involved in interacting with patients day-to-day," technical challenges: "I found using so many new platforms eg Jupyter notebooks, BPMN so quickly challenging...," content complexity, and limited interactions online (Table 10).

XSL•FO
RenderX

Table . Barriers to effective participation.

Theme	Responses (n=259), n (%)
Time constraints or keeping up with materials	109 (42.1)
Lack of knowledge and experience	56 (21.6)
Family and personal commitments	37 (14.3)
Technical challenges	23 (8.9)
Health care terminology and clinical knowledge	14 (5.4)
Work commitments	14 (5.4)
Course structure and content	6 (2.3)

What Changes or Improvements Would Participants Suggest to the Short LHS Coursework?

While the majority of participants found beneficial elements to the course, there are always improvements that can be made. Areas in which changes were suggested were course structure, duration, and timing, suggesting concerns around the pace of the course and the amount of information and breadth covered: "it feels like a lot of materials are being cramped into 1 session and it was hard to appreciate the differences between the models" and the timing of delivery after a long work day; the usability of some learning tools, such as Jupyter Notebooks, difficulties with learning management platform navigation, more revision activities to reinforce learning and a desire for more printable or downloadable resources; questionable benefit of group work and collaborative work where students wanted more support and time to hear instructor expertise: "I feel there was too much reliance on group work and not enough input and guidance from the experts"; course delivery—online format, questioning whether networking opportunities were lost online; prerequisite skills required, given the difficulty of some content (Table 11).

Table . Participants suggested improvements to the course.

Theme	Responses (n=158), n (%)
Course content and structure—curriculum, quality, volume of material, level of complexity, clarity, usefulness, effectiveness, engagement, and applicability	64 (40.5)
Course logistics and administration—course duration, pace, delivery modality, pre-requisites, and learning platforms	37 (23.4)
Learning tools and materials—usability and accessibility	15 (9.5)
Group work and collaboration activities—diversity, effectiveness, and interaction	30 (19.0)
Instructor interactions in-class—interaction, engagement, and support	12 (7.6)

Discussion

Principal Findings

Despite the concept originating in 2007 [24], there is a lack of reports evaluating LHS education programs. In this evaluation, we discuss the findings of 2 years of implementation and iteration of an interdisciplinary Applied LHS professional development course (343/400, 85.8%, presurvey respondents; 200/400, 50%, postsurvey respondents), to a diverse range of professionals working and studying in health care, with an interest in digital health. Most of our participants were from Australia, where LHS was a novel but emerging concept [15,25-27]. The participants found the course engaging and relevant to their work. Participants highlighted specific benefits, barriers, and applications to this course and the LHS framework on their work.

Most health systems are actively seeking to increase the use of data and digital technology to drive improved health care delivery and health outcomes. A major ingredient needed to achieve that lofty goal is a workforce that knows how to not only thrive within the rapidly digitizing world but also how to innovate to improve value-driven care. Training a diverse workforce in the digital transformation of health poses an overwhelming number of choices about the most important learning objectives, competencies, and skills. The LHS framework [16] placed boundaries around the grand vision and enabled us to concretely tell a story that resonated with the goals of potential learners while lending itself to hands-on activities that invite learners to be part of that story.

In addition to the advantages of multidisciplinary curriculum development, the LHS framework was also a key part of the value of the course to interdisciplinary learners. We launched this course as a pilot and hand-selected 50 participants from a much larger pool of applicants with the aim of multidisciplinary involvement and of creating buzz around the course to encourage enrolment for a fee-paying version of the course. Medical directors, research leads, clinicians, and managers brought learnings from the course to hallway discussions and team meetings in their workplaces about how they could apply the LHS framework in specific projects. In addition to a better

XSL•FO
RenderX

understanding of how a project could go from idea to implementation and evaluation using the LHS principles, the framework provided a shared lexicon, a set of approaches like the creation of a learning community, and a toolkit of methods that learners could envision being used in their work. Their excitement was contagious, and a large proportion of our enrolees have come from organizations who continue to sponsor entire interdisciplinary teams of people to take the course together, because they see the value of the framework as a connector across disparate teams, such as clinicians, IT or EMR analysts, and health intelligence units, seeking to work toward a shared goal.

Overall, the course attracted a wide range of professionals at different levels (eg, medical students to directors of emergency departments), professions (eg, nursing and social work), consumers, researchers, and disciplines (eg, IT professionals). In this study, participants highly valued the interdisciplinary nature and collaborative learning activities in the course. Based on previous educational research, we purposefully sorted the groups for a diversity of professions and kept the participants within the same groups for the majority of the course to encourage relationship building. The interdisciplinary aspect of this course was a strength of our education model as it mimics the type of interdisciplinary practice required for complex LHS and digital health initiatives [28].

From several written comments and weekly surveys, we found that different disciplines struggled at different points within the course. For example, people without a research background found the data analysis topic and using Jupyter Notebooks the most challenging aspect of the course, whereas those with a nonclinical background struggled the most with mapping clinical workflows and implementation. Although we used these struggles as teaching moments to demonstrate the need for an interdisciplinary team in LHS, our experience indicates the need to improve our interdisciplinary education model. Previous education researchers and motivational theorists have established that optimal challenge is a key ingredient for engagement and learning [29]. If the material is too easy or too difficult, then learners disengage and, thus, do not learn the material. Many educators have described the challenge of designing a course for optimal challenge among a large cohort of uniprofessional courses [30]. However, our experience is that this challenge is even more dramatic in a one-size-fits-all model in an interdisciplinary course. The content we taught is still appropriate for all audiences, but each person may require more or less self-directed preparatory work as part of the flipped classroom model. Future researchers and educators should investigate how to continue serving an interdisciplinary audience while creating optimal challenges for all participants. For example, in future iterations, we will explore the use of generative artificial intelligence tools to personalize the self-directed online modules for participants' previous knowledge and professional context.

The participants' self-described digital health roles before and after the course only went through minor changes. There was a small conversion in participants who started out seeing themselves as leaders and then later described their roles as connectors. This phenomenon may have been due to instructors telling participants about the importance of connector roles within the LHS framework. Another reason for this effect may be the Dunning-Kruger effect [31]. The Dunning-Kruger effect is when individuals with low exposure to a topic often overestimate their abilities due to a lack of metacognitive awareness. As they gain more knowledge, they become more aware of the limitations. Despite the potential for the Dunning-Kruger effect, the lack of significant changes in participant digital health identity was in contrast to a similar evaluation of our parallel LHS education offering—a 1-year LHS fellowship program for clinicians [15]. In the fellowship program, half of the participants began the program by describing their role as champions and leaders, and then, by the middle of the program, all of the participants described their role as champions and leaders. This potential effect may be due to the benefits of the fellowship program; the fellowship is more experiential, project based, and explicitly focused on leadership development. Since self-identities are an important mediator of future performance [32], future educators and researchers should continue to investigate how LHS educational programs influence participants' self-described roles in the LHS framework and digital health.

Strengths, Limitations, and Future Directions

Overall, we achieved commendable survey response rates, suggesting a high level of engagement from participants. This study uniquely contributes to the existing literature by evaluating an interdisciplinary LHS education program—a domain previously underexplored. Our comprehensive approach encompassed both pre- and postcourse survey data, leveraging learning theories such as self-efficacy theory and the Kirkpatrick evaluation framework to inform our evaluation. Moreover, our qualitative analysis offers valuable insights into participants' perceptions, enriching our understanding of their experiences. However, a limitation is our current inability to capture the upper levels of the Kirkpatrick model, specifically how the LHS course may have influenced participants' workplace behaviors and the subsequent outcomes of those behaviors. In the long term, we aim to evaluate the impact this course and other LHS education offerings have had on individuals and their health organizations' journeys toward a learning health system and individual's career progression. We aim to do this by conducting follow-up, in-depth interviews with participants and organizational sponsors and thematically analyzing the changes that have occurred over time.

Achieving an LHS requires a symbiotic partnership between researchers and health services—by bridging theory and real-world application, future innovations emerging from an LHS will be evidence based and clinically relevant. To increase academic-practice collaboration, our LHS educational offerings aim to grow the understanding of LHS principles and skills in our health services partners and to provide insight into the enablers and barriers for their digital transformation. The shared LHS framework and increased mutual understanding from these programs are increasing trust and collaborative opportunities, leading toward joint translational LHS innovation programs within the health services. We hope that future educators and academic leaders see promise in our emerging LHS education evaluation work [15], other descriptions of LHS education

XSL•FO
RenderX

initiatives [6-11], and the success of LHS initiatives in health care practice [33-35].

By providing a professional development short course, we were able to serve a large market of health professionals who would not otherwise have participated in an expensive university degree. While some professionals like medical specialists receive a continuing medical education fund, most other disciplines are not provided with funding for professional development. Additionally, a major source of participants was partner organizations supporting and sending groups of staff through the program, to learn together as cohorts to develop communities of practice. In this scenario, enrollment was funded by their employers. This is crucial, as at the national and international level, we require a critical mass of appropriately skilled workforce to leverage LHS principles in improving the quality and value of health care delivery.

An interdisciplinary LHS short course has also provided a testbed for applying new technologies to learning. For instance, in the last iteration of the course, we experimented with generative AI feedback on the participants' learning. In their working groups, participants developed an evaluation plan. They fed their plans into ChatGPT, which we provided with structured, custom prompts to provide feedback and rate the quality of the plans. Although some students found the feedback to be generic, the depth of the feedback was dependent upon the richness of the data initially fed to the machine. In large group settings, where there are limited instructors and limited time to provide in-depth feedback to each interdisciplinary group or participant, ChatGPT may be a useful tool to assist with providing formative feedback. The use of this will be further explored in future iterations of the course.

Conclusions

Overall, the Applied Learning Health Systems course received significant positive feedback from interdisciplinary learners. They found the course to be well structured, engaging, and a valuable learning experience. The qualitative comments emphasized the importance of delivering courses that not only provide knowledge but also inspire and motivate learners, and provide concrete tools to apply in their workplaces. A significant number of participants expressed interest in future courses and opportunities for further learning, underscoring the potential for expanding and diversifying course offerings in the future. There is still a great deal of education that needs to be provided to upskill the workforce adequately enough to undertake digital health transformation, but it begins with a shared vision, a common language, and a mutual framework to follow.

Acknowledgments

We are grateful for the dedication of several instructors and contributors to the Applied Learning Health Systems program: Dr Mahima Kalla, Dr Kara Burns, Dr Chris McMaster, Associate Professor Brian Chapman, Dr Douglas Capurro, Reuben Daniels, Jennifer Morris, Dr Vlada Rozova, Anna Fedyukova, Dr Kit Huckvale, Professor Kathleen Gray, Associate Professor Graeme Hart, and Dr Debbie Passey.

Data Availability

All relevant data generated or analyzed during this study are included in this published article. The raw data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

SD played a key role in data curation, formal analysis, investigation, project administration, and visualization, and wrote the original draft, contributing to its review and editing. NZ focused on formal analysis. DC contributed by reviewing and editing the manuscript, ensuring its quality and coherence. WC conceptualized the project and also participated in the review and editing process. KL contributed to the investigation, methodology, supervision, writing the original draft, as well as reviewing and editing the final manuscript.

Conflicts of Interest

None declared.

References

1. Friedman CP. What is unique about learning health systems? Learn Health Syst 2022 Jul;6(3):e10328. [doi: 10.1002/lrh2.10328] [Medline: 35860320]
2. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. Sci Transl Med 2010 Nov 10;2(57):57cm29. [doi: 10.1126/scitranslmed.3001456] [Medline: 21068440]
3. Friedman CP, Allee NJ, Delaney BC, et al. The science of learning health systems: foundations for a new journal. Learn Health Syst 2017 Jan;1(1):e10020. [doi: 10.1002/lrh2.10020] [Medline: 31245555]
4. Detmer DE. Interprofessional clinical informatics education and practice: essentials for learning healthcare systems worldwide. J Interprof Care 2017 Mar;31(2):187-189. [doi: 10.1080/13561820.2016.1250554] [Medline: 28129005]
5. Gray K, Gilbert C, Butler-Henderson K, Day K, Pritchard S. Ghosts in the machine: identifying the digital health information workforce. Stud Health Technol Inform 2019;257:146-151. [Medline: 30741187]

XSL•FO
RenderX

6. McMahon M, Bornstein S, Brown A, Tamblyn R. Training for impact: PhD modernization as a key resource for learning health systems. Healthc Policy 2019 Oct;15(SP):10-15. [doi: 10.12927/hcpol.2019.25983] [Medline: 31755856]
7. Sim SM, Lai J, Aubrecht K, et al. CIHR health system impact fellows: reflections on“driving change” within the health system. Int J Health Policy Manag 2019 Jun 1;8(6):325-328. [doi: 10.15171/ijhpm.2018.124] [Medline: 31256564]
8. Grant RW, Schmittdiel JA, Liu VX, Estacio KR, Chen YFI, Lieu TA. Training the next generation of delivery science researchers: 10-year experience of a post-doctoral research fellowship program within an integrated care system. Learn Health Syst 2024 Jan;8(1):e10361. [doi: 10.1002/lrh2.10361] [Medline: 38249850]
9. Kalra A, Adusumalli S, Sinha SS. Cultivating skills for success in learning health systems: learning to lead. J Am Coll Cardiol 2017 Nov 7;70(19):2450-2454. [doi: 10.1016/j.jacc.2017.09.1086] [Medline: 29096813]
10. Wysham NG, Howie L, Patel K, et al. Development and refinement of a learning health systems training program. EGEMS (Wash DC) 2016;4(1):1236. [doi: 10.13063/2327-9214.1236] [Medline: 28154832]
11. Kohn MS, Topaloglu U, Kirkendall ES, Dharod A, Wells BJ, Gurcan M. Creating learning health systems and the emerging role of biomedical informatics. Learn Health Syst 2022 Jan;6(1):e10259. [doi: 10.1002/lrh2.10259] [Medline: 35036547]
12. Robinson CH, Thompto AJ, Lima EN, Damschroder LJ. Continuous quality improvement at the frontline: one interdisciplinary clinical team’s four-year journey after completing a virtual learning program. Learn Health Syst 2022 Oct;6(4):e10345. [doi: 10.1002/lrh2.10345] [Medline: 36263266]
13. Kasaai B, Thompson E, Glazier RH, McMahon M. Early career outcomes of embedded research fellows: an analysis of the health system impact fellowship program. Int J Health Policy Manag 2023;12:7333. [doi: 10.34172/ijhpm.2023.7333] [Medline: 37579439]
14. Lozano PM, Lane-Fall M, Franklin PD, et al. Training the next generation of learning health system scientists. Learn Health Syst 2022 Oct;6(4):e10342. [doi: 10.1002/lrh2.10342] [Medline: 36263260]
15. Dushyanthen S, Perrier M, Chapman W, Layton M, Lyons K. Fostering the use of learning health systems through a fellowship program for interprofessional clinicians. Learn Health Syst 2022 Oct;6(4):e10340. [doi: 10.1002/lrh2.10340] [Medline: 36263261]
16. Dushyanthen S, Choo D, Perrier M, et al. Designing an interprofessional online course to foster learning health systems. Stud Health Technol Inform 2024 Jan 25;310:1241-1245. [doi: 10.3233/SHTI231163] [Medline: 38270013]
17. Jones-Bonofiglio KD, Willett T, Ng S. An evaluation of flipped e-learning experiences. Med Teach 2018 Sep;40(9):953-961. [doi: 10.1080/0142159X.2017.1417577] [Medline: 29271281]
18. Barr H, Koppel I, Reeves S, Hammick M, Freeth D. Effective interprofessional education. In: Approaching Learning and Teaching: Blackwell Publishing Ltd; 2005:95-104. [doi: 10.1002/9780470776445]
19. Sherer M, Maddux JE, Mercandante B, Prentice-Dunn S, Jacobs B, Rogers RW. The self-efficacy scale: construction and validation. Psychol Rep 1982 Oct;51(2):663-671. [doi: 10.2466/pr0.1982.51.2.663]
20. Pajares F. Self-efficacy beliefs in academic settings. Rev Educ Res 1996 Dec;66(4):543-578. [doi: 10.2307/1170653]
21. Greenberg-Worisek AJ, Shippee ND, Schaffhausen C, et al. The learning health system competency appraisal inventory (LHS-CAI): a novel tool for assessing LHS-focused education needs. Learn Health Syst 2020 Apr;5(2):10.1002/lrh2.10218. [doi: 10.1002/lrh2.10218] [Medline: 33889729]
22. Forrest CB, Chesley FD, Tregear ML, Mistry KB. Development of the learning health system researcher core competencies. Health Serv Res 2018 Aug;53(4):2615-2632. [doi: 10.1111/1475-6773.12751] [Medline: 28777456]
23. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
24. Institute of Medicine (US) Roundtable on Evidence-Based Medicine; Olsen LA AD, Olsen LA, Aisner D, McGinnis JM. The Learning Healthcare System: Workshop Summary: National Academies Press (US); 2007. [doi: 10.17226/11903]
25. Enticott JC, Melder A, Johnson A, et al. A learning health system framework to operationalize health data to improve quality care: an Australian perspective. Front Med (Lausanne) 2021;8:730021. [doi: 10.3389/fmed.2021.730021] [Medline: 34778291]
26. Ellis LA, Sarkies M, Churruca K, et al. The science of learning health systems: scoping review of empirical research. JMIR Med Inform 2022 Feb 23;10(2):e34907. [doi: 10.2196/34907] [Medline: 35195529]
27. Dammery G, Ellis LA, Churruca K, et al. The journey to a learning health system in primary care: a qualitative case study utilising an embedded research approach. BMC Prim Care 2023 Jan 19;24(1):22. [doi: 10.1186/s12875-022-01955-w] [Medline: 36653772]
28. Friedman C, Rubin J, Brown J, et al. Toward a science of learning systems: a research agenda for the high-functioning learning health system. J Am Med Inform Assoc 2015 Jan;22(1):43-50. [doi: 10.1136/amiajnl-2014-002977] [Medline: 25342177]
29. Nakamura J, Csikszentmihalyi M. The concept of flow. In: Handbook of Positive Psychology 2002:89-105. [doi: 10.1093/oso/9780195135336.001.0001]
30. Cooper JL, Robinson P. The argument for making large classes seem small. New Drctns for Teach & Learn 2000 Mar;2000(81):5-16. [doi: 10.1002/tl.8101]
31. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one’s own incompetence lead to inflated self-assessments. J Pers Soc Psychol 1999 Dec;77(6):1121-1134. [doi: 10.1037//0022-3514.77.6.1121] [Medline: 10626367]

XSL•FO
RenderX

32. Nolen SB, Horn IS, Ward CJ. Situating motivation. Educ Psychol 2015 Jul 3;50(3):234-247. [doi: 10.1080/00461520.2015.1075399]
33. Enticott J, Johnson A, Teede H. Learning health systems using data to drive healthcare improvement and impact: a systematic review. BMC Health Serv Res 2021 Mar 5;21(1):200. [doi: 10.1186/s12913-021-06215-8] [Medline: 33663508]
34. Casey JD, Courtright KR, Rice TW, Semler MW. What can a learning healthcare system teach us about improving outcomes? Curr Opin Crit Care 2021 Oct 1;27(5):527-536. [doi: 10.1097/MCC.0000000000000857] [Medline: 34232148]
35. Budrionis A, Bellika JG. The learning healthcare system: where are we now? A systematic review. J Biomed Inform 2016 Dec;64:87-92. [doi: 10.1016/j.jbi.2016.09.018] [Medline: 27693565]

Abbreviations

EMR: electronic medical record
IT: information technology
LHS: learning health system

Edited by TDA Cardoso; submitted 31.10.23; peer-reviewed by A Lakdawala, B Senst; revised version received 18.03.24; accepted 26.04.24; published 14.01.25.

Please cite as:
Dushyanthen S, Zamri NI, Chapman W, Capurro D, Lyons K
Evaluation of an Interdisciplinary Educational Program to Foster Learning Health Systems: Education Evaluation
JMIR Med Educ 2025;11:e54152
URL: https://mededu.jmir.org/2025/1/e54152
doi:10.2196/54152

© Sathana Dushyanthen, Nadia Izzati Zamri, Wendy Chapman, Daniel Capurro, Kayley Lyons. Originally published in JMIR Medical Education (https://mededu.jmir.org), 14.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Recruiting Medical, Dental, and Biomedical Students as First Responders in the Immediate Aftermath of the COVID-19 Pandemic: Prospective Follow-Up Study

Nicolas Schnetzler[1,2], MSc; Victor Taramarcaz[1,2], MSc; Tara Herren[1,2], MSc; Eric Golay[2], MAS; Simon Regard[2,3], MD; François Mach[4], MD; Amanta Nasution[1,2], BSc; Robert Larribau[1,2], MD; Melanie Suppan[1,5], MD, MSc; Eduardo Schiffer[1,5,6], MD; Laurent Suppan[1,2], MD

[1]Department of Anaesthesiology, Pharmacology, Intensive Care and Emergency Medicine, Faculty of Medicine, University of Geneva, Geneva, Switzerland

[2]Division of Emergency Medicine, Department of Acute Care Medicine, Geneva University Hospitals, Rue Gabrielle-Perret-Gentil 2, Geneva, Switzerland

[3]Cantonal Physician Division, Cantonal Health Office, State of Geneva, Geneva, Switzerland

[4]Cardiology Department, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

[5]Division of Anaesthesiology, Department of Acute Care Medicine, Geneva University Hospitals, Geneva, Switzerland

[6]Unit of Development and Research in Medical Education (UDREM), Faculty of Medicine, Geneva, Switzerland

Corresponding Author:
Laurent Suppan, MD
Department of Anaesthesiology, Pharmacology, Intensive Care and Emergency Medicine, Faculty of Medicine, University of Geneva, Geneva, Switzerland

Abstract

Background: Basic life support improves survival prognosis after out-of-hospital cardiac arrest, but is too rarely provided before the arrival of professional rescue services. First responder networks have been developed in many regions of the world to decrease the delay between collapse and initiation of resuscitation maneuvers. Their efficiency depends on the number of first responders available and many networks lack potential rescuers. Medical, dental, and biomedical students represent an almost untapped source of potential first responders, and a first study, carried out during the COVID-19 pandemic, led to the recruitment of many of these future professionals even though many restrictions were still in effect.

Objective: The objective of this study was to determine the impact of an enhanced strategy on the recruitment of medical, dental, and biomedical students as first responders in the immediate aftermath of the COVID-19 pandemic.

Methods: This was a prospective follow-up study, conducted between November 2021 and March 2022 at the University of Geneva Faculty of Medicine, Geneva, Switzerland. A web-based study platform was used to manage consent, registrations, and certificates. A first motivational intervention was held early in the academic year and targeted all first-year medical, dental, and biomedical students. Participants first answered a questionnaire designed to assess their initial basic life support knowledge before following an e-learning module. Those who completed the module were able to register for a face-to-face training session held by senior medical students. A course certificate was awarded to those who completed these sessions, enabling them to register as first responders on the Save a Life first responder network. Since the number of students who had enlisted as first responders 2 months after the motivational intervention was markedly lower than expected, a second, unplanned motivational intervention was held in an attempt to recruit more students.

Results: Out of a total of 674 first-year students, 19 (2.5%) students had registered as first responders after the first motivational intervention. This was significantly less than the proportion achieved through the initial study (48/529, 9.1%; $P<.001$). The second motivational intervention led to the enrollment of 7 more students (26/674, 3.9%), a figure still significantly lower than that of the original study ($P<.001$). At the end of the study, 76 (11.3%) students had been awarded a certificate of competence.

Conclusions: Contrary to expectations, an earlier presentation during the academic year outside the COVID restriction period did not increase the recruitment of medical, dental, and biomedical students as first responders in the immediate aftermath of the COVID-19 pandemic. The reasons underlying this drop in motivation should be explored to enable the design of focused motivational interventions.

(JMIR Med Educ 2025;11:e63018) doi:10.2196/63018

XSL•FO
RenderX

KEYWORDS

basic life support; out-of-hospital cardiac arrest; cardiopulmonary resuscitation; e-learning; blended learning; first responder; undergraduate medical education; student motivation; motivational strategies; medical student; COVID-19; pandemic; life support; survival prognosis; biomedical students; dental students; motivational interventions

Introduction

Background

Basic life support (BLS) improves survival prognosis after out-of-hospital cardiac arrest (OHCA) but is too rarely provided before the arrival of professional rescue services [1-6]. Without BLS, the probability of survival decreases by 10% for each minute that passes [7]. Thus, professional rescue is of limited worth if BLS has not been provided either by bystanders or by first responders [8-10]. Indeed, several studies have demonstrated that initiation of BLS maneuvers by nonprofessionals improves survival and neurological outcomes [1,11,12].

Increasing global awareness regarding the importance of quickly initiating BLS maneuvers after OHCA will take time, and barriers to action often prevent bystanders from initiating cardiopulmonary resuscitation [1,13-15]. To overcome this limitation, first responder systems have been developed in many regions of the world. These systems rely on BLS-certified professional or nonprofessional rescuers who accept a call to respond to OHCA alarms if they happen to be nearby.

In Geneva, Switzerland, the Save a Life project was initiated in October 2019 by the Swiss Emergency Responder Association, with the objective of developing a regional network of first responders [7]. When an OHCA is identified by the emergency medical call center, an alert is displayed on the Save a Life first responder app. If a first responder is near enough and agrees to intervene, the position of the nearest automatic external defibrillator (AED) is displayed along with the exact location of the intervention. The main limitations of this system are the limited number of first responders, their availability, and their geographical distribution.

To improve the number of first responders, Taramarcaz et al [16] designed a process to recruit first-year medical students. Their study took place while the COVID-19 pandemic was still ongoing, and many restrictions were still in effect. In addition, the motivational intervention designed to catch the students' interest was held online rather than in an auditorium and took place rather late after the beginning of the academic year and close to a critical exam session. Thus, the authors hypothesized that an intervention taking place earlier in the academic year, and without the constraints imposed by the COVID-19 pandemic, could lead to higher participation rates and the recruitment of a higher proportion of medical students as first responders [16].

Objective

The objective of this study was to determine the impact of the modifications proposed by Taramarcaz et al [16] on first responder recruitment.

Methods

Study Design

This prospective follow-up study was conducted between November 2021 and March 2022 and followed a structure and sequence similar to that described by Taramarcaz et al [16].

The study platform used for the initial study was reset and reused for this follow-up study. Given the use of a web-based platform, methods and results are reported according to the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) guidelines when appropriate [17]. Data were managed in accordance with the European General Data Protection Regulation [18]. A more detailed description of the tools used can be found in Multimedia Appendix 1.

The learning path was identical to that described in Taramarcaz et al's [16] study and followed a flipped classroom design: after ensuring that no exclusion criteria were present, the first-year medical, dental, and biomedical students of the University of Geneva Faculty of Medicine (UGFM) answered a questionnaire designed to assess their initial BLS knowledge before following an e-learning module. After completing this module, they were able to register for a face-to-face training session held by senior medical students. The estimated time required to complete the e-learning and practice session was about an hour and a half. This duration was chosen as being long enough for learning and skill retention while avoiding an overt demand on their busy schedule. The participants who completed the entire learning path were awarded a BLS-AED course certificate enabling them to register as first responders on the Save a Life first responder network. The whole process, including the certification, was entirely free of charge, and there was no obligation for students to participate. The only incentive was to obtain a BLS-AED certificate.

Ethical Considerations

Since the regional ethics committee (Commission cantonale d'éthique de la recherche, Geneva, Switzerland) had already acknowledged that this design did not fall within the scope of the Swiss federal law on research involving human beings (Req-2020 - 01143), no further ethical assessment was required or requested.

Recruitment

A motivational intervention was performed live on November 29, 2021. This intervention was animated by 2 senior medical students and took place at the end of a basic medical science course. The presentation contained a short, humorous introductory video, a description of the project, and an overview of the Save a Life network. The last slide included a QR code and a URL link to the study platform as indicated in the research protocol [19]. On the same day, all potential participants received the same information by email through the class's

XSL•FO
RenderX

mailing list. To promote participation, a second motivational email was sent to the entire class on December 10.

Since participation was markedly lower than expected by the end of December, a second intervention was planned, this time at the start of a course on atherosclerosis given by the head of the cardiology department at Geneva University Hospital. This intervention took place on January 10, 2022, and a final reminder email was sent on January 13, 2022. For the second intervention, different support material was used, and various real-life scenarios were included, showcasing how BLS knowledge could enable them to act in the case of OHCA.

Enrollment

The QR code and URL provided during the motivational interventions and through the invitation emails redirected the students to an introductory page detailing the study's objectives and procedures. Those willing to participate were asked to answer 2 questions designed to detect the presence of either of 2 exclusion criteria: being registered as first responder and not being a UGFM student. If neither exclusion criterion was met, the students were redirected to a consent form (Table 1) including a disclaimer about data handling and security. Those who agreed were asked to create an account and to provide minimal personal information (first name, last name, and email address) for contact purposes and to allow for the creation of nominative BLS-AED certificates. The students who refused to participate and those who met either exclusion criteria were also given the possibility to follow the learning path and to receive a certificate allowing them to join a first responder system.

After completing the registration process, participants were asked to fill out a precourse questionnaire designed to gather demographic data and determine their precourse BLS knowledge (Table 2).

Table . Screening questionnaire and consent form (reused from Taramarcaz et al [16]).

Survey page, field, and question	Type of question
Page 1	
Already filled the questionnaire or exclusion criteria	
Already a first responder	Yes or no
Demographics	
Student at UGFM[a]	Yes or no
If no: current professional status	Open
Consent	
Agree to participate	Yes or no
If no: reasons for refusal	MAQ[b]
If no: access to the e-learning module	Yes or no

[a]UGFM: University of Geneva Faculty of Medicine.

[b]MAQ: multiple answer question.

XSL•FO
RenderX

Table . Precourse questionnaire (reused from Taramarcaz et al [16]).

Survey page, field, and question	Type of question
1: Demographics	
Year of birth	Open (Regex[a])
Gender	MCQ[b]
Medical, biomedical or dental medicine student	MCQ
Former student or graduate of another health care profession	MCQ
Target Specialty	MCQ
2: General BLS[c] knowledge	
Ever heard of BLS or ACLS[d] before	Yes/no
Meaning of AED[e,f]	Open
Year of the last BLS guidelines update	Open (Regex)
Phone number of the emergency medical communication center[f]	Open
3: Prior BLS experience	
Prior BLS training	MAQ[g]
Wish for additional BLS training	Yes/no
4: Specific BLS knowledge	
Criteria used to recognize OHCA[f,h]	MAQ
BLS-sequence[f]	Ordering
Artery for pulse assessment[f]	MCQ
Compression depth[f]	MCQ
Compressions: ventilation ratio[f]	MCQ
Compression rate[f]	MCQ
Compression-only CPR[fi]	Yes/no
Foreign body airway obstruction [f]	MCQ
5: Confidence	
Precourse confidence to act in an OHCA situation	Likert scale (1-5)

[a]A Regex validation rule was used to avoid invalid entries.
[b]MCQ: multiple choice question (only one answer accepted).
[c]BLS: basic life support.
[d]ACLS: advanced cardiovascular life support.
[e]AED: automatic external defibrillator.
[f]Items used to calculate the 10-point score (initial BLS knowledge).
[g]MAQ: multiple answer question (more than one answer accepted).
[h]OHCA: out-of-hospital cardiac arrest.
[i]CPR: cardiopulmonary resuscitation.

E-Learning and Practice Sessions

The interactive e-learning module used in Taramarcaz et al's [16] study was reused without any changes since it still matched the objectives, respected the Swiss Resuscitation Council's guidelines, and had not received any negative feedback from the students. This module was designed to last about 30 minutes, but no time limit was set and students were able to resume at will. A screen enabling participants to register for near-peer animated practice sessions was displayed upon completion of this e-learning module.

Practice sessions lasted 1 hour and were limited to 4 participants. A total of 32 sessions (128 slots) were planned between December 6, 2021 and March 11, 2022. The instructor-to-participant ratio (1:4) was kept unchanged to maintain high-quality training even though the COVID-19 restrictions had been lifted. The senior medical students who

XSL•FO
RenderX

animated these near-peer-led practice sessions were all certified as BLS-AED instructors according to the Swiss Resuscitation Council's guidelines. Most of the students who had already participated as instructors in Taramarcaz et al's [16] study (15/17, 88%) agreed to resume their involvement and 5 new instructors were trained. While all instructors were to ensure that the objectives had been met by using a standardized checklist, they were free to adapt the structure of their training sessions according to the participants' profiles.

Final Questionnaire and Certification

An email embedding a link to a postcourse questionnaire was sent to the students who successfully completed the practice sessions (Table 3). Participation in this questionnaire was mandatory to obtain a nominative BLS-AED certificate. These certificates, which had a 1-year validity, enabled participants to enroll as first responders on the Save a Life platform.

Table . Postcourse questionnaire (reused from Taramarcaz et al [16]).

Survey page, field, and question	Type of question
1: Opinion	
Appreciation	Yes/no
If yes: positive thoughts	MAQ[a]
If no: negative thoughts	MAQ
General comments	Free text
2: Confidence	
Postcourse confidence for OHCA[b] management	Likert scale (1-5)
Factors contributing to confidence	Likert scale (1-5)
Factors contributing to lack of confidence	Likert scale (1-5)
Other comments on confidence	Free text
3: First responders	
Intention to register as first responder	Yes/no
If yes: contributing factors	Likert scale (1-5)
If no: impeding factors	Likert scale (1-5)
Other factors	Free text
4: Improvement	
Suggestion for improvement	Free text

[a]MAQ: multiple answer question.
[b]OHCA: out-of-hospital cardiac arrest.

Adaptations From the Implementation Study

In line with this study's objectives, the main changes from the implementation study were that the initial presentation to first-year students and the training sessions were held earlier in the academic year [16], with the hypothesis that this would increase the number of registrations as first-year students would be further away from their final exams. Thus, the project was presented on November 29, 2 months earlier than the original study.

Despite this adaptation, and contrarily to our hypothesis, the number of participants was markedly lower than that in the original study. A second, initially unplanned intervention was therefore carried out in early January 2022, and constitutes the second major adaptation from the original implementation study.

Another difference was that biomedical students were also invited to participate in this study. Finally, the practice sessions were held between December 2021 and March 2022 in this study while they had taken place between January and April 2021 in the implementation study. The number of slots remained unchanged.

Outcomes

The primary outcome was the proportion of students who had registered as first responders before the second intervention took place, that is, by January 9, 2022. Secondary outcomes were the proportion of students who had registered following the second intervention, the overall proportion of students who had registered as first responders by May 1, 2022, and attrition at each step of the study [20]. The difference in self-reported confidence in performing BLS maneuvers was also assessed.

Statistical Analysis

Data curation and analysis were carried out using STATA/BE (version 17.0; StataCorp LLC). Descriptive statistics were used to describe the evolution of the number of students at each step of the learning path. Given the sample size, parametric tests were used when appropriate. A P value of less than .05 was considered statistically significant.

XSL•FO
RenderX

The chi-square test was used to assess the difference in student recruitment distribution between this study and Taramarcaz et al's [16]. This was carried out by reusing the original data file, which is freely available online as a Multimedia Appendix 1 of the original study. Since biomedical students had not been invited to participate in the original study, a sensitivity analysis was carried out by excluding them.

Potential differences between students who registered after following the first motivational intervention and those who registered after following the second one were looked for by applying a *t* test on the 10-point BLS score and by comparing attrition at each step. No weighting was used to compute the 10-point BLS knowledge score. A *t* test was performed to look for a difference between this score and interest in following BLS training.

A *t* test was also used to investigate whether there was a statistically significant difference between postcourse confidence and enrollment in the first responder network.

Given the presence of cells with very limited numbers (<5), Fischer tests were applied to analyze the factors influencing self-confidence and the desire to join the Save a Life first responder network.

Results

The 2021 - 2022 academic year included a total of 674 first-year students at UGFM in human and dental medicine and in biomedical sciences. The proportion of students who had registered as first responders after the first motivational intervention was 2.5% (19/674), significantly less than after Taramarcaz et al's [16] implementation study (48/529, 9.1%; $P<.001$). The second motivational intervention led to the enrollment of 7 more students (26/674, 3.9%). This figure is still significantly lower than that observed in Taramarcaz et al's study ($P<.001$) [16]. Even after excluding biomedical students from the analysis, the figure remained significantly lower (25/600, 4.2%) than in Taramarcaz et al's [16] study ($P=.001$).

A total of 502 (74.5%) students followed the link directing them to the study platform, of whom 447 (66.3%) students completed the screening questionnaire. Only 133 (19.7%) students registered on the platform and 76 (11.3%) students received a BLS-AED certificate at the end of the learning program. In the postcourse questionnaire, 68.4% (52/76) of students who obtained the certificate indicated a desire to join the network of first responders, but only 34.2% (26/76) of students followed through. Figure 1 shows participation at each step of the study.

There was a statistically significant relationship between prior BLS knowledge and e-learning completion ($P=.007$), practical session attendance ($P<.001$), and obtention of a BLS certificate ($P=.003$). Conversely, there was no statistically significant relationship between prior BLS-AED knowledge and enrollment in the Save a Life network ($P=.05$) or interest in the program ($P=.94$).

Students' confidence in their ability to initiate BLS maneuvers was significantly increased after following the learning path ($P<.001$, Multimedia Appendix 2). There was no statistical link between postcourse confidence and registration on the Save a Life platform ($P=.09$).

Postcourse satisfaction was 100% (76/76), as was the probability that students who had completed the learning path would recommend it to other students.

Figure 2 shows that a better understanding of health issues, a feeling of mastery of the subject, and an improvement in knowledge regarding resuscitation all contributed to promoting participant confidence.

Stress and fear of doing wrong were the 2 main factors reported as limiting one's confidence in performing BLS maneuvers (Figure 3).

Four factors promoting student willingness to register on the Save a Life platform were identified: feeling able to perform cardiopulmonary resuscitation, the possibility of making a difference, the stakes, and the desire to help (Figure 4).

Stress was the main factor preventing participants from registering as first responders (Figure 5).

XSL•FO
RenderX

Figure 1. Study flowchart. AED: automatic external defibrillator; BLS: basic life support.

XSL•FO
RenderX

Figure 2. Factors promoting student confidence in their ability to perform basic life support (BLS) maneuvers.

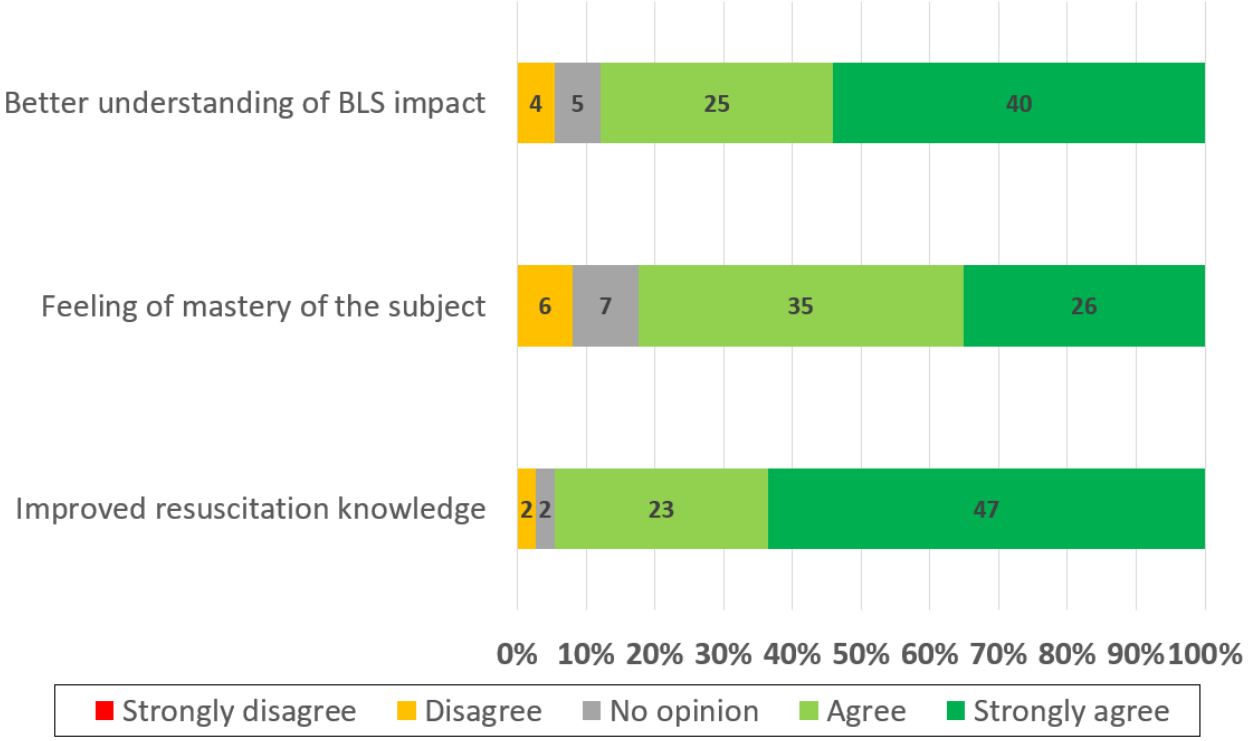
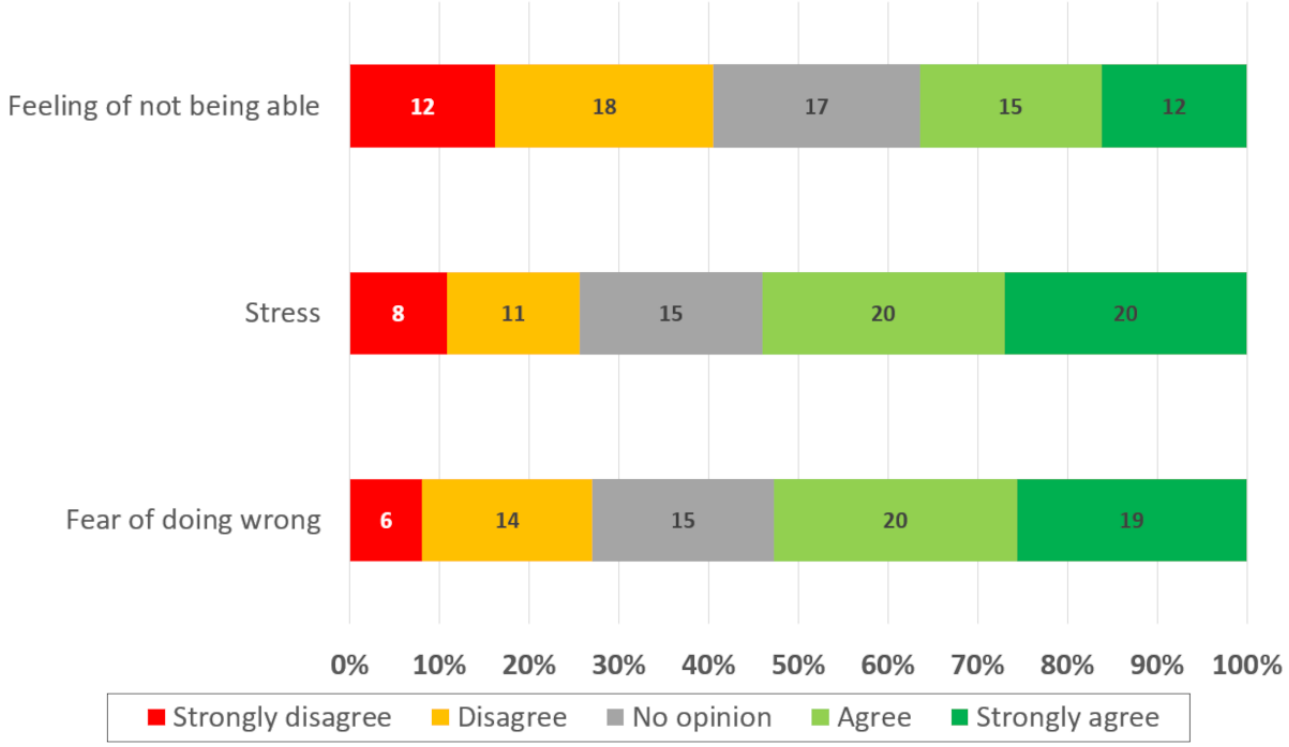


Figure 3. Factors limiting student confidence in performing basic life support maneuvers.



XSL•FO
RenderX

Figure 4. Factors promoting student willingness to register on the Save a Life platform. CPR: cardiopulmonary resuscitation.

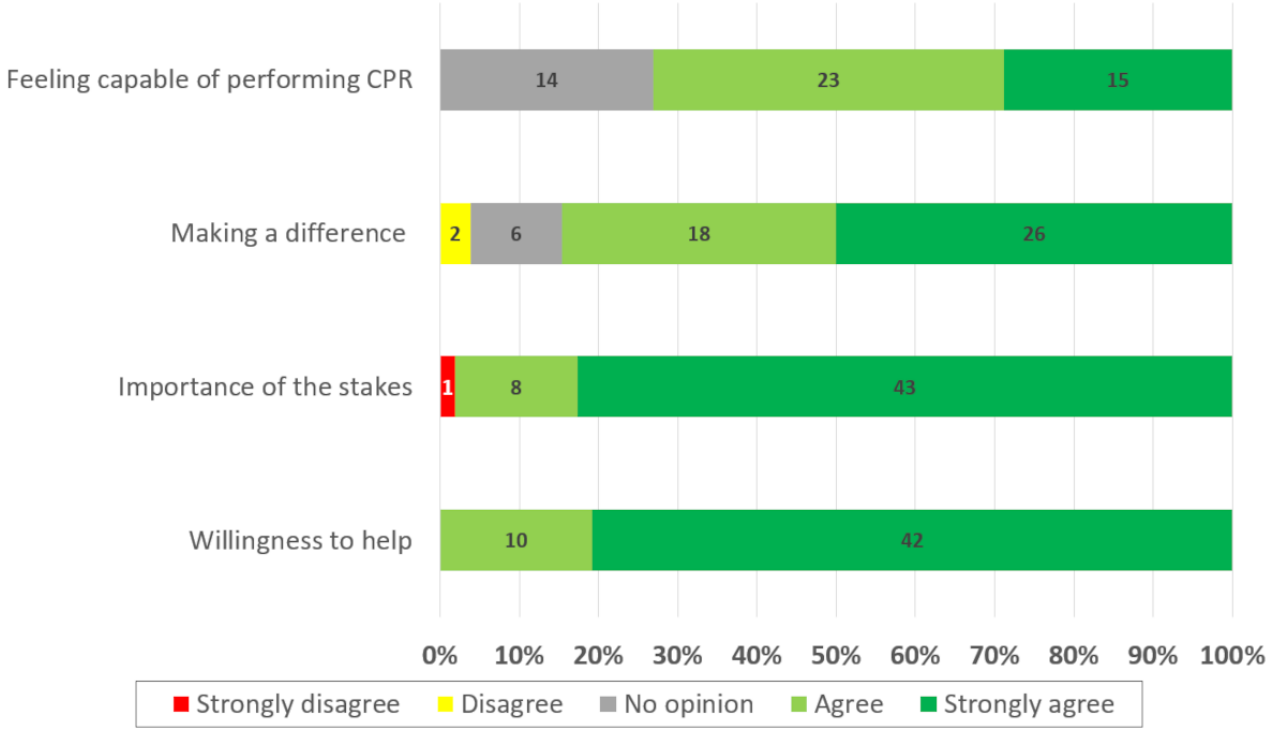
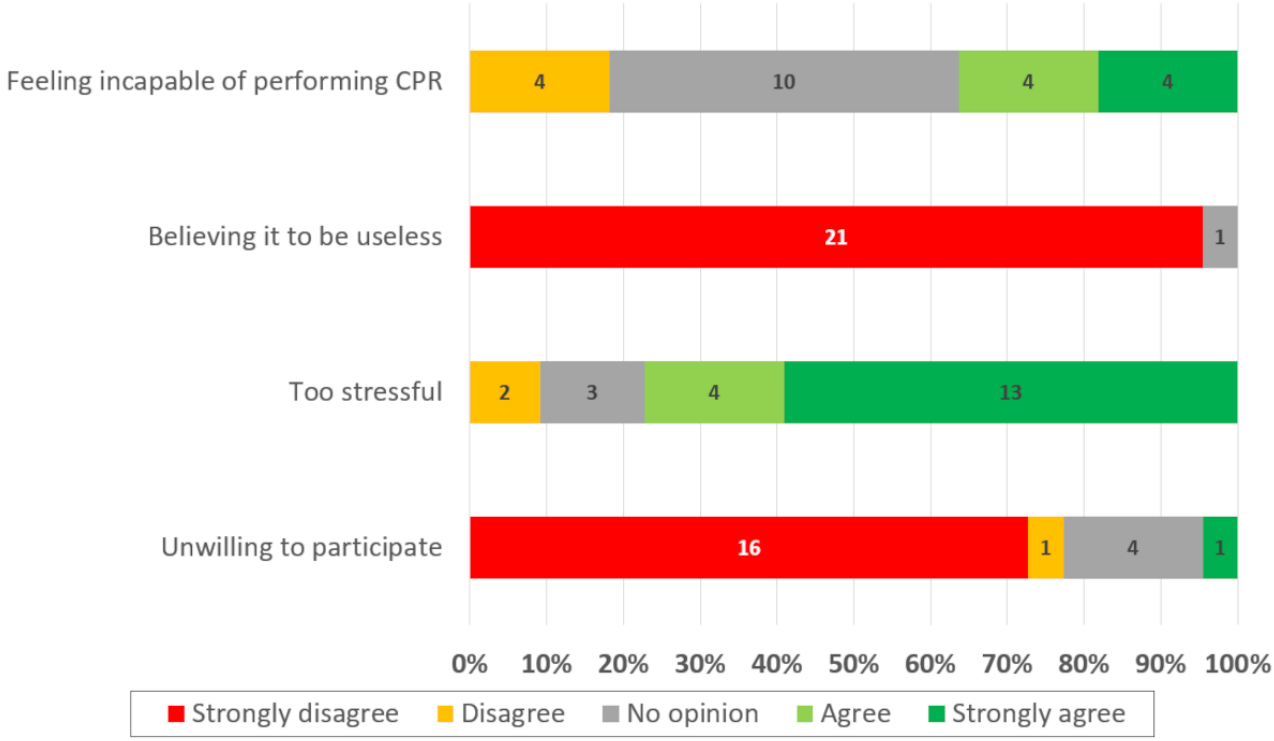


Figure 5. Factors limiting student willingness to register on the Save a Life platform. CPR: cardiopulmonary resuscitation.



Discussion

Main Considerations

Despite the modifications carried out according to the hypotheses outlined in the original study [16], and even after a second motivational intervention, recruitment was markedly lower than expected: indeed, in the original implementation study [16], the proportion of people who had registered on the platform at the end of the project was more than 2 times higher. The target set in the initial protocol was to recruit 10% of first-year students, a goal that has not yet been achieved.

These results deserve to be analyzed from a motivational point of view. The original study took place during the COVID-19 pandemic period, when student dynamics were probably

XSL•FO
RenderX

different, and students may have been more inclined to participate in a presential activity, due to the fact that they had no choice but to spend their first year remotely. The COVID period was a major awareness and health involvement on the part of medical students. Their involvement in medical tasks having a perceptible impact on the future of patients affected by the pandemic undoubtedly positively influenced their motivational reinforcement. This paradox probably partially explains the low recruitment observed in this study, carried out outside the pandemic context. Other factors may also have influenced student motivation during the COVID pandemic: during this period, the health care system was highly regarded by the population, and the project may have given the students a sense of belonging [21]. The feeling of being useless in the face of what was happening and the desire to help may also have been stronger at this time [21]. Conversely, the end of the restrictions may have decreased their motivation to take part in such a project, and students may have been keen to resume many of the activities they had been deprived of [22].

The profile of the teacher endorsing the motivational intervention may also have played a role since teachers can have a significant influence on their students [23]. Since most medical students are interested in the clinical field, any advice, opinion, or encouragement given by a clinician could have a particularly important influence on students [24]. Clinicians can also share their interest and experience in a subject [25], and their support can foster student interest in a particular field [24]. Indeed, human beings strive to feel connected to those they admire, and the sense of belonging that a prestigious clinician radiates can influence student motivation [24]. In addition, first-year students are more motivated by success, prestige, and money, compared with the upper years, who are more focused on the personal gratification of their activity [24,26]. Moreover, student motivation fluctuates over the years, both qualitatively and quantitatively. Understanding its evolution can help encourage students to enjoy their learning and possibly improve their performance [27].

According to the theory of self-determination, there are several types of motivation, depending on what influences it and what goals it aims to achieve: intrinsic motivation, extrinsic motivation, and amotivation [26]. Intrinsic motivation is linked to personal interest in or pleasure inherent to the activity. Extrinsic motivation aims at a goal, a consequence separable from the subject, such as a reward or the absence of inconvenience [26]. Extrinsic motivation can be described as a continuum through which a process of internalization takes place, finally resulting in integrating action towards self-determination [26]. In the educational environment, motivation can be seen as having 3 determinants: the perception of the value of an activity, its skill, and its controllability [28].

A clinician's valorization of the abilities and importance that each student can have in the health care system at their own level can influence the perception of their abilities, and their involvement and motivation [23]. A clinician's speech on public health issues can have a greater impact and radiate a positive perception of the values involved [24,28].

Despite these different aspects motivating first-year students to participate in an optional learning program can still be difficult since it does not bring them any short-term benefits, in this case passing their exams. According to Dweck, students pursue learning goals as well as performance goals [29]. In the short term, when the risk of success is low, students would restrict themselves to the performance goal and neglect activities they consider ineffective for success [28,30].

A considerable proportion of students did not continue with the learning path after completing the questionnaire assessing their BLS knowledge. These students' scores were lower than average, and their lack of knowledge may have had an impact on their perception of their skills for future activities, decreasing their self-confidence and motivation to continue their learning program [28]. This could be addressed by introducing BLS courses at school since, in Geneva, most schoolchildren receive only little, if any, first aid training before attending courses mandatory to obtain a driving license. Furthermore, training schoolchildren has been shown to improve OHCA outcomes [31]. Another option could be to remove this questionnaire from future studies to avoid any attrition linked to its administration.

Once enrolled in the learning program, students follow a certifying course, but registration on the Save a Life platform remains optional, and students therefore need further motivation to enlist as first responders. Participants agreed that perfecting their knowledge, mastering the subject, and understanding the health issues linked to early resuscitation all improved their self-confidence. This is in line with Viau and Louis's [28] opinion, that is, that the perception of the value of an activity and of one's own skills influence motivation.

Understanding the social impact of a first responder network could enable potential participants to internalize the values involved, and, according to the self-determination theory, increase motivation [26]. Thus, the societal impact of the Save a Life project could be further highlighted in future motivational interventions. This could improve recruitment since respondents unanimously agreed that the desire to help influenced their probability of registering as first responders.

Students reported that the main factors limiting their willingness to register as first responders were stress and the fear of making a mistake. Stress goes against the feeling of controllability of a situation, which is essential to self-confidence [28]. Addressing this issue will require further exploration, but a first step could be to point out the low risk of harm to the patient when practicing BLS maneuvers [14,15] and the clear benefits of early resuscitation early in the presentation [8-10].

Aspects modulating self-confidence need to be highlighted in future presentations to students, but also during the training program, to encourage these students as much as possible to join the Save a Life network. Their abilities and knowledge should be encouraged, and the efforts and gains they can make in the management of OHCA should be recognized. Any fears or doubts they may have must also be addressed during the learning path, and the effect of these motivational enhancements will need to be assessed in the next few years.

XSL•FO
RenderX

Limitations

Since the very low participation rate could not be anticipated, the design of this study had to be adapted. Even though the second motivational intervention was endorsed by a clinician while the first was endorsed by a specialist in basic medical science, the effect of each specific intervention could not be assessed given the design of this study. A randomized controlled trial could be considered to explore the effects of endorsement by either type of specialist. In addition, the motivational interventions themselves were also different, and the effect of specifically designed and theory-based motivational interventions would also deserve to be determined. Finally, the impact of specific factors on motivation was only assessed among the students who had followed the learning path, thereby leading to a selection bias. Therefore, participatory research should be considered to help identify better recruitment strategies, and focus groups held to gather a more thorough and less biased understanding of students' motivation and barriers to participation.

Conclusions

Contrary to expectations, an earlier presentation during the academic year outside the COVID restriction period did not increase the recruitment of medical students as first responders, which was more than 2 times lower than in the implementation study even after further motivational interventions. A thorough quantitative and qualitative exploration of motivational factors should be carried out to determine potential ways of improving the recruitment of first-year medical students as first responders.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Web-based platform.
[DOCX File, 15 KB - mededu_v11i1e63018_app1.docx]

Multimedia Appendix 2
Evolution of self-confidence in practice basic life support maneuver.
[PNG File, 25 KB - mededu_v11i1e63018_app2.png]

References

1. Panchal AR, Bartos JA, Cabañas JG, et al. Part 3: adult basic and advanced life support: 2020 American heart association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2020 Oct 20;142(16_suppl_2):S366-S468. [doi: 10.1161/CIR.0000000000000916] [Medline: 33081529]
2. Semeraro F, Greif R, Böttiger BW, et al. European resuscitation council guidelines 2021: systems saving lives. Resuscitation 2021 Apr;161:80-97. [doi: 10.1016/j.resuscitation.2021.02.008] [Medline: 33773834]
3. Olasveengen TM, Mancini ME, Perkins GD, et al. Adult basic life support: 2020 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations. Circulation 2020 Oct 20;142(16_suppl_1):S41-S91. [doi: 10.1161/CIR.0000000000000892] [Medline: 33084391]
4. Weisfeldt ML, Sitlani CM, Ornato JP, et al. Survival after application of automatic external defibrillators before arrival of the emergency medical system. J Am Coll Cardiol 2010 Apr;55(16):1713-1720. [doi: 10.1016/j.jacc.2009.11.077]
5. Arrêts cardio-respiratoires préhospitaliers: revue de la littérature et évolution du pronostic à genève entre 2009 et 2012 [Pre-hospital cardiopulmonary arrest: literature review and prognosis trends in Geneva between 2009 and 2012]. : Université de Genève; 2018 URL: https://access.archive-ouverte.unige.ch/access/metadata/9b59894e-ebb0-46cb-9c99-69897632b06b/download [accessed 2025-05-05]
6. Perkins GD, Graesner JT, Semeraro F, et al. European resuscitation council guidelines 2021: executive summary. Resuscitation 2021 Apr;161:1-60. [doi: 10.1016/j.resuscitation.2021.02.003] [Medline: 33773824]
7. Rapport d'activité [annual report]. : Save-a-Life; 2021 URL: https://www.save-a-life.ch/wp-content/uploads/2022/04/Save-a-Life-Rapport-dactivites-2021.pdf [accessed 2023-03-21]
8. Stiell IG, Wells GA, DeMaio VJ, et al. Modifiable factors associated with improved cardiac arrest survival in a multicenter basic life support/defibrillation system: OPALS study phase I results. Ann Emerg Med 1999 Jan;33(1):44-50. [doi: 10.1016/s0196-0644(99)70415-4] [Medline: 9867885]
9. Fordyce CB, Hansen CM, Kragholm K, et al. Association of public health initiatives with outcomes for out-of-hospital cardiac arrest at home and in public locations. JAMA Cardiol 2017 Nov 1;2(11):1226-1235. [doi: 10.1001/jamacardio.2017.3471] [Medline: 28979980]
10. Dami F, Fuchs V, Berthoz V, Carron PN. Régulation médicale: mise au point 2018 et développements futurs [medical dispatch: 2018 update and future developments]. Ann Fr Med Urgence 2018 Nov;8(6):376-382. [doi: 10.3166/afmu-2018-0089]

XSL•FO
RenderX

11. Iwami T, Kitamura T, Kiyohara K, Kawamura T. Dissemination of chest compression-only cardiopulmonary resuscitation and survival after out-of-hospital cardiac arrest. Circulation 2015 Aug 4;132(5):415-422. [doi: 10.1161/CIRCULATIONAHA.114.014905] [Medline: 26048093]
12. Yasunaga H, Horiguchi H, Tanabe S, et al. Collaborative effects of bystander-initiated cardiopulmonary resuscitation and prehospital advanced cardiac life support by physicians on survival of out-of-hospital cardiac arrest: a nationwide population-based observational study. Crit Care 2010;14(6):R199. [doi: 10.1186/cc9319] [Medline: 21050434]
13. Regard S, Rosa D, Suppan M, et al. Evolution of bystander intention to perform resuscitation since last training: web-based survey. JMIR Form Res 2020 Nov 30;4(11):e24798. [doi: 10.2196/24798] [Medline: 33252342]
14. Haley KB, Lerner EB, Pirrallo RG, Croft H, Johnson A, Uihlein M. The frequency and consequences of cardiopulmonary resuscitation performed by bystanders on patients who are not in cardiac arrest. Prehosp Emerg Care 2011;15(2):282-287. [doi: 10.3109/10903127.2010.541981] [Medline: 21250928]
15. Cassan P. Recommendations on basic cardiopulmonary resuscitation: main points. J Eur Urgences Reanim 2021;33(2):65-72. [doi: 10.1016/j.jeurea.2021.05.003]
16. Taramarcaz V, Herren T, Golay E, et al. A short intervention and an interactive e-learning module to motivate medical and dental students to enlist as first responders: implementation study. J Med Internet Res 2022 May 18;24(5):e38508. [doi: 10.2196/38508] [Medline: 35583927]
17. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet e-surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34. [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
18. Complete guide to GDPR compliance. GDPR.eu. URL: https://gdpr.eu/ [accessed 2022-03-21]
19. Suppan L, Herren T, Taramarcaz V, et al. A short intervention followed by an interactive e-learning module to motivate medical students to enlist as first responders: protocol for A prospective implementation study. JMIR Res Protoc 2020 Nov 6;9(11):e24664 [FREE Full text] [doi: 10.2196/24664] [Medline: 33155574]
20. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11. [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
21. Zhang R, Pei J, Wang Y, et al. COVID-19 outbreak improves attractiveness of medical careers in Chinese senior high school students. BMC Med Educ 2022 Apr 4;22(1):241. [doi: 10.1186/s12909-022-03309-7] [Medline: 35379234]
22. Barrense-Dias Y, Urben S, Chok L, Schechter D, Suris JC. Exploration du vécu de la pandémie et du confinement dus à la COVID-19 des adolescent·es et des parents [exploration the COVID-19 pandemic and lockdown experience among teenagers and parents]. : Centre universitaire de médecine générale et santé publique (Unisanté); 2021. [doi: 10.16908/ISSN.1660-7104/320]
23. Guillemette F, Leblanc C. Favoriser l'expression de la motivation chez les étudiants [encouraging motivation among students]. Université du Québec. 2013. URL: https://pedagogie.uquebec.ca/le-tableau/favoriser-lexpression-de-la-motivation-chez-les-etudiants [accessed 2025-04-16]
24. Kusurkar RA, Ten Cate TJ, van Asperen M, Croiset G. Motivation as an independent and a dependent variable in medical education: a review of the literature. Med Teach 2011;33(5):e242-e262. [doi: 10.3109/0142159X.2011.558539] [Medline: 21517676]
25. Viau R. Des conditions à respecter pour susciter la motivation des élèves [conditions to be respected to motivate pupils]. Correspondance 2000 [FREE Full text]
26. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
27. Orsini C, Binnie VI, Wilson SL. Determinants and outcomes of motivation in health professions education: a systematic review based on self-determination theory. J Educ Eval Health Prof 2016 May;13:19. [doi: 10.3352/jeehp.2016.13.19]
28. Viau R, Louis R. Vers une meilleure compréhension de la dynamique motivationnelle des étudiants en contexte scolaire [towards a better understanding of the motivational dynamics of students in a school context]. Can J Educ 1997;22(2):144. [doi: 10.2307/1585904]
29. Dweck CS. Motivational processes affecting learning. American Psychologist 1986;41(10):1040-1048. [doi: 10.1037//0003-066X.41.10.1040]
30. Fundations for a Psychology of Education: Routledge; 1989.
31. Böttiger BW, Semeraro F, Wingen S. "Kids Save Lives": educating schoolchildren in cardiopulmonary resuscitation is a civic duty that needs support for implementation. J Am Heart Assoc 2017 Mar 14;6(3):e005738. [doi: 10.1161/JAHA.117.005738] [Medline: 28292747]

Abbreviations

AED: automatic external defibrillator
BLS: basic life support
OHCA: out-of-hospital cardiac arrest
UGFM: University of Geneva Faculty of Medicine

XSL•FO
RenderX

Edited by A Bahattab; submitted 07.06.24; peer-reviewed by GJ Noordergraaf, J Llego; revised version received 14.03.25; accepted 15.03.25; published 24.04.25.

Please cite as:
Schnetzler N, Taramarcaz V, Herren T, Golay E, Regard S, Mach F, Nasution A, Larribau R, Suppan M, Schiffer E, Suppan L Recruiting Medical, Dental, and Biomedical Students as First Responders in the Immediate Aftermath of the COVID-19 Pandemic: Prospective Follow-Up Study
JMIR Med Educ 2025;11:e63018
URL: https://mededu.jmir.org/2025/1/e63018
doi:10.2196/63018

© Nicolas Schnetzler, Victor Taramarcaz, Tara Herren, Eric Golay, Simon Regard, François Mach, Amanta Nasution, Robert Larribau, Melanie Suppan, Eduardo Schiffer, Laurent Suppan. Originally published in JMIR Medical Education (https://mededu.jmir.org), 24.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Comparison of Learning Outcomes Among Medical Students in Thailand to Determine the Right Time to Teach Forensic Medicine: Retrospective Study

Ubon Chudoung, BSc; Wilaipon Saengon, BBA; Vichan Peonim, MD; Wisarn Worasuwannarak, LLB, MSc, MD

Department of Pathology, Faculty of Medicine Ramathibodi Hospital, Mahidol University, 270 Rama VI Road, Thung Phaya Thai, Bangkok, Thailand

Corresponding Author:
Wisarn Worasuwannarak, LLB, MSc, MD
Department of Pathology, Faculty of Medicine Ramathibodi Hospital, Mahidol University, 270 Rama VI Road, Thung Phaya Thai, Bangkok, Thailand

Abstract

Background: Forensic medicine requires background medical knowledge and the ability to apply it to legal cases. Medical students have different levels of medical knowledge and are therefore likely to perform differently when learning forensic medicine. However, different medical curricula in Thailand deliver forensic medicine courses at different stages of medical study; most curricula deliver these courses in the clinical years, while others offer them in the preclinical years. This raises questions about the differences in learning effectiveness.

Objective: We aimed to compare the learning outcomes of medical students in curricula that either teach forensic medicine at the clinical level or teach it at the preclinical level.

Methods: This was a 5-year retrospective study that compared multiple-choice question (MCQ) scores in a forensic medicine course for fifth- and third-year medical students. The fifth-year students' program was different from that of the third-year students, but both programs were offered by Mahidol University. The students were taught forensic medicine by the same instructors, used similar content, and were evaluated via examinations of similar difficulty. Of the 1063 medical students included in this study, 782 were fifth-year clinical students, and 281 were third-year preclinical students.

Results: The average scores of the fifth- and third-year medical students were 76.09% (SD 6.75%) and 62.94% (SD 8.33%), respectively. The difference was statistically significant (Kruskal-Wallis test: $P<.001$). Additionally, the average score of fifth-year medical students was significantly higher than that of third-year students in every academic year (all P values were <.001).

Conclusions: Teaching forensic medicine during the preclinical years may be too early, and preclinical students may not understand the clinical content sufficiently. Attention should be paid to ensuring that students have the adequate clinical background before teaching subjects that require clinical applications, especially in forensic medicine.

(JMIR Med Educ 2025;11:e57634) doi:10.2196/57634

KEYWORDS

multiple-choice question; MCQ; forensic medicine; preclinic; clinic; medical student

Introduction

Forensic medicine is a crucial field that intersects with the legal system. It involves the collection, analysis, interpretation, and presentation of evidence in legal cases [1]. Forensic medicine plays an essential role in assisting courts with making correct decisions by providing reliable and timely information. It also plays a critical role in protecting peoples' rights by ensuring that their legal, civil, and human rights are upheld throughout the legal process [2]. Furthermore, studying forensic medicine is important for medical students in different countries, as they are equipped with the necessary knowledge and skills to accurately assess and document injuries and provide expert opinions on causes of death and other relevant medical information that may have legal implications [3-6].

This subject is included among the professional subjects that every Thai medical student must study to comply with the Criminal Procedure Code of Thailand, which requires physicians working in public hospitals to be able to perform postmortem inquests with police in cases where no forensic physician is available [7]. The Medical Council of Thailand has included forensic medicine as a mandatory subject in every doctor of medicine program.

The doctor of medicine programs in Thailand are 6-year programs conducted after graduating from high school. They are generally divided into 3 years at the preclinical level (first through third year) and another 3 years at the clinical level (fourth through sixth year). The teaching of each university's curriculum differs in detail depending on various factors, such as the number of students, number of teachers, location, and

XSL•FO
RenderX

service characteristics. Forensic medicine is subject to these differences.

Studying forensic medicine involves dealing with dead bodies, crime scenes, and traumatic injuries that can be emotionally and mentally stressful for some students [8]. A study from Saudi Arabia revealed that medical students have poor attitudes toward and awareness of the importance of forensic medicine [9]. Additionally, forensic medicine courses cover a wide range of topics, such as anatomy, physiology, pathology, toxicology, psychology, and jurisprudence, which can be difficult to master and integrate [10,11].

Students with different levels of medical knowledge may experience different forensic medicine course outcomes. In Thailand, most medical curricula are currently designed to teach forensic medicine to medical students at the clinical level (fifth year) [12-14]. However, some curricula have been designed to teach forensic medicine to medical students at the preclinical or early clinical level (third or fourth year) [15]. There are no clear guidelines regarding the level of students who should be taught forensic medicine.

This study aims to compare the learning outcomes of medical students in a curriculum that teaches forensic medicine at the clinical level and those of medical students in a curriculum that teaches forensic medicine at the preclinical level.

Methods

Study Design

This retrospective study was conducted to compare multiple-choice question (MCQ) scores of fifth- and third-year medical students from two medical curricula that teach forensic medicine. Both groups of students studied forensic medicine with the same instructors, used similar content, and were assessed via MCQ examinations with similar difficulty levels. The scores indicated the participants' learning outcomes.

Setting and Participants

Samples

Our samples included (1) medical students in a curriculum that teaches forensic medicine at the clinical level (fifth year) through the Doctor of Medicine Program at Ramathibodi Hospital, Mahidol University (782 students), and (2) medical students in a curriculum that teaches forensic medicine as the last subject at the preclinical level (third year) through the Joint Program for Producing More Doctors for Rural Areas, Mahidol University (281 students).

Sample Size Calculation

The sample size was designed to compare 2-sided differences in the MCQ percentage scores between third- and fifth-year medical students studying forensic medicine. The null hypothesis (H_0) was that the MCQ percentage scores between third- and fifth-year medical students would not be significantly different. The alternative hypothesis (H_1) was that the MCQ percentage scores between third- and fifth-year medical students would be significantly different.

We calculated the sample size according to a 5% type 1 error (α) and an 80% study power ($1 - \beta$). The significant difference ($\mu 1 - \mu 2$) and SD (σ) were set at 10 and 11, respectively, based on MCQ score data for medical students who studied forensic medicine from 2010 to 2014. The required sample size was 38 (19 participants in each group; Multimedia Appendix 1) [16]. However, this study included more participants than the calculated sample size.

Intervention

Teaching Method

Both groups of medical students received on-site theoretical lectures before completing the MCQs. The content included basic knowledge of forensic pathology (including postmortem inquest, identification, time of death estimation, crime scene investigation, unnatural death, and sudden unexpected death), clinical forensic medicine (including patients who are wounded, child abuse, sexual assault, and forensic psychiatry), forensic evidence, forensic genetics, forensic toxicology, and medical law and ethics. Third-year medical students studied for 30 hours. Fifth-year medical students studied for 15 hours, using similar content that was more concise, and had the opportunity to visit a court for 3 hours. Neither group had the opportunity to attend crime scene investigations or autopsies (which they would attend later). This teaching method was performed regularly, and the authors did not intervene with any of the participants.

MCQ Examinations

For examinations, all teaching staff (4 staff members) created 5-option MCQs with a single best answer according to the topics they taught, including basic knowledge of forensic pathology (40% of questions), clinical forensic medicine (30% of questions), forensic evidence (5% of questions), forensic genetics (5% of questions), forensic toxicology (5% of questions), and medical law and ethics (15% of questions). The tests were designed to ensure that medical students are able to perform basic postmortem inquests, examine various types of forensic patients, produce accurate medicolegal reports, have basic knowledge of law and ethics, and understand the process of testifying in court. The MCQ examinations were structured via a balanced approach for cognitive function, allocating approximately 25% of the examination to knowledge, 30% to comprehension, 25% to application, and 20% to analysis level, according to the Bloom taxonomy. This distribution is maintained consistently from year to year. The examination was intended to have a moderate level of difficulty. Third-year medical students completed a 100-question examination in 2 hours, and fifth-year medical students completed an 80-question examination in 1.5 hours. Based on an analysis of the examination, most of the items had a difficulty level (p) in the range of 0.4 to 0.7 and a discriminatory power (r) in the range of 0.1 to 0.5. Internal consistency reliability (Kuder-Richardson Formula 20) was in the range of 0.6 to 0.7.

Data Collection

In this study, the data were collected retrospectively for 5 years, from academic years 2010 through 2014.

XSL•FO
RenderX

Statistical Analysis

For the comparison between the two groups, we used the means and SDs of the MCQ scores to test this study's hypothesis that the learning outcome is different between third- and fifth-year students. Kruskal-Wallis and Mann-Whitney *U* tests were used for continuous variables with normal and nonnormal distributions, respectively [17]. The significance level was set at 5% ($P<.05$). The program used for data analysis was SPSS software (version 26; IBM Corp).

Ethical Considerations

This study was approved by the Ethical Clearance Committee on Human Rights Related to Research Involving Human Subjects, Faculty of Medicine Ramathibodi Hospital, Mahidol University (MURA 2015/213). The need for informed consent was waived by the Ethical Clearance Committee on Human Rights Related to Research Involving Human Subjects, Faculty of Medicine Ramathibodi Hospital, Mahidol University. Data were collected by using an anonymous method—assigning numbers to all participants instead of names. No compensation was provided to participants.

Results

From the collection of MCQ scores of medical students from academic years 2010 to 2014 who were taught forensic medicine, the scores of 1063 students were used in this study. The scores were divided into scores of third-year medical students (n=281) and scores of fifth-year medical students (n=728), as shown in Table 1.

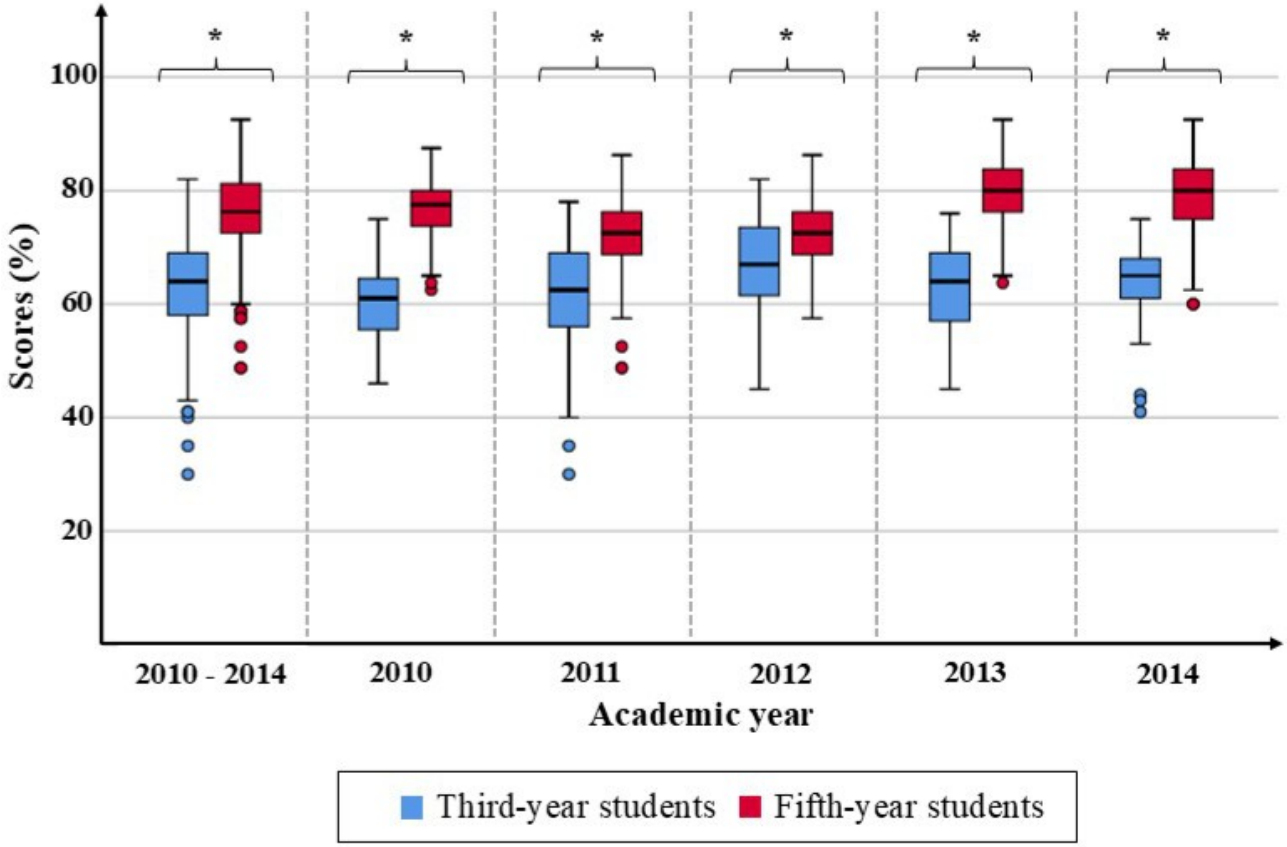
Table . Number of students in each academic year (N=1063).

Students		Academic year					Total
		2010	2011	2012	2013	2014	
Third-year students, n (%)							
	Male	30 (2.8)	35 (3.3)	34 (3.2)	33 (3.1)	33 (3.1)	165 (15.5)
	Female	21 (2)	23 (2.2)	21 (2)	23 (2.2)	28 (2.6)	116 (10.9)
Fifth-year students, n (%)							
	Male	81 (7.6)	94 (8.8)	87 (8.2)	94 (8.8)	101 (9.5)	457 (43)
	Female	53 (5)	64 (6)	71 (6.7)	64 (6)	73 (6.9)	325 (30.6)
Total, n (%)		185 (17.4)	216 (20.3)	213 (20)	214 (20.1)	235 (22.1)	1063 (100)

When comparing students' scores, it was found that fifth-year medical students had an average score of 76.09% (SD 6.75%), which was higher than that of third-year medical students (mean 62.94%, SD 8.33%). The difference was statistically significant (Kruskal-Wallis test: $P<.001$). In addition, when comparing the average scores in each academic year, it was found that the average score of fifth-year medical students was significantly higher than that of third-year students in every academic year (Mann-Whitney *U* test: all *P* values were <.001), as shown in Figure 1.

XSL•FO
RenderX

Figure 1. Comparing scores of third-year and fifth-year students. *Statistically significant (Mann-Whitney *U* test: *P*<.001).



Discussion

Principal Findings

According to this study's findings, fifth-year medical students achieved significantly higher marks on MCQs than those achieved by third-year medical students, despite the latter having more opportunities to prepare and take examinations due to their longer duration of study. The fact that the two groups of medical students had different scores may be due to their different levels of basic knowledge of medicine. Fifth-year medical students study basic clinical subjects. Therefore, they may have more comprehensive and complete basic medical knowledge and may be able to apply it to prove facts about legal cases better than third-year medical students who have not completed their basic clinical subjects. These results are consistent with a study in Italy, which showed that students' awareness of forensic medicine improved in the fifth or sixth year of a forensic medicine course [18].

When analyzing the data by academic year, fifth-year medical students still had higher MCQ scores than those of third-year medical students, with statistical significance for each academic year. These data show that the difference in MCQ scores was unlikely due to different medical students from year to year.

In forensic medicine, students should have the opportunity to learn about real cases, including examinations of legal patients, autopsies, and crime scene examinations. This would improve students' understanding of applying and ability to apply medical knowledge to legal applications. A study in India revealed that a court visit in a real scenario was the method that generated the most interest, and student-led objective tutorials comprised the method that best facilitated enhanced learning; the "model answer" method was also found to be an effective method for teaching forensic medicine [19]. Furthermore, a study in Mexico showed that crime scene investigation laboratory visits are an innovative method of learning that may help broaden medical students' perspectives on forensic sciences and help them understand the multidisciplinary processes of crime investigation [20].

By integrating forensic medicine into the medical curriculum, students also gain a deeper awareness of the complexities surrounding child abuse. Training on this topic not only enhances students' diagnostic skills but also instills a sense of responsibility to act in the best interests of the child, ensuring that they are better prepared to contribute to the early detection, intervention, and prevention of child abuse in their future careers [21].

This study used only MCQ scores from theoretical teaching, which may not measure all of the knowledge and skills of students. Although MCQs can test higher-order thinking, they are typically limited to the "application" and "analysis" levels of the Bloom taxonomy [22]. The use of MCQs is often driven by practical concerns, such as large class sizes, rather than pedagogical reasons. Although MCQs have their place, they may restrict the scope of teaching and require careful consideration to align with higher-order learning objectives [23]. Thus, a combination of test methods can be used. A study from Nepal found that objective structured practical examination is an acceptable and well-received method for medical students [24].

XSL•FO
RenderX

Integrating some content of clinical subjects via vertical integration for preclinical medical students may help to enhance their knowledge and understanding of forensic medicine. A previous study on learning environments found that undergraduate medical students from Egypt who received integrated curriculum teaching experienced a more positive learning environment [6]. Further, a similar study from Malaysia showed that integrated teaching positively affects medical students' learning environment [25]. These studies are also consistent with guidelines from the Medical Council of Thailand for developing medical curricula in Thailand, which support horizontal and vertical integration teaching [26]; that is, clinical teachers should teach about clinical experiences from the beginning and integrate basic medical science knowledge into the clinical years.

Limitations

A limitation of this study was its retrospective design; that is, past MCQ scores were analyzed to evaluate the medical curricula at the time of writing. No systematic interventions were conducted to test the hypothesis. In addition, this study used only MCQ scores; therefore, it may not include every learning outcome of the forensic medicine course.

Recommendations

Students' basic medical knowledge should be considered when teaching and learning subjects that require clinical application, especially in forensic medicine, which applies medical knowledge to law. Teaching such subjects to preclinical-level students, whose medical knowledge remains incomplete, may be too ambitious. It may be appropriate to integrate introductory content from clinical subjects to increase knowledge and understanding. In comparison, clinical-level students with complete basic knowledge may be more suitable for such clinical subjects.

Conclusion

Forensic medicine requires basic medical knowledge and the ability to apply this knowledge in legal cases. Students' basic medical knowledge should be considered when planning the teaching and learning of this subject. Teaching forensic medicine in the preclinical years may be too early, and doing so may result in students being unable to sufficiently understand the clinical content.

Acknowledgments

We would like to thank our colleagues, including the staff and officers of the Division of Forensic Medicine, for the support they have given us.

Data Availability

The datasets used and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: UC
Data curation: UC
Formal analysis: UC
Investigation: UC
Methodology: WW
Project administration: WW
Supervision: VP, WW
Validation: WW
Visualization: WW
Writing – original draft: UC
Writing – review & editing: UC, WS, VP, WW

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sample size calculation.
[DOCX File, 15 KB - mededu_v11i1e57634_app1.docx]

References

1. Shepherd R. Simpson's Forensic Medicine, 12th edition: Arnold; 2003.
2. Payne-James J. History and development of clinical forensic medicine. In: Clinical Forensic Medicine: A Physician's Guide: Humana Press; 2005:1-36. [doi: 10.1385/1-59259-913-3:001]

XSL•FO
RenderX

3. Tóth D, Petrus K, Heckmann V, Simon G, Poór VS. Application of photogrammetry in forensic pathology education of medical students in response to COVID-19. J Forensic Sci 2021 Jul;66(4):1533-1537. [doi: 10.1111/1556-4029.14709] [Medline: 33764562]
4. Sosa-Reyes AM, Villavicencio-Queijeiro A, Suzuri-Hernández LJ. Interdisciplinary approaches to the teaching of forensic science in the Forensic Science Undergraduate Program of the National Autonomous University of Mexico, before and after COVID-19. Sci Justice 2022 Nov;62(6):676-690. [doi: 10.1016/j.scijus.2022.08.006] [Medline: 36400489]
5. Marambe KN, Edussuriya DH, Somaratne PDIS, Piyaratne C. Do medical students who claim to be using deep learning strategies perform better at the Forensic Medicine examination? South-East Asian Journal of Medical Education 2009 Jun 30;3(1):25-30. [doi: 10.4038/seajme.v3i1.464]
6. Fayed MM, Abdo SA, Sharif AF. Preclinical and clinical medical students' perception of the learning environment: a reference to the Forensic Medicine and Clinical Toxicology course. Adv Med Educ Pract 2022 Apr 23;13:369-406. [doi: 10.2147/AMEP.S354446] [Medline: 35494484]
7. Thailand Criminal Procedure Code sections 148-156. Thai Law Forum. 2024 Jul 3. URL: http://www.thailawforum.com/thailand-criminal-procedure-code-sections-148-156/ [accessed 2025-01-28]
8. Papadodima SA, Sergentanis TN, Iliakis RG, Sotiropoulos KC, Spiliopoulou CA. Students who wish to specialize in forensic medicine vs. their fellow students: motivations, attitudes and reactions during autopsy practice. Adv Health Sci Educ Theory Pract 2008 Nov;13(4):535-546. [doi: 10.1007/s10459-007-9065-3] [Medline: 17486420]
9. Madadin MS. Assessment of knowledge about, attitudes toward, and awareness of a forensic medicine course among medical students at the University of Dammam. J Forensic Leg Med 2013 Nov;20(8):1108-1111. [doi: 10.1016/j.jflm.2013.10.003] [Medline: 24237831]
10. Wyatt JP, Squires T, Norfolk G, Payne-James J. Oxford Handbook of Forensic Medicine: Oxford University Press; 2011. URL: https://academic.oup.com/book/29998 [accessed 2026-01-17] [doi: 10.1093/med/9780199229949.001.0001]
11. Levinson SA, Muehlberger CW. An introductory course in legal medicine for medical students. Acad Med 1934 Sep;9(5):293-301. [doi: 10.1097/00001888-193409000-00006]
12. Faculty of Medicine Ramathibodi Hospital, Mahidol University. Doctor of Medicine Program, Revised Curriculum 2020 [Article in Thai]. Mahidol University. 2020. URL: https://www.rama.mahidol.ac.th/meded/sites/default/files/public/img2024/course/%E0%B8%A1%E0%B8%84%E0%B8%AD.2%20%E0%B8%AB%E0%B8%A5%E0%B8%B1%E0%B8%81%E0%B8%AA%E0%B8%B9%E0%B8%65%E0%B8%A3%20%E0%B8%9E%E0%B8%9A-63-pdf%20%E0%B8%A3%E0%B8%A7%E0%B8%A1-110863.pdf [accessed 2024-08-24]
13. Doctor of Medicine Program. Chiang Mai University. 2023. URL: https://www.cmu.ac.th/en/Faculty/course_detail/939ca112-78cc-444c-bfa4-5328c0484b7e [accessed 2024-08-12]
14. Chulalongkorn University. Doctor of Medicine Program (Revised Curriculum 2017) [Article in Thai]. Chulalongkorn University. 2017. URL: https://bhumibol-med.com/Media/media-2017-11-19-04-35-32.pdf [accessed 2024-08-12]
15. Clinical Medical Education Center, Sawanpracharak Hospital. Course Specification: Forensic Medicine 1 (NVEF411). Revised January 13, 2014. [Article in Thai]. Mahidol University. 2014. URL: https://www.rama.mahidol.ac.th/patho/sites/default/files/public/file/NVEF411.pdf [accessed 2024-08-12]
16. Noordzij M, Tripepi G, Dekker FW, Zoccali C, Tanck MW, Jager KJ. Sample size calculations: basic principles and common pitfalls. Nephrol Dial Transplant 2010 May;25(5):1388-1393. [doi: 10.1093/ndt/gfp732] [Medline: 20067907]
17. De Muth JE. Overview of biostatistics used in clinical research. Am J Health Syst Pharm 2009 Jan 1;66(1):70-81. [doi: 10.2146/ajhp070006] [Medline: 19106347]
18. Aulino G, Beccia F, Siodambro C, et al. An evaluation of Italian medical students attitudes and knowledge regarding forensic medicine. J Forensic Leg Med 2023 Feb;94:102484. [doi: 10.1016/j.jflm.2023.102484] [Medline: 36640545]
19. Gupta S, Parekh UN, Ganjiwale JD. Student's perception about innovative teaching learning practices in forensic medicine. J Forensic Leg Med 2017 Nov;52:137-142. [doi: 10.1016/j.jflm.2017.09.007] [Medline: 28922654]
20. Eraña-Rojas IE, López Cabrera MV, Ríos Barrientos E, Membrillo-Hernández J. A challenge based learning experience in forensic medicine. J Forensic Leg Med 2019 Nov;68:101873. [doi: 10.1016/j.jflm.2019.101873] [Medline: 31627125]
21. Aulino G, Beccia F, Rega M, et al. Child maltreatment and management of pediatric patients during COVID-19 pandemic: knowledge, awareness, and attitudes among students of medicine and surgery. A survey-based analysis. Front Public Health 2022 Sep 20;10:968286. [doi: 10.3389/fpubh.2022.968286] [Medline: 36203705]
22. Ehsan SB. Effectiveness of MCQs in assessing higher order cognition. Biomedica 2017;33(4):269-272 [FREE Full text]
23. Liu Q, Wald N, Daskon C, Harland T. Multiple-choice questions (MCQs) for higher-order cognition: perspectives of university teachers. Innovations in Education and Teaching International 2023 Jun 8;61(4):802-814. [doi: 10.1080/14703297.2023.2222715]
24. Menezes RG, Nayak VC, Binu VS, et al. Objective structured practical examination (OSPE) in forensic medicine: students' point of view. J Forensic Leg Med 2011 Nov;18(8):347-349. [doi: 10.1016/j.jflm.2011.06.011] [Medline: 22018165]
25. Yusoff MSB, Jaa'far R, Arzuman H, Arifin WN, Pa MNM. Perceptions of medical students regarding educational climate at different phases of medical training in a Malaysian medical school. Education in Medicine Journal 2013 Sep 1;5(3):e30-e41. [doi: 10.5959/eimj.v5i3.146]

XSL•FO
RenderX

26. Medical Council Announcement No. 76/2023 on the criteria for requesting to open/improve the Doctor of Medicine program and certify medical school institutions in 2023 [Article in Thai]. The Medical Council of Thailand. 2023. URL: https://tmc.or.th/index.php/News/Announcement/1089 [accessed 2024-08-12]

Abbreviations

MCQ: multiple-choice question

Edited by B Lesselroth; submitted 21.02.24; peer-reviewed by D Reham, G Aulino; revised version received 12.08.24; accepted 17.12.24; published 10.02.25.

Please cite as:
Chudoung U, Saengon W, Peonim V, Worasuwannarak W
Comparison of Learning Outcomes Among Medical Students in Thailand to Determine the Right Time to Teach Forensic Medicine: Retrospective Study
JMIR Med Educ 2025;11:e57634
URL: https://mededu.jmir.org/2025/1/e57634
doi:10.2196/57634

© Ubon Chudoung, Wilaipon Saengon, Vichan Peonim, Wisarn Worasuwannarak. Originally published in JMIR Medical Education (https://mededu.jmir.org), 10.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding Community Health Care Through Problem-Based Learning With Real-Patient Videos: Single-Arm Pre-Post Mixed Methods Study

Kiyoshi Shikino[1,2,3], MD, MHPE, PhD; Kazuyo Yamauchi[1,2], MD, MHPE, PhD; Nobuyuki Araki[1,2], MD, PhD; Ikuo Shimizu[2,4], MD, MHPE, PhD; Hajime Kasai[2,4], MD, PhD; Tomoko Tsukamoto[2,3], MD, PhD; Hiroshi Tajima[2,4], MD, PhD; Yu Li[2,3], MD, PhD; Misaki Onodera[4], PhD; Shoichi Ito[1,2,4], MD, PhD

[1]Chiba University Graduate School of Medicine, Community-Oriented Medical Education, Chiba, Japan

[2]Health Professional Development Center, Chiba University Hospital, Chiba University, Chiba, Japan

[3]Department of General Medicine, Chiba University Hospital, Chiba University, Chiba, Japan

[4]Department of Medical Education, Chiba University Graduate School of Medicine, Chiba University, Chiba, Japan

Corresponding Author:
Kiyoshi Shikino, MD, MHPE, PhD
Chiba University Graduate School of Medicine
Community-Oriented Medical Education
1-8-1, Inohana, Chu-ou-ku
Chiba, 2608670
Japan
Phone: 81 43 222 7171
Email: kshikino@gmail.com

Abstract

Background: Japan faces a health care delivery challenge due to physician maldistribution, with insufficient physicians practicing in rural areas. This issue impacts health care access in remote areas and affects patient outcomes. Educational interventions targeting students' career decision-making can potentially address this problem by promoting interest in rural medicine. We hypothesized that community-based problem-based learning (PBL) using real-patient videos could foster students' understanding of community health care and encourage positive attitudes toward rural health care.

Objective: This study investigated the impact of community-based PBL on medical students' understanding and engagement with rural health care, focusing on their knowledge, skills, and career orientation.

Methods: Participants were 113 fourth-year medical students from Chiba University, engaged in a transition course between preclinical and clinical clerkships from October 24 to November 2, 2023. The students were randomly divided into 16 groups (7-8 participants per group). Each group participated in two 3-hour PBL sessions per week over 2 consecutive weeks. Quantitative data were collected using pre- and postintervention questionnaires, comprehension tests, and tutor-assessed rubrics. Self-assessment questionnaires evaluated the students' interest in community health care and their ability to envision community health care settings before and after the intervention. Qualitative data from the students' semistructured interviews after the PBL sessions assessed the influence of PBL experience on clinical clerkship in community hospitals. Statistical analysis included median (IQR), effect sizes, and P values for quantitative outcomes. Thematic analysis was used for qualitative data.

Results: Of the 113 participants, 71 (62.8%) were male and 42 (37.2%) female. The total comprehension test scores improved significantly (pretest: median 4.0, IQR 2.5-5.0; posttest: median 5, IQR 4-5; $P<.001$; effect size r=0.528). Rubric-based assessments showed increased knowledge application (pretest: median 8, IQR 7-9; posttest: median 8, IQR 8-8; $P<.001$; r=0.494) and self-directed learning (pretest: median 8, IQR 7-9; posttest: median 8, IQR 8-8; $P<.001$; r=0.553). Self-assessment questionnaires revealed significant improvements in the students' interest in community health care (median 3, IQR 3-4 to median 4, IQR 3-4; $P<.001$) and their ability to envision community health care settings (median 3, IQR 3-4 to median 4, IQR 3-4; $P<.001$). Thematic analysis revealed key themes, such as "empathy in patient care," "challenges in home health care," and "professional identity formation."

Conclusions: Community-based PBL with real-patient videos effectively enhances medical students' understanding of rural health care settings, clinician roles, and the social needs of rural patients. This approach holds potential as an educational strategy

XSL•FO
RenderX

to address physician maldistribution. Although this study suggests potential for fostering positive attitudes toward rural health care, further research is needed to assess its long-term impact on students' career trajectories.

(JMIR Med Educ 2025;11:e68743) doi:10.2196/68743

KEYWORDS

community health care; community-oriented medical education; mixed method; problem-based learning; real-patient video

Introduction

Japan faces a significant health care delivery challenge owing to uneven physician distribution, notably affecting rural areas and community hospitals [1,2]. This maldistribution exacerbates community hospitals' challenges [3-5]. This issue is not confined to Japan; it impacts countries worldwide [6-11]. In 2019, the Ministry of Health, Labour and Welfare introduced the physician uneven distribution index as part of an intervention policy addressing prefectural geographical disparities in physician distribution [1,12-14]; it assesses the extent of physician maldistribution by evaluating prefectural medical supply and demand.

To combat the physician maldistribution, community hospital training has been integrated into second-year resident physicians' compulsory curriculum [15-17], highlighting the necessity of preparing future physicians with the competencies required to effectively meet rural communities' health care needs. Moreover, introducing community medicine principles early in medical education is an acknowledged need [18].

Japanese medical schools have begun to proactively adopt problem-based learning (PBL) as a foundational step before clinical rotations [19]. PBL emphasizes real-life medical scenarios, cultivating students' clinical reasoning and decision-making skills [20]. PBL prepares students for clinical rotations with an enriched understanding of community health care's challenges and prospects [21]. This approach bolsters medical students' clinical training and supports the alleviation of physician maldistribution by promoting community or rural medicine careers.

However, although PBL has been implemented in medical education settings, its integration with community-oriented medicine in addressing physician maldistribution remains underexplored. Furthermore, despite the global relevance of physician maldistribution, studies focusing on innovative educational interventions targeting this issue are limited [22-24].

We hypothesized that incorporating real-patient videos into community-focused PBL would significantly improve students' capacity to make well-informed career choices and identify with positive role models. This study addressed the aforementioned research gap by examining this approach's effectiveness within Japanese medical education.

Methods

Study Design

This study used an explanatory sequential mixed methods design following a pragmatic approach [25-27], capitalizing on quantitative and qualitative designs' strengths while minimizing their shortcomings. Furthermore, it allowed researchers to better understand experimental results while incorporating participants' perspectives. The National Institutes of Health advises a mixed methods approach "to improve the quality and scientific power of data" and to better address the complex issues facing health sciences today, including health professional education [28,29]. In the qualitative analysis, medical students' reflection papers were text-mined to analyze the word frequency in community-oriented PBL. Next, we conducted individual interviews with medical students during their clinical clerkship.

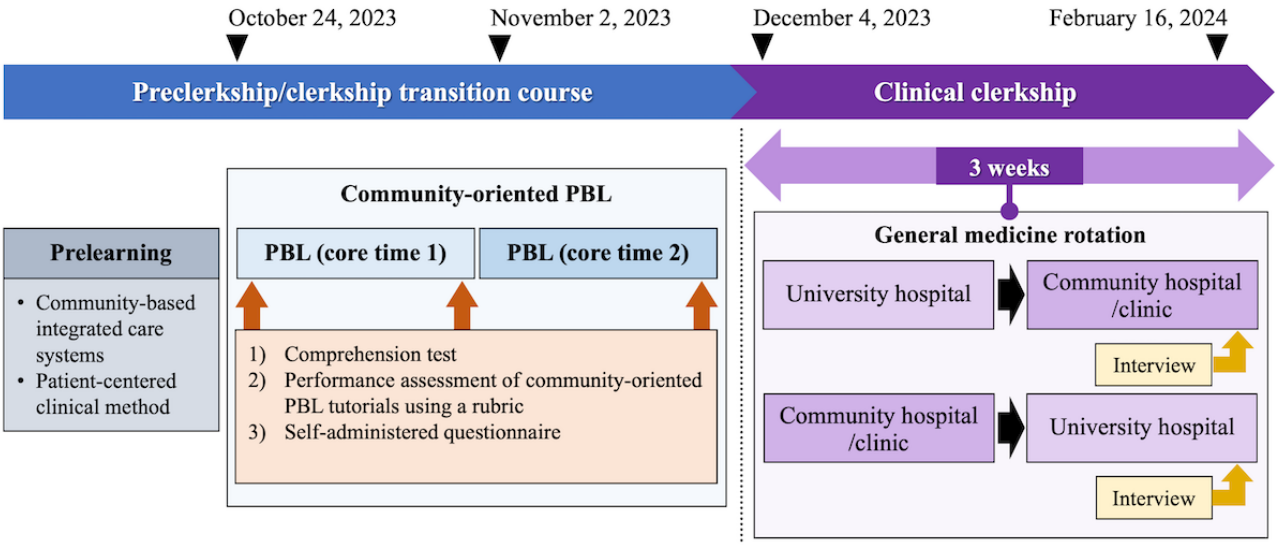
Participants and Trial Design

This study was conducted as part of the undergraduate medical curriculum at Chiba University, Japan. Community-oriented PBL was conducted from October 24 to November 2, 2023, as part of a preclerkship/clerkship transition course [30], a 5-week preparatory education period before clinical clerkship. Additionally, PBL is integrated into the Chiba University medical school's curriculum in years 1-4, and all students experience it. Participants were 113 fourth-year medical students who had attended lectures and received simulation training in basic and clinical medicine. To minimize potential biases and ensure an even distribution of student characteristics, participants were randomly divided into 16 groups of 7-8 each using the random number generation function in Microsoft Excel. The 16 groups received community-oriented PBL of a patient case using a community-based integrated care system with real-patient videos.

Quantitative data were gathered using the comprehension test, a tutor assessment with rubrics during the core time, and pre- and postintervention questionnaires to assess community health care perceptions (Figure 1). Additionally, a qualitative evaluation assessed community health care perceptions using a free-text reflection paper after PBL and follow-up interviews during clinical clerkships in community hospitals or clinics. The study used a mixed methods, sequential explanatory design to integrate the results [25,27,31].

XSL•FO
RenderX

Figure 1. Timeline from PBL in preclerkship/clerkship transition course to general medicine rotation in clinical clerkship. PBL: problem-based learning.



Community-Oriented PBL Educational Intervention

Real-patient videos were meticulously prepared to enhance the learning experience authenticity, simulating real-life scenarios in community health care settings (Multimedia Appendix 1). Prepared in collaboration with medical education experts, community health care professionals, and audiovisual production specialists, these videos aimed to accurately depict home health care characteristics, including medical interviews, physical examinations, and in-home patient interactions. Real patients participated in the production process under strict ethical guidelines to ensure authenticity and respect for patient privacy, emphasizing community health care's unique challenges and dynamics.

Each real-patient video was carefully scripted and filmed to represent common community health care situations, encompassing multiple scenes depicting different patient care stages and interactions, ranging from initial patient assessments in a community hospital setting to follow-up in-home visits. The video durations ranged from 3 to 5 minutes. Along with the videos, patient information sheets and tasks were presented to the students (Multimedia Appendix 2). One of the patient's primary conditions was underlying diabetes, and they presented with severe lower leg edema. The case involved transitioning from acute care to a chronic care hospital, followed by the introduction of home visit medical services. Patient consent was obtained for video use, and students were instructed to adhere to confidentiality guidelines.

During the community-oriented PBL sessions, students collectively viewed the real-patient videos in designated classrooms equipped with audiovisual facilities, allowing for simultaneous viewing on shared screens. Before watching the videos, students were divided into small groups and assigned a tutor to facilitate discussions and learning activities. The tutors observed whether the students could achieve the learning objectives and facilitated discussions. The tutors, randomly selected faculty members in medical education and community-oriented medical education, were given standardized instructions and materials before the sessions to ensure consistency and effectiveness [32].

Community-oriented PBL sessions were divided into 2 sessions per case, each lasting approximately 3 hours. In the first session (core time 1), students were presented with the patient's history and physical examination findings. In the second session (core time 2), the investigation findings and treatment plans were discussed. The same case scenario was used for all students.

Quantitative Measures

Comprehension Test

The comprehension test assessed the minimum essential knowledge required for problem solving in PBL, focusing on holistic medicine, patient-centered care, and the International Classification of Functioning, Disability, and Health (ICF). It comprised 5 multiple-choice questions (Q1-Q5, Multimedia Appendix 3). The test was administered as a pretest at the beginning of core time 1 and a posttest after core time 2, allowing for a comparison of test scores.

The comprehension test items were developed specifically for this study and have not been used in other contexts. They were based on the core learning objectives of the PBL sessions, which included understanding the structure of community health care systems, application of the ICF, and principles of patient-centered care. To ensure clarity and alignment with learning objectives, the questions underwent cognitive debriefing by faculty members in medical education. This process involved reviewing each question for relevance, accuracy, and comprehensibility, with feedback incorporated into the final version to enhance content validity.

Rubric-Based Performance Assessment

In addition to the comprehension test, each student's performance during the PBL sessions was assessed using a rubric. The rubric was determined based on previous studies after the authors discussed the validity of the criteria [33,34].

XSL•FO
RenderX

It evaluated 5 key dimensions: knowledge application, comprehensive care process, self-regulated learning, learning motivation, and communication skills with peers. Self-regulated learning refers to the ability of students to plan, monitor, and reflect on their learning process, fostering autonomy and adaptability in problem-solving contexts [35]. Each of the 5 dimensions was quantitatively evaluated on a 10-point scale (Multimedia Appendix 4). Performance assessments were conducted before and after the educational intervention to measure changes in the competencies.

Self-Administered Questionnaire

Students completed questionnaires before and after community-oriented PBL (Multimedia Appendix 5). They were assigned identification numbers to preserve their anonymity. Data were collected using a self-administered 5-point Likert-scale questionnaire ranging from 1 (strongly disagree) to 5 (strongly agree). The criteria were informed by previous studies and refined by the authors through discussions in focus groups [36,37]. After community-oriented PBL, 2 items ("I am interested in community health care" and "I can envision a community health care setting") were surveyed. The items assessed the students' interest in community health care and their ability to visualize a community health care setting.

Sample Size

This study also served as an educational program for fourth-year medical students in a basic clinical clerkship course. Altogether, 113 medical students from 12 groups were recruited. For quantitative data, the sample size required a 2-tailed *t* test of the difference between the pre- and post-PBL means, assuming a significance level of .05, a power of 0.8, and an effect size of 0.5. When the Mann-Whitney *U* test was conducted with those values, the required sample size was 54 in each group, totaling 108.

Data Analysis

All statistical analyses of quantitative data were conducted using SPSS Statistics for Microsoft Windows version 29.0 (IBM Corp), with a significance level under 5% for each analysis. The comprehension test results, including total scores and individual question responses (Q1-Q5), were analyzed using the Wilcoxon signed rank test for paired total scores. Additionally, the McNemar test was used to compare pre- and post-PBL correct response rates for individual questions. For rubric-based performance assessment, the Wilcoxon signed rank test was used to compare scores from core time 1 and core time 2 for total scores and individual rubric items. Effect sizes were calculated for all analyses: r values were derived from z scores for the Wilcoxon signed rank test, and the Cohen w value was calculated for the McNemar test.

Qualitative Measures

Follow-Up Interviews During Clinical Clerkships

Semistructured interviews (average duration: 20 minutes) with individual medical students were conducted by authors KS, KY, and NA. All sessions were recorded and transcribed verbatim, and interviews were conducted iteratively. An interview guide containing open-ended questions was constructed deductively based on the research question and thematic analysis findings. This guide was modified after the first 9 interviews to address emerging and previously unexplored themes in subsequent interviews. The interview participants received no gifts for participating.

Interview transcripts were analyzed using a template analysis approach [38,39]. An inductive code template was defined based on the research questions, thematic analysis findings, and interview guide. The initial template was developed through independent coding (performed by authors KS and IS) of the first 9 interviews. The template was further developed by coding the subsequent interviews. Regarding version 2 of the template, after coding 3 interviews, KS, NA, and IS agreed that the template adequately covered all texts. KS and IS individually coded the remaining transcripts using the template. At this stage, authors KY and SI discussed all further changes or additions to the template until they reached a consensus. After coding all 12 interviews, no additional changes were made to the templates. The final code template was further confirmed by analyzing the remaining 12 transcripts, which can be interpreted as a sign that code saturation was reached [40].

A qualitative evaluation was conducted to assess the acquisition of higher-order intellectual skills in which an interview was conducted after PBL and clinical clerkship. In the clinical clerkship interview, community-oriented PBL's effectiveness in improving clinical performance in home visit care was investigated. The interviewers (KS, KY, and NA) discussed the content and developed an interview guide. Students were asked the following open-ended question: "What is the effectiveness of community-oriented PBL for clinical clerkship?" The interviews were administered by 3 faculty members to 13 students from community-oriented PBL groups during their clinical clerkship. All target students had experienced home visits in their clinical clerkship 1-3 months after community-oriented PBL. The interviewers were trained facilitators from the faculty overseeing community-oriented PBL and conducted thematic analysis. Two researchers (KS and NA) independently read and coded the transcripts. Researcher triangulation was conducted in which the same 2 researchers conducted the analysis and consensus building.

KS and IS, who have extensive experience in qualitative research, defined and regularly discussed the themes and subthemes from the data to ensure the results' reliability. The cognitive process dimensions to which they corresponded were also evaluated.

Ethical Considerations

This study was approved by the Ethics Review Committee of the Graduate School of Medicine, Chiba University (approval number 3425). The procedures for obtaining informed consent were explained to the medical students, who were also informed that this study would not affect their grades. All data collected in this study were anonymized to ensure privacy and confidentiality. Participants did not receive any compensation for their participation in this study.

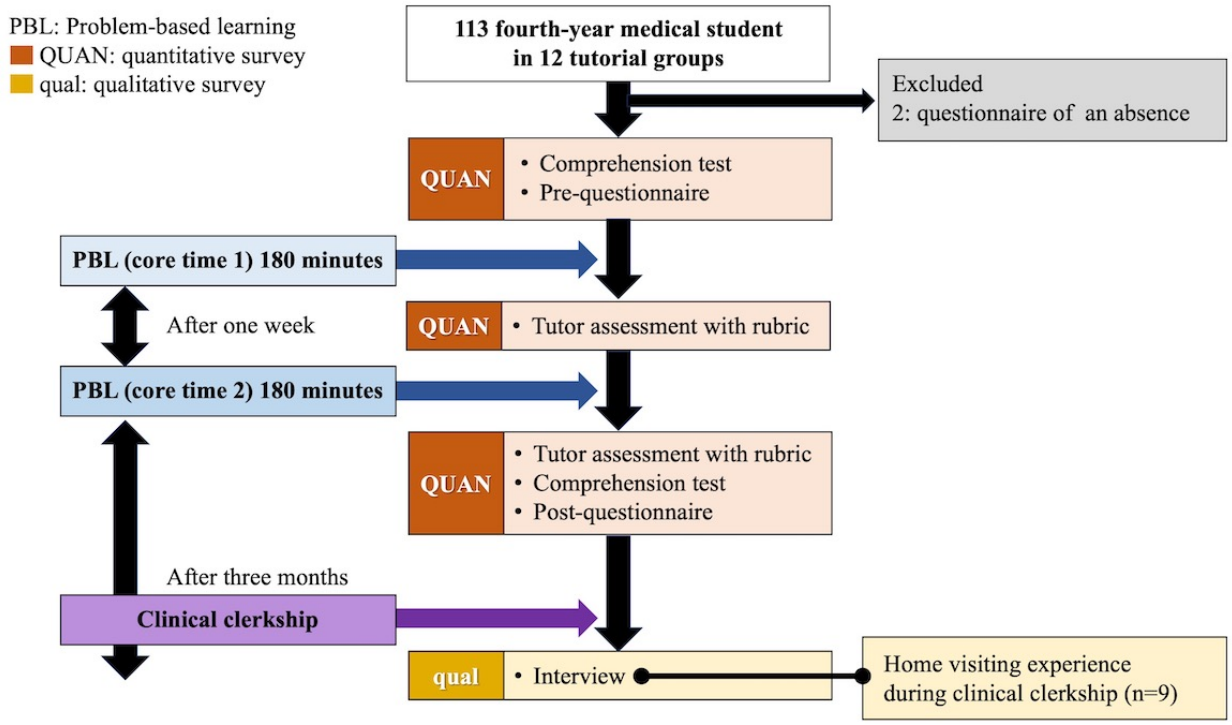
XSL•FO
RenderX

Results

Participant Characteristics

In total, 113 medical students participated in PBL. Of the 113 participants, 71 (62.8%) were male and 42 (37.2%) female. In addition, 2 (1.8%) students were excluded from the survey because they were absent for a core time session; 111 (98.2%) students participated in the quantitative evaluation. The study flowchart is shown in Figure 2.

Figure 2. Study flow diagram.



Quantitative Measures

Comprehension Test

The total comprehension test scores of the students significantly improved after the community-oriented PBL intervention (P<.001, effect size r=0.528). The total pretest scores had a median of 4.0 (IQR 2.5-5.0), whereas the posttest scores improved to a median of 5 (IQR 4-5). Table 1 provides the pre- and posttest analysis results of each question (Q1-Q5). All questions except Q4 showed significant improvements in the percentage of correct answers.

Table 1. Correct answer rates for individual questions in the comprehension test (N=113).

Question number	Pretest correct answers, n (%)	Posttest correct answers, n (%)	*P* value	Effect size (*r*)
1	104 (92.0)	112 (99.1)	.02	2.333
2	69 (61.1)	107 (94.7)	<.001	5.600
3	81 (71.7)	106 (93.8)	<.001	4.640
4	84 (74.3)	93 (82.3)	.19	2.180
5	77 (68.1)	99 (87.6)	<.001	4.310

Rubric-Based Performance Assessment of Community-Oriented PBL Tutorials

Table 2 presents the results of the rubric-based assessment of the students' performance in core time 1 and core time 2. The students' total scores significantly improved from core time 1 (median 38, IQR 33-43) to core time 2 (median 40, IQR 35-44), with P<.001 and an effect size (r) of 0.516. Regarding individual rubric items, significant improvements were observed in domains such as acquiring knowledge applicable to community health care, developing comprehensive care processes, and fostering self-directed learning. In contrast, no significant improvement was observed in the motivation to learn, whereas interpersonal skills showed a small but statistically significant enhancement.

XSL•FO
RenderX

Table 2. Performance assessment of community-oriented PBL[a] tutorials using a rubric.

Performance items	Core time 1, median (IQR)	Core time 2, median (IQR)	*P* value	Effect size (*r*)
Total score (0-50)	38 (33-43)	40 (35-44)	<.001	0.516
Domains				
1. Acquire knowledge that can be easily recalled and applied in community health care settings (0-10).	8 (7-9)	8 (8-8)	<.001	0.494
2. Develop an effective, comprehensive community care process (0-10).	8 (6-8)	8 (8-8)	<.001	0.532
3. Develop self-directed learning methods (0-10).	8 (7-9)	8 (8-8)	<.001	0.553
4. Motivate myself to learn (0-10).	8 (6-8)	8 (6-8)	.11	0.151
5. Acquire good interpersonal skills (0-10).	8 (7-9)	8 (7-9)	.03	0.201

[a]PBL: problem-based learning.

Self-Administered Questionnaire

Results indicated significant changes in students' perceptions of community health care after participating in community-oriented PBL. For the statement "I am interested in community health care," the preintervention median score was 3 (IQR 3-4), which increased to 4 (IQR 3-4) post intervention, with a statistical significance of $P<.001$ ($U=4446.5$). Similarly, the median score for the statement "I can envision a community health care setting" improved from 3 (IQR 3-4) preintervention to 4 (IQR 3-4) postintervention, also showing a significant difference, with $P<.001$ ($U=2589.5$).

Qualitative Measures

Follow-Up Interviews During Clinical Clerkships

We explored community-oriented PBL's impact on medical students' experiences during home visit consultations. In total, 12 (10.6%) medical students who had not experienced home visits before PBL consented to participate in the interview immediately after acquiring home visit experience during a community clinical clerkship. Through qualitative thematic analysis of the interviews, 7 main themes emerged: "building readiness for home care visit participation," "understanding and navigating the home care environment," "professional and personal growth," "interprofessional collaboration and team dynamics," "challenges and opportunities in home care," "community engagement and regional health care systems," and "ethical considerations and end-of-life care" (Table 3).

XSL•FO
RenderX

Table 3. Follow-up interviews and thematic analysis.

Theme	Subthemes
Building readiness for home care visit participation	• Broadened understanding of patient care beyond medical intervention • Enhanced preparedness for real-world clinical situations • Shift in perspective from theoretical knowledge to practical application
Understanding and navigating the home care environment	• Insights into the holistic approach required in home visits • Observations on the complexities of home care, including resource limitations and patient lifestyles • Recognition of the importance of patient and family communication
Professional and personal growth	• Aspiration to contribute meaningfully to patient care • Development of empathy and emotional intelligence • Recognition of the multifaceted role of health care providers in patient support
Interprofessional collaboration and team dynamics	• Importance of teamwork and a multidisciplinary approach in patient care • Learning from and contributing to the health care team • Navigating professional roles and patient relationships
Challenges and opportunities in home care	• Adapting PBL[a] knowledge to address specific patient needs • Confronting and managing unique patient care challenges • Opportunities for innovative care practices in constrained environments
Community engagement and regional health care system	• Enhancing community-oriented medical practice through targeted PBL • Gaining insights into community health care needs and resources • Understanding the impact of regional characteristics on health care delivery
Ethical considerations and end-of-life care	• Deepened understanding of end-of-life care preferences and practices • Navigating ethical dilemmas in patient care decisions • Valuing patient autonomy and quality of life in care planning

[a]PBL: problem-based learning.

Discussion

Principal Findings

This study demonstrated that integrating real-patient videos into community-oriented PBL improves medical students' knowledge, skills, and attitudes toward community health care. Comprehension test results showed significant improvements in students' understanding of core concepts, including community-based integrated care systems (Q1), the ICF framework (Q2 and Q3), and holistic, patient-centered care (Q5). These findings highlight students' enhanced theoretical knowledge essential for community health care practice.

Rubric-based performance assessments revealed notable improvements in 3 key domains:

- Knowledge application (item 1): Students showed improved abilities to recall and apply knowledge in community health care scenarios.
- Developing comprehensive care processes (item 2): Scores reflected stronger skills in designing patient-centered care plans tailored to community settings.
- Self-directed learning (item 3): Students demonstrated enhanced autonomy in planning, monitoring, and reflecting on their learning tasks.

Although interpersonal skills (item 5) improved slightly, no significant changes were observed in the motivation to learn (item 4), indicating areas for potential curriculum enhancement.

Self-assessment questionnaires revealed increased interest in community health care and an improved ability to envision a community health care setting. These results suggest that the intervention can positively influence students' attitudes and readiness for community health care practice, potentially guiding their career interests toward rural areas.

Qualitative analysis of students' reflections underscored themes such as readiness for home care visits, professional and personal growth, and community engagement. Students reported a deeper understanding of the complexities of community health care, fostering empathy and patient-centered approaches essential for effective practice in underserved areas.

Implications of Findings

This study provided valuable insights into how community-oriented PBL, enhanced by real-patient videos, fosters medical students' ability to conceptualize their future professional roles. The qualitative data indicated that students develop a deeper understanding of the principles and complexities of rural care, including holistic approaches, patient-centered decision-making, and the importance of interprofessional collaboration. These findings suggest that the intervention successfully prepares students to engage with the challenges and rewards of rural and community-based practice.

Importantly, the qualitative analysis revealed that many students began to envision themselves as contributors to rural health care systems. Themes such as "professional and personal growth" and "community engagement" highlighted students' recognition of their potential roles in underserved areas. This shift was

XSL•FO
RenderX

supported by their increased interest in community health care, as measured by the "questionnaire for perceptions of community health care self-assessment." Postintervention, students reported greater interest and confidence in envisioning community health care settings.

However, although the data indicated a significant attitudinal shift, we lack sufficient evidence to confirm a direct impact on medical students' long-term career intentions to pursue rural care roles after graduation. Future studies should include longitudinal tracking to assess whether the observed changes in perceptions and interests translate into tangible career decisions. Additionally, research is needed to validate the predictive validity of the self-assessment questionnaire in forecasting students' career trajectories.

This intervention lays a strong foundation for addressing the global challenge of physician maldistribution by bridging theoretical knowledge with practical applications. However, further investigation is required to understand its long-term influence on medical workforce trends and rural health care outcomes.

Comparison With the Literature

Our findings will contribute to a growing body of evidence supporting the efficacy of PBL in medical education. Previous studies have demonstrated PBL's ability to enhance clinical reasoning and decision-making skills [19,21,41]. However, our research added a unique dimension by integrating real-patient videos, which provide authentic learning experiences and contextualize medical knowledge within the framework of community-oriented care. Similar studies have reported that experiential learning approaches, such as case-based learning with audiovisual materials, improve students' engagement and retention of knowledge [41].

The significant gains in self-directed learning observed in this study echo findings from Matsuyama et al [35], who emphasized the role of contextual attributes in promoting self-regulated learning. Additionally, the qualitative themes identified in our analysis, such as the importance of interprofessional collaboration and navigating the home care environment, align with the existing literature on the competencies required for effective community health care practice [42-48].

Limitations

Our study has some limitations. First, it was conducted as part of the curriculum of a single medical school, potentially limiting the findings' applicability to other institutions and geographical settings. Future research should involve multiple institutions to enhance the results' generalizability. Additionally, the demographic and cultural context of the participants may not fully represent broader populations, especially in countries with different health care challenges and educational frameworks. Second, the absence of a control group makes it difficult to attribute observed improvements solely to the intervention of incorporating real-patient videos into PBL. Our study evaluated a "bundle" of educational strategies, including real-patient videos, the PBL framework, faculty interventions, and testing conditions. We could not measure these components' differential effects or potential synergistic interactions. Although a randomized controlled trial (RCT) could offer stronger evidence, implementing RCTs in educational settings poses ethical and logistical challenges, such as withholding valuable learning resources from a control group or ensuring equivalent baseline characteristics. To address these limitations, we recommend that future research consider alternative designs to balance rigor and feasibility. Third, the scalability of real-patient videos poses a significant challenge. Producing high-quality real-patient videos requires substantial time, resources, and collaboration among medical educators, health care professionals, and audiovisual specialists. These demands may limit the feasibility of widespread adoption. We suggest collaborative efforts, such as interschool partnerships and the development of shared digital repositories, to distribute production costs and enhance scalability. Fourth, the mixed methods design relied on self-reported measures and reflections, which could introduce a response bias. Future studies could benefit from incorporating objective measures of clinical performance and patient care outcomes. Fifth, the effectiveness of clinical reasoning education via hybrid PBL may vary depending on instructors' teaching skills. Despite standardized training for tutors, differences in tutor effectiveness may have influenced the consistency of outcomes. We propose further methods to ensure uniformity in future implementations, such as advanced tutor workshops and peer evaluations. Sixth, because the study participants were fourth-year medical students at a single Japanese institution, the results may not be directly generalizable to other populations, such as residents or general physicians, or contexts outside Japan, underscoring the need for further validation. Seventh, the comprehension test used in this study, although developed specifically for the program's educational objectives, was not formally piloted with a separate cohort. Instead, the test underwent cognitive debriefing with faculty members to ensure clarity, relevance, and alignment with the intended learning objectives. Although this process enhanced content validity, the absence of a formal pilot test may limit the ability to fully validate the test's reliability and generalizability. Similarly, the 2 items in the self-assessment questionnaire—designed to evaluate medical students' interest in and ability to envision community health care—were developed based on previous studies and focus group discussions but have not been validated using established scales. This lack of formal validation for the comprehension test and self-assessment questionnaire limits the generalizability and robustness of the findings. Finally, although the qualitative data provided valuable insights into students' conceptualization of professional roles and their preparedness for community health care settings, the lack of longitudinal data limits our ability to assess the long-term impact of these interventions on career trajectories or practice in rural or underserved areas. Future research should include follow-up assessments to evaluate the sustained influence of such educational interventions on students' career decisions and professional development.

Conclusion

Integrating real-patient videos into a community-oriented PBL curriculum shows significant promise in fostering medical students' interest and competencies in community and rural medicine. Our study demonstrated improvements in knowledge

XSL•FO
RenderX

acquisition and application, as indicated by enhanced rubric and comprehension test scores. Moreover, qualitative analysis revealed PBL's effectiveness in developing essential skills and shaping medical students' perceptions toward community health care. Although these changes may not directly translate to career decisions, they represent an essential step toward fostering awareness of rural health care needs and aligning medical students' competencies with the demands of underserved areas. This approach highlights the potential of combining real-patient videos with PBL as an innovative educational strategy to address physician maldistribution and support rural health care systems.

Data Availability

The datasets generated and analyzed during the study are not publicly available, because they include participants' personal data, but they are available from the corresponding author upon reasonable request.

Authors' Contributions

KS, KY, and SI planned, designed, and conceived the study. KS drafted the manuscript. KS, KY, NA, and IS recruited participants. KS, KY, and NA conducted the interviews. KS and NA analyzed the initial coding. KS, KY, IS, and SI analyzed the final coding and interpreted the data. KS and KY performed statistical analyses. HK conceived the figures. KS, KY, NA, IS, TT, HT, LY, and SI participated as tutors. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Community-oriented problem-based learning tutor guide.
[DOCX File , 23 KB - mededu_v11i1e68743_app1.docx]

Multimedia Appendix 2
Information and task sheet.
[DOCX File , 25 KB - mededu_v11i1e68743_app2.docx]

Multimedia Appendix 3
Comprehension test questions.
[DOCX File , 19 KB - mededu_v11i1e68743_app3.docx]

Multimedia Appendix 4
Rubric for community-oriented problem-based learning.
[DOCX File , 25 KB - mededu_v11i1e68743_app4.docx]

Multimedia Appendix 5
Questionnaires for perceptions of community health care self-assessment.
[DOCX File , 22 KB - mededu_v11i1e68743_app5.docx]

References

1. Takayama A, Poudyal H. Incorporating medical supply and demand into the index of physician maldistribution improves the sensitivity to healthcare outcomes. J Clin Med 2021 Dec 28;11(1):155 [FREE Full text] [doi: 10.3390/jcm11010155] [Medline: 35011896]
2. Hara K, Kunisawa S, Sasaki N, Imanaka Y. Examining changes in the equity of physician distribution in Japan: a specialty-specific longitudinal study. BMJ Open 2018 Jan 08;8(1):e018538 [FREE Full text] [doi: 10.1136/bmjopen-2017-018538] [Medline: 29317415]
3. Kobayashi Y, Takaki H. Geographic distribution of physicians in Japan. Lancet 1992 Dec 05;340(8832):1391-1393. [doi: 10.1016/0140-6736(92)92569-2] [Medline: 1360099]
4. Toyabe S. Trend in geographic distribution of physicians in Japan. Int J Equity Health 2009 Mar 03;8(1):5 [FREE Full text] [doi: 10.1186/1475-9276-8-5] [Medline: 19257879]
5. Tanihara S, Kobayashi Y, Une H, Kawachi I. Urbanization and physician maldistribution: a longitudinal study in Japan. BMC Health Serv Res 2011 Oct 08;11(1):260 [FREE Full text] [doi: 10.1186/1472-6963-11-260] [Medline: 21982582]
6. Wu J. Measuring inequalities in the demographical and geographical distribution of physicians in China: generalist versus specialist. Int J Health Plann Manage 2018 Oct 20;33(4):860-879. [doi: 10.1002/hpm.2539] [Medline: 29781216]

XSL•FO
RenderX

7. Karan A, Negandhi H, Nair R, Sharma A, Tiwari R, Zodpey S. Size, composition and distribution of human resource for health in India: new estimates using National Sample Survey and Registry data. BMJ Open 2019 May 27;9(4):e025979 [FREE Full text] [doi: 10.1136/bmjopen-2018-025979] [Medline: 31133622]
8. Morris S, Sutton M, Gravelle H. Inequity and inequality in the use of health care in England: an empirical investigation. Soc Sci Med 2005 Mar;60(6):1251-1266. [doi: 10.1016/j.socscimed.2004.07.016] [Medline: 15626522]
9. Guttmann A, Shipman SA, Lam K, Goodman DC, Stukel TA. Primary care physician supply and children's health care use, access, and outcomes: findings from Canada. Pediatrics 2010 Jun;125(6):1119-1126. [doi: 10.1542/peds.2009-2821] [Medline: 20498170]
10. Mick SS, Lee SD, Wodchis WP. Variations in geographical distribution of foreign and domestically trained physicians in the United States: 'safety nets' or 'surplus exacerbation'? Soc Sci Med 2000 Jan;50(2):185-202. [doi: 10.1016/s0277-9536(99)00183-5] [Medline: 10619689]
11. Winkelmann J, Muench U, Maier CB. Time trends in the regional distribution of physicians, nurses and midwives in Europe. BMC Health Serv Res 2020 Oct 12;20(1):937 [FREE Full text] [doi: 10.1186/s12913-020-05760-y] [Medline: 33046077]
12. Organization for Economic Co-operationDevelopment (OECD). Doctors. Paris: OECD Publishing; 2018.
13. Sato H. Demand, supply and shortages of physicians: a critical analysis on the current government?s method. J Health Welf Policy 2020;3:39-48.
14. Ministry of Health, Labour and Welfare Study Group on Supply and Demand of Medical Staff. Doctor Supply and Demand Subcommittee 4th interim report. Ministry of Health, Labour and Welfare. 2019. URL: https://www.mhlw.go.jp/content/12601000/000504403.pdf [accessed 2024-04-24]
15. Tago M, Shikino K, Hirata R, Watari T, Yamashita S, Tokushima Y, et al. General medicine departments of Japanese universities contribute to medical education in clinical settings: a descriptive questionnaire study. Int J Gen Med 2022;15:5785-5793 [FREE Full text] [doi: 10.2147/IJGM.S366411] [Medline: 35774114]
16. Teo A. The current state of medical education in Japan: a system under reform. Med Educ 2007 Mar;41(3):302-308. [doi: 10.1111/j.1365-2929.2007.02691.x] [Medline: 17316216]
17. Kozu T. Medical education in Japan. Acad Med 2006;81(12):1069-1075. [doi: 10.1097/01.acm.0000246682.45610.dd]
18. Ohde S, Deshpande GA, Takahashi O, Fukui T. Differences in residents' self-reported confidence and case experience between two post-graduate rotation curricula: results of a nationwide survey in Japan. BMC Med Educ 2014 Jul 12;14(1):141 [FREE Full text] [doi: 10.1186/1472-6920-14-141] [Medline: 25016304]
19. Davis MH. AMEE Medical Education Guide No. 15: problem-based learning: a practical guide. Med Teach 1999 Jul 03;21(2):130-140. [doi: 10.1080/01421599979743] [Medline: 21275726]
20. Haith-Cooper M. Problem-based learning within health professional education. What is the role of the lecturer? A review of the literature. Nurse Educ Today 2000 May;20(4):267-272. [doi: 10.1054/nedt.1999.0397] [Medline: 10827097]
21. Forbes H, Syed M, Flanagan O. The role of problem-based learning in preparing medical students to work as community service-oriented primary care physicians: a systematic literature review. Cureus 2023 Sep;15(9):e46074 [FREE Full text] [doi: 10.7759/cureus.46074] [Medline: 37900379]
22. Bhandary S. Problem-based learning curriculum and process assessment system for the undergraduate competency-based medical education: experiences from Nepal. Arch Med Health Sci 2021;9(2):331-336. [doi: 10.4103/amhs.amhs_282_21]
23. Kibret S, Teshome D, Fenta E, Hunie M, Taye MG, Fentie Y, et al. Medical and health science students' perception towards a problem-based learning method: a case of Debre Tabor University. AMEP 2021 Jul;12:781-786. [doi: 10.2147/amep.s316905]
24. Hou S. Integrating problem-based learning with community-engaged learning in teaching program development and implementation. Univ J Educ Res 2014 Jan;2(1):1-9. [doi: 10.13189/ujer.2014.020101]
25. Barbour RS. The case for combining qualitative and quantitative approaches in health services research. J Health Serv Res Policy 1999 Jan 01;4(1):39-43. [doi: 10.1177/135581969900400110]
26. Malterud K. The art and science of clinical knowledge: evidence beyond measures and numbers. Lancet 2001 Aug;358(9279):397-400. [doi: 10.1016/s0140-6736(01)05548-9]
27. Côté L, Turgeon J. Appraising qualitative research articles in medicine and medical education. Med Teach 2005 Jan 03;27(1):71-75. [doi: 10.1080/01421590400016308] [Medline: 16147774]
28. Dowding D. Review of the book best practices for mixed methods research in the health sciences. Qual Soc Work 2013 Jul 16;12(4):541-545. [doi: 10.1177/1473325013493540a]
29. Creswell JW, Plano CV. Designing and Conducting Mixed Method Research. 3rd Edition. Los Angeles: Sage Publications; 2017.
30. Ovitsh RK, Gupta S, Kusnoor A, Jackson JM, Roussel D, Mooney CJ, et al. Minding the gap: towards a shared clinical reasoning lexicon across the pre-clerkship/clerkship transition. Med Educ Online 2024 Feb 06;29(1):2307715. [doi: 10.1080/10872981.2024.2307715]
31. Malterud K. Qualitative research: standards, challenges, and guidelines. Lancet 2001 Aug;358(9280):483-488. [doi: 10.1016/s0140-6736(01)05627-6]
32. Savery JR. Overview of problem-based learning: definitions and distinctions. Interdiscip J Probl Based Learn 2006 May 22;1(1):3. [doi: 10.7771/1541-5015.1002]

XSL•FO
RenderX

33. Houlden RL, Collier CP, Frid PJ, John SL, Pross H. Problems identified by tutors in a hybrid problem-based learning curriculum. Acad Med 2001 Jan;76(1):81. [doi: 10.1097/00001888-200101000-00021] [Medline: 11154202]
34. Wang S, Yan D, Hu X, Liu J, Liu D, Wang J. Comparison of attitudes toward the medical student-led community health education service to support chronic disease self-management among students, faculty and patients. BMC Med Educ 2023 Jan 11;23(1):17 [FREE Full text] [doi: 10.1186/s12909-023-04008-7] [Medline: 36631772]
35. Matsuyama Y, Nakaya M, Okazaki H, Leppink J, van der Vleuten C. Contextual attributes promote or hinder self-regulated learning: a qualitative study contrasting rural physicians with undergraduate learners in Japan. Med Teach 2017 Nov 26;40(3):285-295. [doi: 10.1080/0142159x.2017.1406074]
36. Recker AJ, Sugimoto SF, Halvorson EE, Skelton JA. Knowledge and habits of exercise in medical students. Am J Lifestyle Med 2021 Oct 12;15(3):214-219 [FREE Full text] [doi: 10.1177/1559827620963884] [Medline: 34025308]
37. Trullàs JC, Blay C, Sarri E, Pujol R. Effectiveness of problem-based learning methodology in undergraduate medical education: a scoping review. BMC Med Educ 2022 Feb 17;22(1):104 [FREE Full text] [doi: 10.1186/s12909-022-03154-8] [Medline: 35177063]
38. Joseph N, Rai S, Madi D, Bhat K, Kotian S, Kantharaju S. Problem-based learning as an effective learning tool in community medicine: initiative in a private medical college of a developing country. Indian J Community Med 2016;41(2):133-140 [FREE Full text] [doi: 10.4103/0970-0218.177535] [Medline: 27051088]
39. King N. Using templates in the thematic analysis of text. In: Cassel C, Symon G, editors. Essential Guide to Qualitative Methods in Organizational Research. London, UK: Sage Publications; 2004:256-270.
40. Hennink MM, Kaiser BN, Marconi VC. Code saturation versus meaning saturation: how many interviews are enough? Qual Health Res 2017 Mar 26;27(4):591-608 [FREE Full text] [doi: 10.1177/1049732316665344] [Medline: 27670770]
41. Ikegami A, Ohira Y, Uehara T, Noda K, Suzuki S, Shikino K, et al. Problem-based learning using patient-simulated videos showing daily life for a comprehensive clinical approach. Int J Med Educ 2017 Feb 27;8:70-76 [FREE Full text] [doi: 10.5116/ijme.589f.6ef0] [Medline: 28245193]
42. Du X, Al Khabuli JOS, Ba Hattab RAS, Daud A, Philip NI, Anweigi L, et al. Development of professional identity among dental students - a qualitative study. J Dent Educ 2023 Jan 02;87(1):93-100. [doi: 10.1002/jdd.13092] [Medline: 36052467]
43. Orsmond P, McMillan H, Zvauya R. It's how we practice that matters: professional identity formation and legitimate peripheral participation in medical students: a qualitative study. BMC Med Educ 2022 Feb 09;22(1):91 [FREE Full text] [doi: 10.1186/s12909-022-03107-1] [Medline: 35139839]
44. Wang J, Wang B, Liu D, Zhou Y, Xing X, Wang X, et al. Video feedback combined with peer role-playing: a method to improve the teaching effect of medical undergraduates. BMC Med Educ 2024 Jan 19;24(1):73 [FREE Full text] [doi: 10.1186/s12909-024-05040-x] [Medline: 38243255]
45. Noverati N, R Naro G, J Fischer R, M Thompson B. Using video and virtual patients in problem-based learning: a scoping review. Med Sci Educ 2020 Dec 14;30(4):1685-1691 [FREE Full text] [doi: 10.1007/s40670-020-01108-7] [Medline: 34457832]
46. Berini CR, Bonilha HS, Simpson AN. Impact of community health workers on access to care for rural populations in the United States: a systematic review. J Community Health 2022 Jun 24;47(3):539-553. [doi: 10.1007/s10900-021-01052-6] [Medline: 34817755]
47. Xu Y, Koh XH, Chua YTS, Tan CGI, Aloweni FAB, Yap BEJ, et al. The impact of community nursing program on healthcare utilization: a program evaluation. Geriatr Nurs 2022 Jul;46:69-79. [doi: 10.1016/j.gerinurse.2022.04.024] [Medline: 35609434]
48. Jack HE, Arabadjis SD, Sun L, Sullivan EE, Phillips RS. Impact of community health workers on use of healthcare services in the United States: a systematic review. J Gen Intern Med 2017 Mar 5;32(3):325-344 [FREE Full text] [doi: 10.1007/s11606-016-3922-9] [Medline: 27921257]

Abbreviations

ICF: International Classification of Functioning, Disability, and Health
PBL: problem-based learning
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by B Lesselroth; submitted 12.11.24; peer-reviewed by RA El Arab, M Agbede; comments to author 19.12.24; revised version received 08.01.25; accepted 14.01.25; published 31.01.25.

Please cite as:
Shikino K, Yamauchi K, Araki N, Shimizu I, Kasai H, Tsukamoto T, Tajima H, Li Y, Onodera M, Ito S
Understanding Community Health Care Through Problem-Based Learning With Real-Patient Videos: Single-Arm Pre-Post Mixed Methods Study
JMIR Med Educ 2025;11:e68743
URL: https://mededu.jmir.org/2025/1/e68743
doi:10.2196/68743
PMID:

©Kiyoshi Shikino, Kazuyo Yamauchi, Nobuyuki Araki, Ikuo Shimizu, Hajime Kasai, Tomoko Tsukamoto, Hiroshi Tajima, Yu Li, Misaki Onodera, Shoichi Ito. Originally published in JMIR Medical Education (https://mededu.jmir.org), 31.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Organizational Leaders' Views on Digital Health Competencies in Medical Education: Qualitative Semistructured Interview Study

Humairah Zainal[1*], PhD; Xin Xiao Hui[1*], MSocSci; Julian Thumboo[2,3*], MBBS, MMed; Fong Kok Yong[2,3*], MBBS, MMed

[1]Health Services Research Unit, Singapore General Hospital, Singapore, Singapore
[2]Department of Rheumatology and Immunology, Singapore General Hospital, Singapore, Singapore
[3]Duke-NUS Medical School, National University of Singapore, Singapore, Singapore
[*]all authors contributed equally

Corresponding Author:
Fong Kok Yong, MBBS, MMed
Department of Rheumatology and Immunology
Singapore General Hospital
10 Hospital Boulevard
Singapore, 168582
Singapore
Phone: 65 6908 8949
Email: fong.kok.yong@singhealth.com.sg

Abstract

Background: Digital technologies (DTs) have profoundly impacted health care delivery globally and are increasingly used in clinical practice. Despite this, there is a scarcity of guidelines for implementing training in digital health competencies (DHC) in medical schools, especially for clinical practice. A lack of sustained integration of DHC risks creating knowledge gaps due to a limited understanding of how DT should be used in health care. Furthermore, few studies have explored reasons for this lag, both within and beyond the medical school curriculum. Current frameworks to address these barriers are often specific to individual countries or schools and focus primarily on curriculum design and delivery. A comprehensive framework is therefore required to ensure consistent implementation of DHC across various contexts and times.

Objective: This study aims to use Singapore as a case study and examine the perspectives of doctors in organizational leadership positions to identify and analyze the barriers to DHC implementation in the undergraduate curriculum of Singapore's medical schools. It also seeks to apply the Normalization Process Theory (NPT) to address these barriers and bridge the gap between health care systems and digital health education (DHE) training.

Methods: Individual semistructured interviews were conducted with doctors in executive and organizational leadership roles. Participants were recruited through purposive sampling, and the data were interpreted using qualitative thematic analysis.

Results: A total of 33 doctors participated, 26 of whom are currently in organizational leadership roles and 7 of whom have previously held such positions. A total of 6 barriers were identified: bureaucratic inertia, lack of opportunities to pursue nontraditional career pathways, limited protective mechanisms for experiential learning and experimentation, lack of clear policy guidelines for clinical practice, insufficient integration between medical school education and clinical experience, and poor IT integration within the health care industry.

Conclusions: These barriers are also present in other high-income countries experiencing health care digitalization, highlighting the need for a theoretical framework that broadens the generalizability of existing recommendations. Applying the NPT underscores the importance of addressing these barriers to effectively integrate DHC into the curriculum. The active involvement of multiple stakeholders and the incorporation of continuous feedback mechanisms are essential. Our proposed framework provides concrete, evidence-based, and step-by-step recommendations for implementation practice, supporting the introduction of DHC in undergraduate medical education.

(JMIR Med Educ 2025;11:e64768) doi:10.2196/64768

XSL•FO
RenderX

KEYWORDS

technology; medical education; curriculum; clinical competence; digital competence; Singapore; digital health; qualitative study; medical school; risk; comprehensive framework; doctor; thematic analysis; information technology; evidence-based; undergraduate; healthcare systems; mobile phone

Introduction

Background

The integration of digital technologies (DTs) into clinical care is transforming health care worldwide [1], underscoring the need to prepare future health care professionals with digital health competencies (DHC) through digital health education (DHE). Despite widespread recognition of the importance of DHC, medical schools worldwide—including those in Singapore—have been lagging in their efforts to implement such training in a meaningful and systematic manner [2-8]. Countries like the United States, the United Kingdom, Canada, and Germany face similar challenges, such as fragmented integration efforts, limited faculty expertise, and curriculum overload, which hinder the consistent incorporation of DHE into undergraduate medical curricula [9-14]. Singapore, a high-income nation in Southeast Asia with advanced education systems and extensive digitalization, provides a compelling case study to explore these barriers. While Singapore's unique sociopolitical and cultural context informs this study, the challenges it faces mirror those encountered by other high-income nations, highlighting the broader international relevance of this research.

Existing efforts to integrate DHE in medical schools, especially for clinical practice, are often disconnected, lacking systematic frameworks and sustained engagement with key stakeholders [2-8]. For example, in Europe and the United States, DHE initiatives are often siloed, leading to significant variability in the quality and scope of training [2,4,6]. Similarly, Australia has faced barriers such as a lack of standardized frameworks for digital health training and challenges in aligning medical education with rapidly evolving health care technologies [3].

These gaps result in inconsistencies in training, with DHC frequently treated as elective content rather than as a core component of medical education. This study applies Normalization Process Theory (NPT)—a framework designed to examine how new practices become embedded within institutions—to provide insights into systematically normalizing DHE and ensuring its sustainable integration [15].

This study addresses the following research questions: (1) What are the institutional and structural barriers to integrating DHE into undergraduate medical curricula? (2) How can the medical school experience be aligned with technological advances? (3) How can NPT be systematically applied to facilitate the effective and sustainable incorporation of DHE? By focusing on Singapore, this study not only provides a deeper understanding of these challenges but also offers insights that can inform global efforts to strengthen DHE integration in medical education.

Exploring the Perspectives of Doctors in Organizational Leadership Roles

Despite their influence on governance and standards, the perspectives of doctors in organizational leadership roles are often overlooked. Existing research that evaluates the opinions of this group of stakeholders primarily addresses challenges in implementing DT in health care, characteristics of effective health systems, and key attributes for health care leaders [16-18].

This study recognizes that doctors in organizational leadership roles possess a strategic understanding of both clinical practice and medical education. Their insights are crucial as they can influence curriculum design, resource allocation, and policy formulation. By leveraging their dual perspectives, the study identifies unique barriers that may not be visible to frontline educators or students.

Furthermore, leaders in health care organizations have the authority to implement change and drive initiatives. Understanding their perspectives ensures that any proposed solutions are both feasible and likely to gain support at the highest levels of the institution. Their endorsement can facilitate smoother implementation and wider acceptance of DHE initiatives.

In addition, these leaders are often involved in broader system-wide decision-making. They are also likely familiar with challenges related to integrating new technologies and practices into established systems. Hence, their experience can provide valuable insights into overcoming any institutional inertia, aligning new initiatives with existing policies, and addressing systemic barriers.

Medical Education in Singapore

Medical education in Singapore is provided by 3 schools, which are Yong Loo Lin School of Medicine (YLL) at the National University of Singapore (NUS), Lee Kong Chian School of Medicine (LKCMedicine) at the Nanyang Technological University (NTU), and Duke-NUS Medical School (Duke-NUS). YLL, established in 1905, and LKCMedicine, founded in 2013, both provide a 5-year undergraduate program. In this program, students spend 2 years studying the basics of medical sciences before undergoing clinical clerkships from the third to the fifth year. YLL was formed to address the critical health care needs of the local population during the colonial period, whereas LKCMedicine was established to meet the increasing health care demands due to an aging population [19]. To boost Singapore's capabilities in translational medicine, Duke-NUS was founded in 2005 through a partnership between NUS and Duke University in the United States. Duke-NUS is a graduate medical school that offers a 4-year MD program, where the first year focuses on basic sciences and the second year on clinical postings. In their third year, students focus on developing research skills, and in their final year, they engage in clinical clerkships [20]. All medical schools in Singapore receive public

XSL•FO
RenderX

funding, and students' tuition fees are subsidized by the government.

Despite their cutting-edge facilities and innovation-driven educational technology, disparities persist between medical school training and clinical application [8,21]. Efforts to integrate DHE courses, such as virtual reality and point-of-care ultrasound, also vary in content and duration across these institutions [8]. Hence, standardizing the curriculum and ensuring consistent training across all medical schools is crucial to bridging the gap between theoretical knowledge and practical skills. This approach would enhance the overall competency of future health care professionals and improve the quality of patient care. In addition, ongoing assessment and adaptation of these programs to incorporate emerging technologies and methodologies are essential for keeping pace with the rapidly evolving medical field.

Methods

Data Collection

A qualitative study was conducted using individual semistructured interviews with doctors who are currently or have previously held organizational leadership positions. Participants were identified by our principal investigator (PI), FKY, based on their leadership roles within public health care organizations, ensuring they possessed the requisite knowledge and experience aligned with the research objectives. Selection criteria focused on senior leaders with expertise in research, clinical education, and development.

This cohort represented a niche group of chief physicians leading public tertiary hospitals that serve as teaching hospitals for undergraduate and postgraduate medical training. Specifically, participants included group chief executive officers from all 3 public health care clusters in Singapore, chairmen of medical boards at public health care institutions, senior administrators from the Ministry of Health, and Directors of Training and Education. To ensure a consistent depth of expertise, participants were required to have a minimum of 5 years of leadership experience within public health care organizations. Those with less than 5 years of such experience were excluded from the study.

Purposive sampling was used to ensure a diverse representation of organizational leaders based on factors such as organizational type, that is, public health care clusters and institutions and functional domains, ie, clinical services and administration. This approach enabled the collection of rich, varied perspectives, enhancing the study's credibility.

Data collection took place from January to April 2021. Participants were invited by the PI by email, which provided a detailed outline of the study's purpose, procedures, potential risks, and benefits. The email also included a consent statement for participants to review and acknowledge before proceeding. In reviewing issues of reflexivity, the threat of potential researcher biases due to the established professional relationship between the PI and the research participants was overcome by having the research fellow, who had no previous relationship with any of the participants, as the interviewer. During the Zoom (Zoom Video Communications) interview, participants were given the opportunity to ask questions, and their verbal consent was recorded at the beginning of each session to ensure informed and voluntary participation. They were reminded of their right to withdraw from the study at any point. It was clarified that data collected prior to withdrawal would still be retained and analyzed to enable a comprehensive evaluation of findings.

To protect participants' anonymity, we assigned code identifiers beginning with "OL" (organizational leader) to each of them. Any identifying information and audio recordings were stored separately from the main dataset in a secure, password-protected file, accessible only to authorized research team members. In reporting results, care was taken to remove or generalize any details that could potentially identify individuals. The data collected and analyzed was used exclusively to inform curriculum development, with no intention to disclose identifiable information.

The interview guide was developed based on the NPT constructs of coherence, cognitive participation, collective action, and reflexive monitoring (Textbox 1). We then adapted the interview guide iteratively to allow participants to share their views on matters that were not initially included in the guide. Generally, the questions sought participants' views on the clinical skills that are still relevant in the digital age (coherence), additional skills that medical students and doctors need for clinical practice amid increasing health care digitalization (coherence), and the clinical skills that are currently being covered in local medical schools (coherence). We also asked participants for their opinions on the clinical skills that should be emphasized more in the medical school curricula (coherence), the challenges of integrating DHE into the compulsory curricula, and suggestions for curriculum improvement to better prepare students for future clinical practice (cognitive participation). To explore how "Collective Action" could be operationalized, we asked how medical schools could improve their collaboration with other stakeholders, particularly professional bodies, health care institutions, and the health care system, to better prepare medical students for clinical practice in the digital era (collective action). We also raised the question of how participants would evaluate the impacts, benefits, and areas for improvement of DHE initiatives within the medical school curriculum (reflective monitoring).

XSL•FO
RenderX

Textbox 1. Interview questions.

- In general, what are the clinical skills that a medical doctor should have?
- Which of these skills are still relevant in the digital age?
- Are there any skills that have been replaced by digital technology, be it partially or completely?
- Against the backdrop of increasing digitalization of health care, what new skills, clinical or otherwise, should a doctor have in order to practice medicine?
- What clinical skills are currently being covered in the local medical schools?
- Which of these skills should be emphasized more in the medical school curriculum?
- In your opinion, how well do the current medical school curricula prepare medical students for the digital aspects of health care? What do you think are some of the challenges in implementing digital health education in medical schools?
- What other improvements can be made to our local medical school curriculum to better prepare the students for clinical practice in light of rapid advances in technology (for example, the advent of artificial intelligence, big data, imaging, smartphone applications, and digital equipment such as handheld ultrasound)?
- How can local medical schools improve their collaborations with professional bodies and health care institutions to prepare medical students for clinical practice in this era of new technology?
- What can the health care system do to support medical students and young doctors in this era of new technology?
- Do you have any other comments on the digital transformations of medicine or health care before we end this interview?

Challenges of contextual differences and stakeholder variation are crucial factors that need to be carefully considered when applying NPT in diverse settings. The study was conducted in Singapore, where the adoption of DT within health care settings has been gradual [7]. This presents challenges for students who may not have adequate exposure to digital systems during their clinical placements. In response to this, the application of NPT should be focused on building digital literacy and ensuring that any intervention is compatible with ongoing efforts to integrate digital solutions into clinical practice. Furthermore, there is also a limited innovation culture in Singapore's health care system [7]. To overcome this, interventions that adopt NPT should incorporate elements designed to stimulate collaboration and mentorship programs with industry professionals. This would help bridge the gap between academic training and the innovation needs of the health care sector.

The study included 33 participants, with the sample size determined based on theoretical and practical considerations. Data collection continued until saturation was reached, ensuring that no new themes or insights emerged from the interviews. This indicates that the sample size was sufficient to capture the relevant perspectives for the study. While practical constraints, such as time and resources, influenced the final number of participants, the primary focus was on ensuring data richness and diversity. This approach allowed for a comprehensive exploration of the research questions.

A total of 30 interviews were conducted and recorded over Zoom due to the physical restrictions brought about by the COVID-19 pandemic, while 3 in-person interviews were held with participants who were located in areas with fewer restrictions at the time or who specifically preferred in-person interaction. All in-person interviews were carried out in accordance with local health guidelines to ensure participant safety. Each interview lasted approximately 40 minutes and was audio-recorded. The transcriptions were derived from the audio recordings of the interviews, which were processed using Otter.ai software (Otter.ai, Inc) before being reviewed for accuracy by the PI and research fellow.

Data Analysis

Thematic analysis using Braun and Clarke's [22] 6-step framework was used to explore barriers that emerged from the data, while a deductive approach based on the constructs of NPT was used to map suggestions for curricula improvement to relevant NPT constructs. To overcome potential interpretive bias and selective perception, coding was conducted by 2 researchers independently. After the initial coding, discrepancies were discussed, and a consensus was reached to refine the codebook and ensure consistency in the application of codes. To enhance credibility and trustworthiness, data were triangulated by comparing the findings across participants from various public health care clusters to identify any consistencies and divergences in opinions. This helped to ensure that the themes captured diverse perspectives and were not unduly influenced by any single group.

In addition, we contextualized the findings by examining studies from other high-income countries undergoing similar digital transformations in health care. Furthermore, we analyzed recently published data reflecting the perspectives of other stakeholders in the health care industry, such as clinical educators and leaders of medical schools, regarding the digital competencies required for future clinical practice [8,21]. In the reporting of findings, we followed the Standards for Reporting Qualitative Research of O'Brien et al [23].

Ethical Considerations

This study was classified as a quality improvement (QI) project focusing on medical education curricula by the Research Integrity, Compliance, and Ethics (RICE) committee of SingHealth. In line with institutional guidelines, QI projects aimed at enhancing existing practices, processes, or programs, such as curriculum development in medical education, do not meet the criteria for human subjects research. As such, the study

XSL•FO
RenderX

was granted an ethical waiver by the SingHealth Centralized Institutional Review Board (2020/2880). This decision was based on the determination that the activities involved posed no more than minimal risk to participants. Despite this waiver, the research adhered strictly to the ethical principles outlined in the World Medical Association's Declaration of Helsinki and institutional guidelines.

Results

A total of 33 participants took part in the study. They included 19 chief medical officers from local public health care institutions, 3 chief executive officers from public health care clusters, 4 senior administrators, and 7 former organizational leaders. Each had at least 5 years of organizational leadership experience and represented various specialties (Table 1).

Participants shared that local medical schools have not yet revamped the curricula to incorporate relevant competencies for the digital age. They identified 6 reasons for the lag in DHC training, some of which extended beyond the medical schools. The analysis of codes, along with the generation of subthemes and themes, is summarized in Table 2. Illustrative quotes from the interviews are provided below.

Table 1. Demographics of participants (N=33).

Characteristics	Participants
Age (years)	
Mean	62
Median	60
Minimum age	44
Maximum age	82
Gender, n (%)	
Male	31 (94)
Female	2 (6)
Years in organizational leadership	
Mean	18.7
Median	18
Discipline, n (%)	
Gastroenterology and hepatology	5 (15.2)
Pediatrics (including pediatrics genetics, pediatric emergency medicine, and pediatric gastroenterology)	4 (12)
General surgery	3 (9.1)
Psychiatry	3 (9.1)
Renal medicine	3 (9.1)
Anesthesiology	2 (6.1)
Geriatric medicine	2 (6.1)
Respiratory medicine	2 (6.1)
Cardiology	2 (6.1)
Orthopedic surgery	2 (6.1)
General medicine	1 (3)
Medical oncology	1 (3)
Ophthalmology	1 (3)
Surgery and urology	1 (3)
Hand and reconstructive microsurgery	1 (3)

XSL•FO
RenderX

Table 2. Codes, subthemes, and themes identified from the coding process.

Codes	Subthemes	Themes
• Lack of time. • Hard to change. • Resistance. • Not open to new technologies. • Not willing to try new technologies. • Academics have to be open.	• Packed curriculum. • Preference for status quo. • Traditional mindset of senior clinicians and faculty.	• Bureaucratic inertia.
• Lack of alternative career pathways. • Lack of role models. • Mindset changes needed.	• Expectations for graduates to become doctors with patient-fronting roles.	• Limited opportunities to pursue traditional career pathways.
• Safe. • Safe sandbox. • Safety nets. • Patient safety. • Safe and creative space. • Nurture and protect. • Talk about the pitfalls and dangers of using technology.	• Lack of safety mechanisms to use DT[a] for educational purposes. • Limited opportunities to experiment with new technologies due to lack of creative space.	• Lack of protective mechanisms for experiential learning and experimentation.
• Clear guidelines. • Clear policies. • Clear intent. • Clear boundaries. • Help students navigate data, fake news, and misinformation. • Data abuse. • Medical ethics. • Respect privacy. • Ethical competency. • Schools presume these (ethical competencies) are common sense.	• Gaps in outlining guidelines and boundaries for technology use. • Gaps in teaching students the pitfalls of using technologies for clinical practice. • Gaps in equipping students with skills in handling data, medical information, and patients' privacy.	• Lack of clear policies and guidelines for clinical practice.
• Interface. • Incorporate teaching facilities within health care institutions. • Correlate. • String information.	• Limited integration of educational and research facilities for medical students within clinical settings. • Lack of feedback on students' performance outcomes. • Lack of compatible data encountered in medical school and residency.	• Lack of integration between medical school education and experience in the health care system.
• Gap between IT and health care. • Nonintegration. • Disorganized. • Slave to the system. • Need to redesign the system. • Put up robust systems. • Involve IT experts. • Facilitating platforms. • Support end users. • Internet separation.	• Health care industry should drive the IT industry.	• Lack of IT integration within the health care industry.

[a]DT: digital technology.

Bureaucratic Inertia

Participants suggested that bureaucratic inertia within both the health care system and medical schools contributed to sporadic and limited training in DT. They attributed this inertia to faculty members' lack of awareness regarding the evolution of clinical practice, their limited expertise in DT, and their resistance to incorporating new competencies, which would require sacrificing some traditional areas of expertise. As shared by OL8 and OL26:

There are senior clinicians who may not be so open to using DT. They are not willing to use different methodologies to solve the same problem. [OL8, Internal medicine, and Respiratory and Critical Care Medicine]

I tried to teach ultrasound in a medical school but with limited success… Unfortunately, it was met with great resistance from people who are traditional. [OL26, Cardiology]

XSL•FO
RenderX

Furthermore, participants perceived that policy makers and senior clinicians were hesitant to invest in DT due to concerns over higher health care costs, further hindering efforts to optimize DT in clinical settings. This perspective is illustrated by the following comment:

> *Some new technologies are almost invariably more expensive and will increase the cost of care.* [OL4, General Surgery]

The above excerpts highlighted systemic barriers to the integration of DT in clinical practice and medical education, emphasizing how institutional inertia and hesitation to invest in new technologies are contributing to the stagnation in clinical training and practice. The reluctance of policy makers and leaders to embrace change and allocate resources for DT exacerbates these challenges, ultimately hindering the evolution of medical education.

Lack of Opportunities to Pursue Nontraditional Career Pathways

Participants also identified limited opportunities to pursue alternative career paths and nonclinical roles, as well as the absence of role models in new technology fields, as significant barriers to implementation. As opined by OL26:

> *I've seen promising students and residents fall through the cracks and give up along the way because we don't have enough career pathways and role models for those in the medical innovation track.* [OL26, Cardiology]

In addition, OL8 highlighted the stigma within the medical community, where students who left medical school to explore nontraditional pathways were often perceived as failures.

> *We lack the definition of what kind of medical graduates we want to train. Other than basic clinical knowledge, I don't think we have defined anything further than that, like a clinician with knowledge of innovation. If a student decides to be an entrepreneur, for example, create a new start-up and drop out of medical school, we should still take that as a success and not a failure.* [OL8, Internal Medicine and Respiratory and Critical Care Medicine]

Without embracing alternative career paths and addressing the stigma associated with leaving traditional medical roles, the health care system risks alienating promising talent and limiting progress in medical innovation. Establishing clear pathways and celebrating diverse career outcomes is essential to cultivating a dynamic and adaptable health care profession.

Lack of Protective Mechanisms for Experiential Learning and Experimentation

In addition, participants noted limited protective mechanisms for experiential learning and experimentation in the health care system. The lack of a "safe and creative space" hindered trainees from engaging in innovative and secure experimentation with DT. Some participants proposed establishing sandboxes where trainees could test ideas with safeguards in place. This would enable them to contribute to clinical practice improvements while receiving proper guidance when mistakes occur. As articulated by OL8 and OL9:

> *The senior clinicians may not be so open to new things. As health care leaders ourselves, we need to embrace the idea of creating a safe sandbox where students [are] allowed to use their imagination to innovate, with all the safety nets in check for patient safety.* [OL8, Internal Medicine, and Respiratory and Critical Care Medicine]

> *What's lacking is a safe space for students and residents. A safe space is a space that offers professional, psychological, and personal safety for them. Measures need to be taken to train, nurture, and protect them rather than condemn them when they do something wrong. The health care system should give them that safe and creative space that ensures they are not bullied, harassed, and ridiculed.* [OL9, Anesthesiology]

Without the establishment of structured and supportive environments for experiential learning, the health care system risks stifling innovation and deterring the next generation of clinicians from engaging with DT. Proactively establishing protected and guided learning environments is essential for fostering a culture of experimentation and ensuring meaningful contributions to clinical advancements.

Lack of Clear Policies to Guide DT Integration in Clinical Practice

Another significant barrier articulated by participants was the lack of clear policies to guide the effective integration of DT in clinical practice. They emphasized the need for well-defined guidelines at both institutional and ministerial levels to support the ethical and professional use of DT. As noted by OL11:

> *The policies that govern digital technologies like telemedicine must be reasonable. Currently, the intent is unclear. At the institutional and ministerial level, there must be clear guidelines and policies that outline the learning and growth in the use of these technologies.* [OL11, Geriatric Medicine]

The lack of comprehensive policies limits awareness of the risks, pitfalls, and ethical considerations associated with DT, deterring its use, particularly among students. OL25 elaborated on the importance of training students in ethics and professionalism to prevent potential misuse of data.

> *In the world of AI and digital medicine, the role of ethics and professionalism are going to be even more important because it opens up easy channels to data abuse, and doctors will have so much data in their hands. So, you need to teach the students medical ethics and values related to patient information and treatment prescription. It's going to be so critical you need to enforce that.* [OL25, Medical Oncology]

The absence of clear, comprehensive policies to govern the use of DT in health care creates ethical and professional ambiguities, deterring adoption and proper training. Establishing well-defined guidelines is critical to mitigating risks, ensuring ethical use,

XSL•FO
RenderX

and preparing future clinicians to navigate the complexities of digital medicine responsibly.

Lack of Integration Between Medical School Education and Clinical Experience

Participants shared that the perceived lack of integration between medical school education and students' clinical experience in the health care system is another barrier to DHE. They attributed this gap to the lack of systems interoperability, which prevents students from accessing and using health care data used in clinical settings and receiving feedback from these systems. As one participant explained, a more integrated system would allow student performance data to correlate with hospital data, enabling continuous feedback and supporting learners' improvement:

> *The biggest gap is that we don't know how students are performing. The data that students are trained for should be similar to the place of practice. If the system is built such that medical school data correlates with say, hospital data, I can string all the information about your learning journey and see how that impacts your performance outcome. From that perspective, we can support the learners better because we give them an environment where they are constantly receiving feedback from the system and seeking new ways to improve themselves. I think that will probably be the most meaningful thing for our learners.* [OL15, Psychiatry]

More broadly, participants noted that the lack of integration between educational and research facilities within health care institutions limits students' clinical immersion. According to OL16, closer collaboration between medical schools and health care institutions is essential for strengthening this connection and enhancing experience, not just physically but through more active interaction between the institutions and health care professionals.

> *I think medical schools should be in the health care institutions. They should interface very closely. One way is to incorporate teaching and research facilities within health care institutions so that the immersion is useful. Currently, our medical schools are within the proximity of the hospital campus. It makes sense, but that's just the physical infrastructure. The people need to be interfaced quite a fair bit.* [OL16, Psychiatry]

The fragmented nature of medical education and clinical training suggests that a more integrated approach, leveraging data-driven feedback mechanisms and collaborative partnerships between academic and health care institutions, is necessary to foster a culture of continuous learning and improvement in health care.

Lack of IT Integration Within the Health Care Industry

Participants also suggested that an integrated IT infrastructure in health care institutions would increase DHE effectiveness and enhance clinical care. However, they highlighted the current lack of interoperability between systems, which hinders the optimization of technical needs. A recurring concern was the IT sector's lack of ability to understand and address the specific needs of health care, with participants noting disorganization and a disconnect between IT and health care practices. As one participant expressed:

> *The gap between the IT and health care industry has not been bridged yet. We have a lot of IT in the health care industry, but a lot of it is record-keeping. It does not integrate [and] information is coming from every direction that is totally disorganized. How, then, can we teach our medical students to be responsible for the patient as a whole? Somebody who has the ability to do IT programming has to follow the doctors on their rounds. I've ever asked my IT colleagues, "Look, is this an IT industry or a health care industry? When they said it's a health care industry, I said, okay, then you have to listen to me and make things work for me, not enslave me to your products."* [OL9, Anesthesiology]

Furthermore, participants emphasized the challenge of internet separation and the need for platforms that allow seamless cross-sharing of information, which they identified as crucial for effective learning environments. As shared by OL12:

> *One of the biggest challenges is Internet separation…The availability and cross-sharing of information are all important facilitating platforms that we have to provide for medical students.* [OL12, Pediatrics]

The lack of integrated IT infrastructure and disjointed systems within health care settings creates significant barriers to enhancing DHE and clinical care, often leaving medical practitioners frustrated with ineffective solutions. To bridge the gap between IT and health care, a more tailored approach is needed, where technological systems are designed to directly support clinical workflows, ensuring both efficiency and improved educational outcomes.

Discussion

Principal Findings

By interviewing doctors in organizational leadership, we gained insider perspectives on gaps in both the medical curricula and the health care system. A total of 6 barriers were identified: bureaucratic inertia, lack of opportunities to pursue nontraditional career paths, limited protective mechanisms for experiential learning, unclear policy guidelines, limited integration between education and clinical experience, and IT integration issues. The findings contributed to the existing literature by showing that DHE barriers were not limited to medical school curricula but involved broader systemic issues. Comprehensive strategies were needed to address these challenges.

By using qualitative interviews, our study uncovered nuances in leadership decision-making that are often missed in quantitative surveys, providing a richer understanding of the factors influencing leadership perspectives. While most studies suggest that organizational leaders prioritize efficiency and sustainability [16-18], our findings reveal that leaders in this

XSL•FO
RenderX

context place a higher emphasis on experimentation and innovation, a factor not traditionally associated with corporate leadership. Furthermore, our research also highlights the growing influence of digital transformation on leadership styles, an area that received limited attention in previous studies focused on traditional management structures. It underscores the importance of adaptive leadership in an era of constant change, suggesting the need for leadership training programs that focus on flexibility.

Many of the barriers identified in this study align with findings from other high-income countries. These include the lack of the necessary information and communication technology (ICT) skills and limited awareness of the potential benefits of DT among some clinicians. For example, in Germany, an empirical study by Ernstmann et al [24] revealed that some primary care doctors perceived eHealth cards as less useful due to their limited ICT expertise and lack of involvement in technological development. These eHealth cards, which store medical data, treatment plans, medications, and electronic patient files, rely on a telematics infrastructure for communication [24]. The study recognized that without robust IT support, comprehensive training for medical professionals, and a standardized national implementation procedure, the acceptance, adoption, and sustained use of eHealth technology by doctors are likely to be hindered [25].

In addition, other studies have shown an increasing proportion of medical school graduates pursuing careers outside full-time clinical practice in some countries [26]. However, findings from countries such as the United States and South Korea indicate that medical school curricula often fail to adequately address the need for programs providing information on nontraditional careers or nonclinical career pathways [27,28]. Despite expressed interest in these career options, medical students often lack awareness of available training opportunities. To attract students to such careers, early outreach programs, combined with appropriate indemnity and support for innovative projects, are essential. These initiatives could be implemented through elective classes, incentives from professional societies, or partnerships with experts [27].

Furthermore, research from countries such as Canada and Taiwan highlights how technological tools can be leveraged to foster experiential learning among medical students. At the University of Ottawa, social accountability experiential logs were developed for third-year medical students to address the social determinants of health, which are often overlooked in clinical learning objectives [29]. These logs guided students in reflecting on clinical encounters and targeting psychosocial skill development, improving clinical confidence, and demonstrating adaptability for other medical schools (Fung et al [29]). Similarly, a Taiwanese study by Liao et al [30] showcased how the mPath (KU Leuven) e-learning tool supported communication skills training by providing a flexible, technology-enhanced learning environment [30]. Features such as remote accessibility, session recordings, peer feedback mechanisms, and visualized analytical reports enabled learners to engage in self-reflection, adapt communication strategies, and enhance subverbal communication skills [30]. Together, these initiatives exemplify how experiential learning tools can address both biomedical and psychosocial challenges in medical education.

The lack of clear laws and policies to guide DT integration in clinical practice is also a barrier in other high-income countries. For instance, health care leaders in Sweden have acknowledged the need for updated policies [16]. They noted that existing laws and regulations have not kept pace with rapid technological advancements and the evolving organization of health care. These policies require revision to ensure clarity regarding liability and accountability, particularly in addressing how errors are managed when artificial intelligence (AI) systems play a role in clinical decision-making [16].

Furthermore, the limited integration between medical education and clinical experience has been highlighted in various studies and reviews. For instance, Pereira et al [31] describe the implementation of a single competency-based Epic onboarding process for medical students in certain US medical schools with rotations across multisystem training sites. This initiative has enabled learners to spend more time in clinical settings with optimized access to electronic health records (EHRs) [31]. While this approach reduces the training burden, curricula could be further enhanced by emphasizing the practical application of EHRs in clinical settings. This includes training students to maintain professionalism and establish rapport with patients while using EHR systems [31]. In addition, Chan and Zary [32] emphasize that providing immediate and formative feedback on students' performance can support the effective use of AI in medical education. However, delivering high-quality feedback in clinical contexts remains a challenge, as it depends on the underlying knowledge base and model of the AI system, which still requires refinement [32].

Previous systematic reviews have consistently identified infrastructure and technical barriers as the most frequently cited barriers to technology integration in health care [33]. These challenges include limitations in health care capacity for technology adoption, inadequate interconnectedness, insufficient network resources, and incompatibility with existing daily workflows [33]. Addressing these barriers requires the active involvement of health care professionals in the development and implementation of health technology tools, which can also enhance their capacity to effectively manage such applications. Furthermore, the reviews emphasize the critical importance of user engagement and collaboration with system developers throughout all phases of design, development, deployment, and continued use [33]. This collaborative approach ensures that the applications are fit for purpose, as they are designed to align with and address health care providers' needs and expectations.

Our findings highlighted structural and bureaucratic barriers beyond medical schools that hindered DHE implementation. Although they are common in high-income countries, no comprehensive framework has been proposed to address them to date. This study applies May and Finch's [15] NPT to suggest ways to bridge these gaps. A summary of how the 4 constructs of NPT can be applied to each of these barriers is found in Table 3.

XSL•FO
RenderX

Table 3. Addressing each identified barrier with the Normalization Process Theory (NPT).

Barriers	NPT contributions
Bureaucratic inertia	• Coherence: enhance understanding and sense-making among stakeholders about the importance and benefits of DHC[a]. This could be achieved by hiring prospective faculty with the skill sets that are relevant to the needs of up-and-coming developments in medicine. • Cognitive participation: engage key stakeholders to foster buy-in and commitment. For example, leaders of medical schools can engage individuals with influence to encourage the integration of digitalization in the core curriculum. These include engaging clinical educators, teachers, and innovators trained in DT[b] in knowledge exchange and talking with the faculty to facilitate the training of DHC and keep them abreast of the latest technological developments in clinical settings. • Collective action: develop strategies to streamline decision-making processes and reduce red tape. • Reflexive monitoring: continuously evaluate and adjust strategies to address bureaucratic resistance and demonstrate early successes to build momentum.
Lack of opportunities to pursue nontraditional career pathways	• Coherence: clarify the relevance of DHC to future career opportunities and the evolving landscape of health care. • Cognitive participation: involve influential faculty and practitioners, such as medical innovators, in promoting the value of alternative career pathways. Medical schools should also provide sufficient training and mentoring opportunities for students who wish to pursue alternative career pathways. • Sufficient resources should be invested in implementing a curriculum that provides students with opportunities to diversify their skill sets, such as skills in clinical informatics relevant to clinical practice. It should include collaborative mentorship where students can explore new fields in DT by forming partnerships with experts from both the clinical and nonclinical fields. • Collective action: integrate DHC into career development programs and highlight role models who have successfully incorporated digital skills. • Reflexive monitoring: gather feedback from students and professionals to continually refine the approach and address concerns about career impact. Relevant recognition should also be given to medical graduates who embark on alternative pathways to encourage the growth of the fields and normalize these pathways for them.
Lack of protective mechanisms for experiential learning and experimentation	• Coherence: emphasize the importance of experiential learning for mastering DHC. • Cognitive participation: foster a culture of experimentation and learning by involving faculty in the design and delivery of experiential learning opportunities. • Collective action: develop and implement policies and resources that support protected time and space for experiential learning and innovation. These include creating more sandboxes and expanding reasonable access to EHRs in clinical settings. • Reflexive monitoring: continuously assess and improve experiential learning programs based on feedback and outcomes.
Lack of clear policy guidelines for clinical practice	• Coherence: clearly articulate the need for and benefits of standardized DHC policies. • Cognitive participation: engage policy makers, clinical leaders, and educators in developing and endorsing clear guidelines. To ensure that the threat of litigation does not hinder technological adoption, professional bodies should establish clear policies that regulate the effective implementation of DT in clinical settings. A technology assessment committee could also be set up to develop guidelines that enable young trainees to use DT effectively and ethically, both for their safety as well as for their patients. • Collective action: implement training and support systems to ensure consistent application of policies across clinical settings. Professional bodies should also work with schools to equip students with knowledge of cybersecurity as well as the limitations and pitfalls of using DT in various circumstances. • Reflexive monitoring: regularly review and update policies based on clinical practice feedback and emerging best practices. Dedicating time to reflect on what can be improved along the way would be a crucial step for schools.
Insufficient integration between medical school education and clinical experience	• Coherence: highlight the importance of integrating DHC across the continuum of medical education. • Cognitive participation: involve both academic and clinical faculty in designing integrated curricula that seamlessly blend theory and practice. • Collective action: develop joint academic-clinical initiatives and placements that reinforce DHC training in real-world settings. • Reflexive monitoring: evaluate the effectiveness of integrated programs and make adjustments to enhance alignment between education and practice.
Limited IT integration within the health care industry	• Coherence: communicate the critical role of IT in supporting DHC and improving health care outcomes. • Cognitive participation: collaborate with IT professionals and health care administrators to prioritize IT integration. To ensure that digital health care technologies can be used safely and effectively by clinicians, new technology or equipment introduced for clinical practice needs to be installed by IT personnel with knowledge of the health care system and with input from health care professionals so that the latter's needs are met. • Collective action: advocate for investments in IT infrastructure and training to support DHC initiatives. At the national level, a move towards interoperability of systems that allow users to share data would also facilitate students' adaptation to new systems in different health care settings. • Reflexive monitoring: continuously assess the state of IT integration and address gaps through ongoing improvement efforts.

XSL•FO
RenderX

[a]DHC: digital health competencies.

[b]DT: digital technology.

Limitations of the Study

This study has several limitations that should be acknowledged. First, the focus on the perspectives of organizational leaders may not fully represent the experiences of frontline educators or students, limiting the generalizability of the findings. Furthermore, interviewing organizational leaders may introduce a bias toward presenting their organizations in a favorable light. They may be reluctant to express views that could be perceived as critical of their organizations. This concern may stem from the constraints they feel due to their roles or the public image of their organizations. As a result, their responses might reflect a more measured or politically cautious perspective. To address this, we incorporated triangulation by cross-referencing their responses with published articles on similar topics. This approach provided a more balanced perspective, though we recognize the inherent limitations in capturing the full organizational dynamics.

Second, the relatively small sample size, while sufficient to achieve thematic saturation, may constrain the breadth of insights. We also recognize that the unique sociopolitical, cultural, and economic context of Singapore may limit the generalizability of our findings to other settings. Singapore's centralized governance and relatively small population create conditions that may differ from other countries. Consequently, while the insights from our study provide valuable lessons, they should be interpreted with caution when applying them to contexts with different governance structures or cultural dynamics.

The third limitation was the gender imbalance among the organizational leaders interviewed, with 94% (31/33) male and only 6% (2/33) female participants. While this reflects the current leadership demographics within public health care institutions, the barriers and challenges identified in our research are rooted in institutional and structural factors rather than individual-level or gender-specific experiences. As such, we do not expect that the gender distribution significantly influenced the findings. However, future research could benefit from a more gender-diverse sample to explore whether different leadership perspectives might offer additional insights or nuances.

Another limitation of our study is the use of Zoom for interviews, which was necessitated by the COVID-19 pandemic and which might have influenced the depth and dynamics of the discussions compared to in-person interviews. In face-to-face settings, nonverbal cues such as body language, eye contact, and physical proximity play a significant role in building rapport and fostering a more comfortable environment for in-depth conversations. These subtle cues can often provide valuable insights into a participant's emotional state, engagement level, and willingness to share more personal or sensitive information. Nonetheless, the insights obtained through Zoom still offer valuable contributions to understanding the barriers to DHE integration.

In addition, we acknowledge that the NPT's focus on individual experiences may not fully capture the diversity of perspectives of multiple stakeholders. To address this, we triangulated our data by comparing the findings across participants from various public health care clusters to identify any consistencies and divergences in opinions. This helped to ensure that the normalization process was not unduly influenced by any single group. By addressing these challenges, we believe our study provides a more nuanced understanding of NPT, particularly in contexts where contextual variations and diverse stakeholder groups are at play. These adaptations strengthen the applicability of NPT and offer valuable insights for its broader use in similar settings.

While we have made considerable efforts to adapt NPT to our specific context, we recognize that there may still be limitations in generalizing our findings across very different settings. Thus, future research should explore how NPT applies in more varied environments with larger sample sizes to further validate our findings. Furthermore, given that normalization is a gradual process, further studies should also conduct longitudinal follow-up assessments to monitor changes over time.

Strengths of the Study

This qualitative study informs us about the institutional and structural barriers present in Singapore's medical school curricula. The diverse sample of this study, spanning various health care institutions and specialties, yielded rich data. Participants possessed extensive organizational leadership experience and were attuned to the needs of contemporary clinical practice. Unlike previous research focusing mainly on institutional inertia and pedagogical strategies [5,34-37], this study uncovered structural barriers as well.

While findings may seem limited to Singapore's context, applying relevant NPT constructs could render results applicable globally since many other high-income countries faced similar challenges in technological development and curriculum digitalization [3,12,38]. Furthermore, the identified barriers necessitated universal solutions extending beyond Singapore.

A potential line of future research would be to gather the views of medical innovators and entrepreneurs to explore other barriers to the effective adoption of DT in health care institutions. Another area would be to evaluate the ways in which DHC training among medical trainees and graduates influences the efficiency and cost-effectiveness of health care delivery. This research could provide valuable insights into how DHC in medical education affects not only the preparedness of new health care professionals but also the overall performance of health care organizations.

Conclusions

Focusing on the perspectives of doctors in organizational leadership roles provides a comprehensive understanding of the barriers to incorporating DHE into Singapore's medical curricula. Their strategic insight, policy influence, experience with system-wide challenges, understanding of the

XSL•FO
RenderX

education-practice gap, resource management capabilities, and expertise in innovation and change management are invaluable for developing practical, effective, and sustainable strategies to address these barriers.

Unlike previous studies focusing solely on gaps within schools, our findings underscored the importance of collaborations with professional bodies and health care institutions to overcome various barriers. By applying NPT, this study provides a structured approach to understanding and overcoming the barriers. It offers a roadmap for other countries facing similar challenges in DHE. However, NPT should be seen as adaptable, requiring regular reevaluation to accommodate dynamic changes in the field.

Acknowledgments

This study was funded by SingHealth Duke-NUS Medicine Academic Clinical Programme under Seah Cheng Siang Distinguished Professorship in Medicine. The authors would like to thank the participants of this study for their invaluable insights and the reviewers of this journal for their helpful comments. They are also grateful to Associate Professor Warren Fong Weng Sen for his helpful feedback on the draft of this manuscript.

Conflicts of Interest

None declared.

References

1. Gopal G, Suter-Crazzolara C, Toldo L, Eberhardt W. Digital transformation in healthcare - architectures of present and future information technologies. Clin Chem Lab Med 2019;57(3):328-335 [FREE Full text] [doi: 10.1515/cclm-2018-0658] [Medline: 30530878]
2. Aungst TD, Patel R. Integrating digital health into the curriculum-considerations on the current landscape and future developments. J Med Educ Curric Dev 2020;7:2382120519901275 [FREE Full text] [doi: 10.1177/2382120519901275] [Medline: 32010795]
3. Edirippulige S, Brooks P, Carati C, Wade V, Smith A, Wickramasinghe S, et al. It's important, but not important enough: eHealth as a curriculum priority in medical education in Australia. J Telemed Telecare 2018 Dec;24(10):697-702 [FREE Full text] [doi: 10.1177/1357633X18793282] [Medline: 30343657]
4. Khurana M. Keeping Pace: the need for digital health education in medical schools. Acad Med 2020;95(11):1629-1630 [FREE Full text] [doi: 10.1097/ACM.0000000000003672] [Medline: 33109967]
5. Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P. Introducing artificial intelligence training in medical education. JMIR Med Educ 2019;5(2):e16048 [FREE Full text] [doi: 10.2196/16048] [Medline: 31793895]
6. Tudor Car L, Kyaw BM, Nannan Panday RS, van der Kleij R, Chavannes N, Majeed A, et al. Digital health training programs for medical students: scoping review. JMIR Med Educ 2021;7(3):e28275 [FREE Full text] [doi: 10.2196/28275] [Medline: 34287206]
7. Zainal H, Xiaohui X, Thumboo J, Yong FK. Exploring the views of Singapore junior doctors on medical curricula for the digital age: a case study. PLoS One 2023;18(3):e0281108 [FREE Full text] [doi: 10.1371/journal.pone.0281108] [Medline: 36862708]
8. Zainal H, Xiaohui X, Thumboo J, Kok Yong F. Digital competencies for Singapore's national medical school curriculum: a qualitative study. Med Educ Online 2023 Dec;28(1):2211820 [FREE Full text] [doi: 10.1080/10872981.2023.2211820] [Medline: 37186901]
9. Gillissen A, Kochanek T, Zupanic M, Ehlers J. Medical students' perceptions towards digitization and artificial intelligence: a mixed-methods study. Healthcare (Basel) 2022 Apr 13;10(4):A [FREE Full text] [doi: 10.3390/healthcare10040723] [Medline: 35455898]
10. Machleid F, Kaczmarczyk R, Johann D, Balčiūnas J, Atienza-Carbonell B, von Maltzahn F, et al. Perceptions of digital health education among European medical students: mixed methods survey. J Med Internet Res 2020;22(8):e19827 [FREE Full text] [doi: 10.2196/19827] [Medline: 32667899]
11. Sit C, Srinivasan R, Amlani A, Muthuswamy K, Azam A, Monzon L, et al. Attitudes and perceptions of UK medical students towards artificial intelligence and radiology: a multicentre survey. Insights Imaging 2020;11(1):14 [FREE Full text] [doi: 10.1186/s13244-019-0830-7] [Medline: 32025951]
12. Sorg H, Ehlers JP, Sorg CGG. Digitalization in medicine: are German medical students well prepared for the future? Int J Environ Res Public Health 2022;19(14):8308 [FREE Full text] [doi: 10.3390/ijerph19148308] [Medline: 35886156]
13. Chen M, Safdar N, Nagy P. Should medical schools incorporate formal training in informatics? J Digit Imaging 2011;24(1):1-5 [FREE Full text] [doi: 10.1007/s10278-009-9249-x] [Medline: 19908095]
14. Hurley KF, Taylor B, Postuma P, Grace P. What are Canadian medical students learning about health informatics. eJHI 2011;6 [FREE Full text]
15. May C, Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. Sociology 2009 Jun 15;43(3):535-554 [FREE Full text] [doi: 10.1177/0038038509103208]

XSL•FO
RenderX

16. Petersson L, Larsson I, Nygren J, Nilsen P, Neher M, Reed JE, et al. Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. BMC Health Serv Res 2022 Jul 01;22(1):850 [FREE Full text] [doi: 10.1186/s12913-022-08215-8] [Medline: 35778736]
17. Enticott J, Braaf S, Johnson A, Jones A, Teede H. Leaders' perspectives on learning health systems: a qualitative study. BMC Health Serv Res 2020;20(1):1087 [FREE Full text] [doi: 10.1186/s12913-020-05924-w] [Medline: 33243214]
18. Alanazi A. Digital leadership: attributes of modern healthcare leaders. Cureus 2022;14(2):e21969 [FREE Full text] [doi: 10.7759/cureus.21969] [Medline: 35282530]
19. Kong Chiang L. FAQs. School of Medicine NTU. URL: https://www.ntu.edu.sg/medicine/about-us/faqs#Content_C020_Col00 [accessed 2022-12-22]
20. Samarasekera D, Ooi S, Yeo S, Hooi S. Medical education in Singapore. Med Teach 2015 Aug;37(8):707-713 [FREE Full text] [doi: 10.3109/0142159X.2015.1009026] [Medline: 25693792]
21. Zainal H, Xin X, Thumboo J, Fong KY. Medical school curriculum in the digital age: perspectives of clinical educators and teachers. BMC Med Educ 2022;22(1):428 [FREE Full text] [doi: 10.1186/s12909-022-03454-z] [Medline: 35659212]
22. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol Jan 2006;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
23. O'Brien BC, Harris I, Beckman T, Reed D, Cook D. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
24. Ernstmann N, Ommen O, Neumann M, Hammer A, Voltz R, Pfaff H. Primary care physician's attitude towards the German e-health card project--determinants and implications. J Med Syst 2009;33(3):181-188 [FREE Full text] [doi: 10.1007/s10916-008-9178-0] [Medline: 19408451]
25. Safi S, Thiessen T, Schmailzl KJ. Acceptance and resistance of new digital technologies in medicine: qualitative study. JMIR Res Protoc 2018;7(12):e11072 [FREE Full text] [doi: 10.2196/11072] [Medline: 30514693]
26. Jeffe DB, Andriole DA, Hageman HL, Whelan AJ. The changing paradigm of contemporary U.S. allopathic medical school graduates' career paths: analysis of the 1997-2004 national AAMC graduation questionnaire database. Acad Med 2007;82(9):888-894 [FREE Full text] [doi: 10.1097/ACM.0b013e31812f797e] [Medline: 17726402]
27. Banerjee R, George P, Priebe C, Alper E. Medical student awareness of and interest in clinical informatics. JAMIA Internet 2015;22(1):e42-e47 [FREE Full text] [doi: 10.1093/jamia/ocu046] [Medline: 25726567]
28. Kim KJ, Park JH, Lee YH, Choi K. What is different about medical students interested in non-clinical careers? BMC Med Educ 2013;13:81 [FREE Full text] [doi: 10.1186/1472-6920-13-81] [Medline: 23731551]
29. Fung OW, Mulholland A, Bondy M, Driedger M, Kendall CE. Implementing experiential learning logs addressing social accountability into undergraduate medical clerkship education. Can Med Educ J 2023;14(2):146-149 [FREE Full text] [doi: 10.36834/cmej.73907] [Medline: 37304626]
30. Liao F, Murphy D, Wu JC, Chen CY, Chang CC, Tsai PF. How technology-enhanced experiential e-learning can facilitate the development of person-centred communication skills online for health-care students: a qualitative study. BMC Med Educ 2022;22(1):60 [FREE Full text] [doi: 10.1186/s12909-022-03127-x] [Medline: 35078482]
31. Pereira AG, Kim M, Seyward M, Nesbitt B, Pitt MB, Minnesota Epic101 Collaborative. Collaborating for competency-a model for single electronic health record onboarding for medical students rotating among separate health systems. Appl Clin Inform 2018;9(1):199-204 [FREE Full text] [doi: 10.1055/s-0038-1635096] [Medline: 29564849]
32. Chan KS, Zary N. Applications and challenges of implementing artificial intelligence in medical education: integrative review. JMIR Med Educ 2019;5(1):e13930 [FREE Full text] [doi: 10.2196/13930] [Medline: 31199295]
33. Borges do Nascimento IJ, Abdulazeem H, Vasanthan LT, Martinez EZ, Zucoloto ML, Østengaard L, et al. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. NPJ Digit Med 2023;6(1):161 [FREE Full text] [doi: 10.1038/s41746-023-00899-4] [Medline: 37723240]
34. Foadi N, Koop C, Mikuteit M, Paulmann V, Steffens S, Behrends M. Defining learning outcomes as a prerequisite of implementing a longitudinal and transdisciplinary curriculum with regard to digital competencies at hannover medical school. J Med Educ Curric Dev 2021;8:23821205211028347 [FREE Full text] [doi: 10.1177/23821205211028347] [Medline: 34368455]
35. Behrends M, Steffens S, Marschollek M. The implementation of medical informatics in the national competence based catalogue of learning objectives for undergraduate medical education (NKLM). Stud Health Technol Inform 2017;243:18-22 [FREE Full text] [Medline: 28883161]
36. Hersh W, Biagioli F, Scholl G, Gold J, Mohan V, Kassakian S. From competencies to competence: model, approach, and lessons learned from implementing a clinical informatics curriculum for medical students. In: Health Professionals' Education in the Age of Clinical Information Systems, Mobile Computing and Social Networks. United States: Academic Press; 2017:269-287.
37. Haag M, Igel C, Fischer M, German Medical Education Society (GMA), Committee “Digitization – Technology-Assisted LearningTeaching”, Joint working group “Technology-enhanced TeachingLearning in Medicine (TeLL)” of the German Association for Medical Informatics, BiometryEpidemiology (gmds)the German Informatics Society (GI). Digital teaching

XSL•FO
RenderX

and digital medicine: a national initiative is needed. GMS J Med Educ 2018;35(3):Doc43 [FREE Full text] [doi: 10.3205/zma001189] [Medline: 30186953]
38. Cutrer WB, Spickard WA, Triola MM, Allen BL, Spell N, Herrine SK, et al. Exploiting the power of information in medical education. Med Teach 2021;43(sup2):S17-S24 [FREE Full text] [doi: 10.1080/0142159X.2021.1925234] [Medline: 34291714]

Abbreviations

AI: artificial intelligence
DHC: digital health competencies
DHE: digital health education
DT: digital technology
EHR: electronic health record
ICT: information and communication technology
LKCMedicine: Lee Kong Chian School of Medicine
NPT: Normalization Process Theory
NTU: Nanyang Technological University
NUS: National University of Singapore
OL: organizational leader
PI: principal investigator
QI: quality improvement
RICE: Research Integrity, Compliance, and Ethics
YLL: Yong Loo Lin School of Medicine

Edited by J Moen; submitted 25.07.24; peer-reviewed by J Walsh, SS Lee; comments to author 18.12.24; revised version received 29.12.24; accepted 20.01.25; published 07.03.25.

Please cite as:
Zainal H, Xiao Hui X, Thumboo J, Kok Yong F
Organizational Leaders' Views on Digital Health Competencies in Medical Education: Qualitative Semistructured Interview Study
JMIR Med Educ 2025;11:e64768
URL: https://mededu.jmir.org/2025/1/e64768
doi:10.2196/64768
PMID:40053774

©Humairah Zainal, Xin Xiao Hui, Julian Thumboo, Fong Kok Yong. Originally published in JMIR Medical Education (https://mededu.jmir.org), 07.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Performance of Plug-In Augmented ChatGPT and Its Ability to Quantify Uncertainty: Simulation Study on the German Medical Board Examination

Julian Madrid[1], MD; Philipp Diehl[1], MD, PhD; Mischa Selig[2,3], PhD; Bernd Rolauffs[2,3], MD, PhD; Felix Patricius Hans[2,4], MD, MSc; Hans-Jörg Busch[2,4], MD, PhD; Tobias Scheef[2,5*], MD; Leo Benning[2,4*], MPH, MD

[1]Department of Cardiology, Pneumology, Angiology, Acute Geriatrics and Intensive Care, Ortenau Klinikum, Klosterstrasse 18, Lahr, Germany

[2]Faculty of Medicine, University of Freiburg, Freiburg, Germany

[3]G.E.R.N. Research Center for Tissue Replacement, Regeneration and Neogenesis, Department of Orthopedics and Trauma Surgery, University of Freiburg, Freiburg, Germany

[4]University Emergency Center, Medical Center, University of Freiburg, Freiburg, Germany

[5]Department of Diagnostic and Interventional Radiology, Medical Center, University of Freiburg, Freiburg, Germany

[*]these authors contributed equally

Corresponding Author:
Julian Madrid, MD
Department of Cardiology, Pneumology, Angiology, Acute Geriatrics and Intensive Care, Ortenau Klinikum, Klosterstrasse 18, Lahr, Germany

Abstract

Background: The GPT-4 is a large language model (LLM) trained and fine-tuned on an extensive dataset. After the public release of its predecessor in November 2022, the use of LLMs has seen a significant spike in interest, and a multitude of potential use cases have been proposed. In parallel, however, important limitations have been outlined. Particularly, current LLMs encounter limitations, especially in symbolic representation and accessing contemporary data. The recent version of GPT-4, alongside newly released plugin features, has been introduced to mitigate some of these limitations.

Objective: Before this background, this work aims to investigate the performance of GPT-3.5, GPT-4, GPT-4 with plugins, and GPT-4 with plugins using pretranslated English text on the German medical board examination. Recognizing the critical importance of quantifying uncertainty for LLM applications in medicine, we furthermore assess this ability and develop a new metric termed “confidence accuracy” to evaluate it.

Methods: We used GPT-3.5, GPT-4, GPT-4 with plugins, and GPT-4 with plugins and translation to answer questions from the German medical board examination. Additionally, we conducted an analysis to assess how the models justify their answers, the accuracy of their responses, and the error structure of their answers. Bootstrapping and CIs were used to evaluate the statistical significance of our findings.

Results: This study demonstrated that available GPT models, as LLM examples, exceeded the minimum competency threshold established by the German medical board for medical students to obtain board certification to practice medicine. Moreover, the models could assess the uncertainty in their responses, albeit exhibiting overconfidence. Additionally, this work unraveled certain justification and reasoning structures that emerge when GPT generates answers.

Conclusions: The high performance of GPTs in answering medical questions positions it well for applications in academia and, potentially, clinical practice. Its capability to quantify uncertainty in answers suggests it could be a valuable artificial intelligence agent within the clinical decision-making loop. Nevertheless, significant challenges must be addressed before artificial intelligence agents can be robustly and safely implemented in the medical domain.

(JMIR Med Educ 2025;11:e58375) doi:10.2196/58375

KEYWORDS

medical education; artificial intelligence; generative AI; large language model; LLM; ChatGPT; GPT-4; board licensing examination; professional education; examination; student; experimental; bootstrapping; confidence interval

XSL•FO
RenderX

Introduction

The GPT—recently updated to its fourth iteration (GPT-4)—is a generative and autoregressive large language model (LLM). It is pretrained on a vast corpus of internet text and fine-tuned on a labeled dataset using a transformer architecture [1-3]. GPT generates coherent and contextually appropriate text. It likely discovered a semantic grammar of language (ie, semantic regularities), enabling it to construct semantically and syntactically correct sentences [4,5]. However, GPT does not perform meaningful computations on symbolic representations [4-8]. The Wolfram language, a Turing-complete computational language, in contrast, allows such symbolic representation. GPT and the Wolfram language combined hence cover 2 different aspects of human cognition [4,9,10]. Combining these features, particularly when computation and symbolic representations are needed, represents a significant step toward general artificial intelligence (AI). This combination has already been successfully used to examine contradictions in Einstein Special Theory of Relativity equations [11].

In the light of these technological advances, LLMs show increasing promise in supporting medical training and practice. However, the models must acquire an in-depth and accurate representation of medical knowledge to be used in these sensitive domains. A medical board examination exemplifies these domains well, as it determines the qualification of medical students to obtain their license to practice medicine.

Our primary outcome is the model's ability to achieve the minimum required score for passing the 2 written parts of the German medical licensing examination. This task poses a different challenge to an LLM than medical board examinations in the English language [12,13], as the performance of such models in other languages and in combination with more recent GPT versions and available plugins has not been explored. In the medical field, where mistakes can have severe consequences, assessing the amount of uncertainty is of paramount importance [14]. It is therefore crucial to gain insights into the depth and structure the LLMs have of the medical knowledge representation and where its limitations lie [15]. Hence, our secondary outcomes were the total correct answer rates, the presence of logical justification of the answer, the presence of information internal to the question, the presence of information external to the question, the confidence GPT displays in its answers, the difficulty of the question, information errors, logical errors, reasoning errors, and the correctness of a second try answer when the first answer was wrong. Insights into these 2 dimensions of outcomes can contribute to facilitating a meaningful use of novel LLM technologies in the medical domain.

Methods

Medical Board Examination Dataset

The German medical board examination consists of 3 steps. The first board examination, taken after 2 years of study, primarily covers basic natural sciences. It comprises 320 questions, which students answer over 2 consecutive days. The second board examination takes place after 6 years of study. It likewise consists of 320 medical questions, which students answer over 3 consecutive days. The third board examination, also after 6 years of study, is an oral examination and was, hence, excluded from this study. The German medical board examination takes place biannually, once in spring and once in fall. As a representative sample, we used the medical board examination from spring 2023. We excluded questions the medical board examination committee deemed inconsistent with the medical literature in the regular post examination review of the content. Additionally, we did not include questions displaying images, as GPT models could not analyze them at the time of our analysis. Furthermore, LLMs are not able to analyze images, GPT4vision which became broadly accessible in the second half of 2023 combines computer vision algorithms—which generate a text description of images—and LLMs to analyze this text. All questions and answers were exported from AMBOSS SE, a German medical education content creator and service provider.

GPT Models and Prompt Engineering

We evaluated several GPT models with varying characteristics using OpenAI's web interface. The models tested included GPT-3.5, GPT-4, GPT-4 integrated with the Wolfram, ScholarAI, and Web Request (WeGPT.ai) plugins, and GPT-4 integrated with the Wolfram, ScholarAI, Web Request plugins, and an additional feature for translating German inputs into English. We did not investigate earlier versions of GPT as they demonstrated lower performance in a similar study on the American medical board examination [12].

Creating a precise and adequate context is crucial for generating expected results [16,17]. Thus, we aimed to be as specific as possible, simulating the context of a medical student taking the medical board examination. The prompts hence included the request to answer each respective question with 5 possible answers, where only 1 answer was correct. We asked the models to justify their choices based on the provided patient case information, and to estimate their confidence in the answer's accuracy as a percentage of maximal confidence (ie, 100%). If the selected answer was incorrect, the GPT models were asked to explain their mistake in a second attempt. For the GPT-4 model with plugin integration, we asked the model to use the available plugins (Wolfram, ScholarAI, and Web Request). For the GPT-4 model with plugin integration and English translation, we first asked the model to translate the input into the English language, and then to use the translated text to perform the abovementioned tasks. All used prompts are available in Multimedia Appendix 1.

Model Testing and Outcome Parameters

For each GPT model, we used the appropriate prompt followed by the question and the possible answers. The investigators then analyzed the GPT's answer to assess the defined primary and secondary outcomes, which were either binary or in proportions. In cases of uncertainty, the investigators (JM, TS, and LB) convened to resolve the issue.

First, the correctness of the answer was recorded (binary variable), followed by the presence of logical justification, the presence of information internal to the question, and the

XSL•FO
RenderX

presence of information external to the question (binary variables).

Next, we recorded the model's confidence in its answer (proportion), and the difficulty of the question, derived from the number of students who answered correctly on the AMBOSS platform (proportion).

To enhance our understanding of where GPT models falter, we sought to classify potential errors. As literature on error types is limited, we conducted a formal analysis to determine distinctive error types and established a formal definition. We propose a classification into 3 categories: information error, logical error, and reasoning error.

The GPT response can be formalized as "answer A" is given "link" because of "information B." There are only three possibilities for errors: (1) "answer A" is incorrect because "information B" is incorrect—termed an information error; (2) "answer A" is incorrect while "information B" is correct, but the link between them is incorrect—termed a logical error; (3) "answer A" is incorrect, "information B" is correct, and the link between "answer A" and "information B" is correct—termed a reasoning error (Figure 1). If the answer provided was incorrect, the investigator informed the GPT of its faulty answer, recorded whether it understood its mistake, and provided the correct answer in a second attempt. In the models with integrated plugin use, the active use of plugins was documented for Wolfram, ScholarAI, and Web Requests (binary variables).

XSL•FO
RenderX

Figure 1. Formal definition of error types; we propose a classification into 3 categories: information error, logical error, and reasoning error. The GPT response can be formalized as "answer A" is given "link" because of "information B." There are only three possibilities for errors: (1) "answer A" is incorrect because "information B" is incorrect—termed an information error; (2) "answer A" is incorrect while "information B" is correct, but the link between them is incorrect—termed a logical error; and (3) "answer A" is incorrect, "information B" is correct, and the link between "answer A" and "information B" is correct—termed a reasoning error.

Data Analysis

Summary statistics were calculated for the outcome variables (Table 1 and Multimedia Appendices 2 and 3). Dichotomous variables were represented by frequency and proportions with 95% CIs, while continuous variables were expressed as mean values with 95% CIs. Uncertainty calculations displayed as 95% CIs were computed via bootstrapping [18].

XSL•FO
RenderX

Table . Characteristics of GPT model answers (N=541).

	GPT-3.5	GPT-4	GPT-4 + plugin	GPT-4 + plugin + translation
Correct answer (proportion±95% CI)	373 (0.69±0.65 to 0.73)	493 (0.91±0.89 to 0.93)	493 (0.91±0.89 to 0.94)	486 (0.9±0.87 to 0.92)
Logical justification (proportion±95% CI)	479 (0.89±0.86 to 0.91)	526 (0.97±0.96 to 0.98)	529 (0.98±0.96 to 0.99)	527 (0.97±0.96 to 0.99)
Question's difficulty mean (±95% CI)	0.288 (0.272 to 0.303)	0.288 (0.272 to 0.303)	0.288 (0.272 to 0.303)	0.288 (0.272 to 0.303)
Error overall (proportion±95% CI)	168 (0.31±0.27 to 0.35)	48 (0.09±0.07 to 0.11)	48 (0.09±0.06 to 0.11)	55 (0.1±0.08 to 0.13)
Presence of internal information (proportion±95% CI)	521 (0.96±0.95 to 0.98)	537 (0.99±0.98 to 1)	537 (0.99±0.98 to 1)	537 (0.99±0.98 to 1)
Presence of external information (proportion±95% CI)	538 (0.99±0.99 to 1)	540 (1±0.99 to 1)	541 (1±1 to 1)	541 (1±1 to 1)
Information error (proportion±95% CI)	37 (0.22±0.16 to 0.29)	5 (0.1±0.02 to 0.19)	5 (0.1±0.02 to 0.2)	7 (0.13±0.05 to 0.22)
Logical error (proportion±95% CI)	61 (0.36±0.29 to 0.43)	18 (0.38±0.25 to 0.52)	12 (0.25±0.125 to 0.375)	19 (0.35±0.22 to 0.47)
Confidence mean (±95% CI)	0.912 (0.904 to 0.918)	0.938 (0.934 to 0.942)	0.919 (0.915 to 0.924)	0.919 (0.915 to 0.923)
Use of plugin Wolfram (proportion±95% CI)	N/A[a]	N/A	50 (0.09±0.07 to 0.12)	47 (0.09±0.06 to 0.11)
Reasoning error (proportion±95% CI)	72 (0.42±0.36 to 0.51)	26 (0.54±0.4 to 0.69)	30 (0.63±0.48 to 0.75)	29 (0.53±0.4 to 0.65)
Correct answer in second try (proportion±95% CI)	90 (0.54±0.46 to 0.61)	32 (0.67±0.52 to 0.79)	36 (0.75±0.63 to 0.88)	33 (0.6±0.47 to 0.73)
Use of plugin ScholarAI (proportion±95% CI)	N/A	N/A	107 (0.2±0.16 to 0.23)	47 (0.09±0.06 to 0.11)
Use of plugin web requests (proportion±95% CI)	N/A	N/A	2 (0.003±0 to 0.01)	25 (0.05±0.03 to 0.06)

[a]N/A: not applicable.

The primary outcome was determined by comparing the performance of the GPT-4 model, integrated with the plugins and the English translation, to the required passing score for the medical board examination, which is 60%. The difference of proportions was calculated with 95% CI using bootstrapping (Multimedia Appendix 4).

Subsequently, secondary outcomes were calculated: the final examination rate for each GPT model was compared to both chance and the required passing score for the medical board examination. The difference of proportions was calculated with 95% CI using bootstrapping (Multimedia Appendix 4).

The proportions of logical justification within the answer, information internal to the answer, and information external to the answer were compared between correct and incorrect responses. The difference of proportions was calculated with 95% CI using bootstrapping (Table 2 and Multimedia Appendix 5).

XSL•FO
RenderX

Table . Analysis of plugin-integrated GPT-4 model answers.

		All correct answers (n=493)	All incorrect answers (n=48)	Difference in proportions or Cohen *d* or Pearson *r* (±95% CI)	Confidence accuracy (±95% CI)
Comparison of GPT models justifications between correct and incorrect answers					
GPT-4 + plugin (N=541)					
	Logical justification (proportion ±95% CI)	493 (1±1 to 1)	36 (0.75±0.63 to 0.88)	0.25 (0.13 to 0.38)[a]	—[b]
	Internal information (proportion ±95% CI)	489 (0.99±0.983 to 998)	48 (1±1 to 1)	0 (-0.01 to 0)[a]	—
Comparison of GPT models justifications between correct and incorrect answers					
	External information (proportion ±95% CI)	493 (1±1 to 1)	48 (1±1 to 1)	0 (0 to 0)[a]	—
Confidence of GPT models compared between correct and incorrect answers					
GPT-4 + plugin (N=541)					
	Confidence mean (±95% CI)	0.923 (0.918 to 0.928)	0.886 (0.87 to 0.901)	-0.69 (-0.99 to -0.39)[c]	0.037 (0.021 to 0.053)
Comparison of question's difficulty of GPT models between correct and incorrect answers					
GPT-4 + plugin (N=541)					
	Question's difficulty mean (±95% CI)	0.279 (0.263 to 0.295)	0.379 (0.327 to 0.438)	0.57 (0.27 to 0.86)[c]	—
Correlation of confidence and question's difficulty for all answers					
GPT-4 + plugin (N=541)				-0.0874 (-0.176 to 0.004)[d]	
	Confidence mean (±95% CI)	0.920 (0.916 to 0.924)	—		—
	Question's difficulty mean (±95% CI)	—	0.288 (0.273 to 0.304)		—
Comparison of correct answers between GPT models (N=541)					
GPT-4 + plugin vs GPT-3.5					
	Correct answer rate (proportion ±95% CI)	373 (0.69±0.65 to 0.73)	493 (0.91±0.89 to 0.94)	0.22 (0.18 to 0.27)[a]	—
GPT-4 + plugin vs GPT-4					
	Correct answer rate (proportion ±95% CI)	493 (0.91±0.89 to 0.94)	493 (0.91±0.89 to 0.94)	0 (-0.03 to 0.03)[a]	—
GPT-4 + plugin vs GPT-4 + plugin + translation					
	Correct answer rate (proportion ±95% CI)	493 (0.91±0.89 to 0.94)	486 (0.9±0.87 to 0.92)	-0.01 (-0.05 to 0.02)[a]	—

[a]Difference in proportions.

[b]Not available.

[c]Cohen *d*.

[d]Pearson *r*.

The model's confidence in its answers was compared between correct and incorrect responses. Additionally, the relationship between the model's confidence in its answers and the difficulty of the question was assessed. Cohen *d* values and 95% CI were computed using a linear regression model and bootstrapping (Table 2 and Multimedia Appendices 6 and 7).

XSL•FO
RenderX

To evaluate the accuracy of the model's confidence in its answers, we developed a parameter termed confidence accuracy (CA). It is defined as follows:

CA = (confidence of correct answers in percentage – confidence of incorrect answers in percentage)/100

Consequently, this parameter can take values from −1 to 1, where 1 accurately reflects the model's uncertainty, 0 indicates no ability to quantify uncertainty, and −1 suggests incorrect quantification.

The difficulty of the question was assessed using real correct response proportions from students available on the AMBOSS platform. The difficulty was assessed as follows:

Difficulty=1 – correct answer proportion

Then, the difficulty of the question was compared between correct and incorrect answers, with Cohen *d* calculated using a linear regression model (Table 2 and Multimedia Appendix 7).

Furthermore, we compared the proportion of correct answers between models (Table 2 and Multimedia Appendix 8).

We compared the proportion of correct answers in the GPT-4 models with the proportion of correct answers in the answers where a plugin has been used. We compared the proportion of plugin usage in GPT models with German and English input. We compared the confidence of the model when using plugins to the confidence of the model overall. We compared the proportion of correct answers when averaging the 4 different models to each model in particular (Multimedia Appendix 9).

In instances where questions were accompanied by images, GPT models sometimes responded by describing the image, although the models could not access the respective images. This phenomenon is known as a type of hallucination [19]. Therefore, we compared the proportion of hallucinations present in each model when answering questions, including image questions. We calculated the proportion of correct answers for each model when keeping the questions with pictures (Multimedia Appendix 9).

We compared the different error proportions between different models. We compared the proportion of logical errors when using the Wolfram plugin to the proportion of errors when using the entire model. We compared correct second-try answers between different models (Multimedia Appendix 9).

The 95% CIs were calculated using bootstrapping. Where necessary, parametric assumptions were tested using quantile-quantile plots for normality and Levene tests for the homogeneity of variances. The independence of question answers was assumed. All statistical analyses were performed in RStudio (version 2023.06.0+421). The significance level for all tests was set a priori at 95% CI.

Results

All tests were performed on the 541 questions of the German medical board examination from spring 2023. Sub analyses were performed on other subgroups, the respective sample sizes are indicated in the appropriate tables. All results for GPT-3.5, GPT-4, GPT-4 + plugin (GPT4P), and GPT-4 + plugin + translation (GPT4PT) are listed in full detail in the tables and the supplementary materials. To ensure legibility, only relevant results are addressed in the results section.

Descriptive statistics with CIs for the first board examination, second board examination, and the overall examination are displayed in Table 1 and Multimedia Appendices 2 and 3.

All models performed significantly better than chance. Furthermore, all GPT models were significantly better than the required proportion to pass the final medical board examination.

All GPT models had a significantly higher proportion of providing a logical justification for correct answers compared to incorrect answers (Table 2 and Multimedia Appendix 5). Yet, there was no statistical significance for the proportion of used internal information for correct and incorrect answers (Table 2 and Multimedia Appendix 5). Similarly, there was no statistical significance for the proportion of used external information for correct and incorrect answers (Table 2 and Multimedia Appendix 5).

Although generally high for both incorrect and correct answers, models had a confidence mean which was significantly higher for correct answers than incorrect answers (Table 2 and Multimedia Appendix 6). This is reflected in CA values significantly different from zero: GPT-3.5 (0.028, 95% CI 0.011 to 0.048), GPT-4 (0.041, 95% CI 0.023 to 0.062), GPT4P (0.037, 95% CI 0.021 to 0.053), and GPT4PT (0.043, 95% CI 0.028 to 0.059).

From all models, only GPT4P made significantly more reasoning errors than logical errors (0.37, 95% CI 0.125 to 0.60). All models made significantly more reasoning errors than information errors: GPT-3.5 (0.21, 95% CI 0.11 to 0.30), GPT-4 (0.44, 95% CI 0.27 to 0.60), GPT4P (0.52, 95% CI 0.31 to 0.71), and GPT4PT (0.40, 95% CI 0.20 to 0.58). All models but GPT4P made significantly more logical errors than information errors: GPT-3.5 (0.14, 95% CI 0.029 to 0.26), GPT-4 (0.27, 95% CI 0.10 to 0.44), and GPT4PT (0.22, 95% CI 0.05 to 0.38). GPT-4 (0.12, 95% CI 0.05 to 0.22) and GPT4P (0.12, 95% CI 0.02 to 0.22) made significantly less information errors than GPT3.5.

The GPT4-based models all performed better than the GPT 3.5 model in providing correct answers as reflected in the difference of correct answer proportions (Table 2 and Multimedia Appendix 8). However, no GPT4-based model was better than another GPT4-based model, as reflected in the difference of correct answer proportions (Table 2 and Multimedia Appendix 8).

Discussion

Primary Outcome

All GPT models assessed performed above the minimum required score of 60%. The GPT-4 models performed particularly well, outperforming most students in the given examinations. Specifically, for the first board examination, all GPT-4 models performed better than 98.6% of students. For

XSL•FO
RenderX

the second board examination, they surpassed 95.8% of students, as detailed in the records of the examining body [20].

In general, there was a significant gap between GPT-3.5 and the GPT-4 models. The more recent models, with substantially more parameters and the capacity to remember longer prompts, appear to increase the accuracy of responses. However, we observed no additional benefit when GPT-4 models were paired with plugins.

The use of plugins did not yield a higher proportion of correct answers than the standard model. It is possible that GPT-4 already achieves a very high rate of accuracy, resulting in a ceiling effect. Hence, the addition of plugins may not offer a significant advantage for the questions prompted.

During our study, we noted that the Wolfram plugin was frequently used for more complex calculations. Yet, in the context of clinically applicable questions, complex mathematical procedures are typically not required and the use of symbolic language is usually not required. Thus, using the Wolfram Alpha plugin is likely more beneficial for questions that involve extensive computations or advanced mathematical problems requiring symbolic representations. The ScholarAI plugin was activated for complex informational queries, but the resulting papers were not consistently useful. Surprisingly, the Internet Access plugin (WeGPT.ai) was the least used. This may be because answering medical questions typically demands expert-level knowledge, and general internet searches do not provide sufficiently specific information. Moreover, since the model has been trained on a vast amount of internet data, it likely already encompasses the knowledge available from the world wide web within its parameters.

We speculated that posing questions in German might hinder the model's access to the broader body of knowledge available in English. However, this was not the case; the GPT model equipped with translation capabilities did not outperform the GPT-4 models without translation features. The GPT model likely abstracts high-level concepts and is not impeded by the language of the queries. This aligns with the LLMs' transformer architecture, which accesses higher-level concepts prior to translating text into another language [21].

Interestingly, the GPT-4 model with translation invoked plugins less frequently than the model without translation. We hypothesize that plugin calls occur at a lower level in the neural network, making them less necessary in English due to the larger available language corpus. In German, the model might need to delve deeper into the latent representation of concepts not tied to a specific language. However, this remains speculative and warrants further research.

Secondary Outcomes

While all models provided a very high proportion of logical justification for correct answers, it was significantly less extensive for incorrect answers. However, upon further analysis, we did not detect a significant difference in the proportion of internal information from the question in the answer or in the use of external information not contained in the question between correct and incorrect answers. One study already assessed the presence of logical justification in answers to United States Medical Licensing Examination questions, where all answers exhibited logical justification regardless of their accuracy [12]. Hence, this metric could not be used as a discriminator for correctness.

We were unable to demonstrate a significant correlation between the model's confidence in an answer and the difficulty level of the question for humans. This suggests that the model's interpretation of question difficulty differs from that of humans. However, as with humans, the model showed improved performance on easier questions compared to more challenging ones. Thus, it appears that the representation of question difficulty is distinct between LLMs and humans.

Conceptual Implications

Use for Medical Education

This performance suggests that LLMs such as GPT could assume a greater role in medical education, as their integration could significantly change the conventional approach to medical education, which has traditionally emphasized the acquisition and maintenance of medical knowledge. The emergence of AI agents with superior information retention abilities, however, prompts a reevaluation of our educational focus. In this light, teaching methodologies could shift toward navigating and structuring available information with respective AI agents. The approach could hence shift from retaining information to learning how to efficiently access information and deeply understand these systems, along with their benefits and drawbacks.

Use in Clinical Practice

The utility of LLMs is not limited to educational settings but also extends to clinical practice. Although LLMs may not be as effective in highly specialized tasks where dedicated machine learning algorithms excel—for instance, XGBoost in identifying pulmonary embolisms [22-24] — LLMs are highly proficient in text processing and information integration from diverse algorithms [25]. This positions them as intelligent medical assistants, capable of transforming complex data into narratives that are comprehensible in a human context. Currently, clinicians have a limited understanding of AI agents and their functions. Clinicians must, therefore, gain a thorough understanding of how various AI agents function, including their strengths and weaknesses.

With insufficient knowledge on the principles of LLM-based assistants, clinicians are at risk of blindly following such assistant's guidance without fully understanding its operations [26,27]. Due to the inherent complexity of LLMs, which often function as a black box, we can only partially monitor their operations at varying levels of complexity and behavior [26]. Given the marginal uncertainty intrinsic to such complex models, the AI agent should not supplant clinicians in decision-making, but rather provide additional informed perspectives.

To serve as a useful assistant, however, the assessment of uncertainty for any output provided by such is crucial. The key attribute enabling this evaluation is the ability to quantify uncertainty, a trait humans are presumed to possess [14]. For

XSL•FO
RenderX

LLM-based assistants to provide a comparable estimate, a standardized measure is needed to gauge the confidence in an AI agent's output. For binary outcomes such as healthy or diseased, metrics such as specificity, sensitivity, and area under the curve are effective. For more complex queries with multiple potential answers—as managed by LLMs—traditional measures such as sensitivity and specificity are inadequate. We therefore developed a new metric called "confidence accuracy" (CA) which correlates the confidence assigned to an answer with its empirical accuracy. This allows for the quantification of uncertainty, crucial for clinical decision-making. Although our work showed that all GPT models have the ability to quantify uncertainty, the expression in percentage does not seem to reflect the confidence for any specific decision (ie, the models were overall largely overconfident). Although statistically different from zero, CA values were consistently close to zero. New LLM methodologies aim to enhance this by incorporating uncertainty estimation [28]. Future AI agents should be fine-tuned using the CA metric in order to improve uncertainty quantification, a critical objective for implementing AI as a supportive tool for physicians in clinical environments.

Identified Errors

We observed that GPT models commit different types of errors, particularly reasoning errors. Reasoning errors typically occur in situations where multiple options are correct, but one is more critical than the other. GPT models over proportionately make reasoning errors likely because this skill is acquired through human experience and is challenging to learn from text-based web sources. The second most common error type in GPT models was logical errors. Since LLMs use a statistical approach to reconstruct human-written text, we anticipated difficulties with logic and mathematics, which require formal symbolic representation [4-8]. We hypothesized that the Wolfram plugin, using the Wolfram language, would mitigate these challenges. Yet, using the Wolfram plugin did not reduce the number of logical errors. Finally, fewer information errors were observed compared to other error types across all GPT models. This likely reflects the strength of these LLMs, which have assimilated a vast corpus of knowledge. In addition to the 3 error types derived from the informational and logical structure of GPT's answers, there are 2 sources of bias that arise prior to answer generation. First, due to the stochastic nature of token generation, there is likely a stochastic bias inherent in all GPT responses. Second, due to in-context generation conditioned by the prompting strategy, a systematic bias probably occurs as well. We attempted to mitigate the stochastic bias by averaging the results from all models and selecting the most common outcome. However, the performance of such averaged models did not surpass that of the GPT-4 models.

To assess whether the GPT models could recognize and correct their own mistakes, we prompted them to attempt another answer after providing incorrect responses. In most instances, the model would acknowledge the mistake and provide the correct answer along with a new explanation. This phenomenon could likely be attributed to the differing mechanics of forward and backward reasoning in LLMs. With forward reasoning, the LLM calculates the probability of the next token without a specific reasoning goal [29]. In contrast, backward reasoning enables the LLM to better contextualize the information. It is crucial to note, however, that we did not request the model to immediately reassess the answer; instead, we informed it of the answer's incorrectness before asking for a reevaluation [29]. Future studies could further investigate the model's ability to self-correct without prior notification of its errors.

In instances where questions were accompanied by images (ie, the model did not have access to the images), GPT models, particularly GPT-3.5, often responded by describing the image that the model had not actually seen. This unexpected information error, known as a hallucination [19], persisted in the GPT-4 models, albeit at a significantly reduced frequency compared to GPT-3.5. Nevertheless, the propensity for overconfidence in entirely fabricated information remains a challenge for the latest generation of LLMs and is a phenomenon not fully understood [30].

Limitations

Technological Limitations of LLMs

Although the results were impressive with GPT outperforming most students in the German medical board examination, it is crucial to remember that these models still possess significant limitations. At the time of our data collection, GPT-4 was incapable of interpreting medical images, such as chest x-rays or histological samples. This is a considerable drawback, given that medical information is inherently multimodal, and the ability to integrate multimodal data will be essential for the adoption of such models in academic and clinical settings. It is anticipated that future GPT iterations and other LLMs will be fully multimodal, which necessitates additional research to evaluate their performance across a more diverse array of questions.

A second concern relates to the stochastic nature of token generation, meaning that answers may vary slightly when questions are posed multiple times [31].

A third concern pertains to the prompt sensitivity of LLMs. This trait can be advantageous as it allows the incorporation of context into the generation of meaningful output and may contribute to the models' Bayesian characteristics [32]. However, prompt sensitivity also increases the risk of systematic errors with repetitive use of the same prompt. Prompt engineering is a discipline that emerged in trying to minimize systematic errors [33,34].

Within the extensive volume of data available online, there are significant risks of bias. Given that LLMs are trained on vast datasets, there is an inherent risk of adopting biases from the underlying data structures. However, fine-tuning through supervised learning on labeled data can help mitigate these risks [35,36].

Limitations of the Use of LLMs in a Medical Context

Despite the seemingly immediate promise of using LLMs in both educational and clinical contexts, the current ethical and regulatory environment needs to be considered to advance the use of these novel technologies safely.

XSL•FO
RenderX

As the representation of medical information of an LLM must not be confused with medical knowledge from a medical professional, it remains crucial to enable students and medical professionals alike to identify LLM-generated outputs as such in order to interpret them very carefully. Different to, for example, a senior medical colleague providing guidance for a clinical decision, an LLM-generated output is neither based on clinical knowledge, nor experience. The risk of such confusion has been described as anthropomorphic projection and efforts for advancing these novel technologies in the medical field need to simultaneously foster the awareness of such phenomena. This differentiation resonates with the provisions of the European Union (EU) on a risk-based assessment approach [37] and, more recently, with the Bletchley Declaration [38]. The latter emphasizes the risks at the "frontier" of AI, at which we operate with the presented project.

While the concerns discussed in the context of medical education—and, more widely, training—are mainly within the realm of AI ethics, more specific limitations apply to the clinical use of these technologies. At the time of our analysis, no commercially available LLM in the EU—including the GPT versions assessed in this work—have an assigned intended medical use, a basic regulatory prerequisite for their use in a clinical context. Without such intended medical use, the Medical Device Regulation (MDR), the regulatory framework for medical devices in the EU, is not applicable. Hence, such a device would not be a medical device in the regulatory sense and could, therefore, not be used in a clinical context without irresponsible safety and liability risks. While it is not the user (eg, researchers or clinicians), but the manufacturer (eg, OpenAI for the ChatGPT models) who assigns an intended medical use—which itself comes with further regulatory requirements—the clinical use of the currently available and mostly all-purpose LLMs remains challenging.

Yet, even developing an LLM with an intended medical use and fulfilling all adjacent regulatory requirements would—as of now—not necessarily resolve the challenge centering around the clinical use of such program, as a key requisite for software as a medical device outlined in the MDR ("devices that incorporate electronic programmable systems, including software, or software that are devices in themselves, shall be designed to ensure repeatability, reliability and performance in line with their intended use." MDR Annex I, Rule 17.1 [39]) is currently considered to be violated, although this question remains subject to debate.

However, the rapid development of technological advances and the concurrent establishment of respective regulations should not be perceived as a "race to get to grips with AI" [40], but should be viewed as a co-evolution to eventually yield the best population-wide benefit from these technological advances. In this light, the emphasis of a "pro-innovation and proportionate governance," as proposed in the Bletchley Declaration, is equally as crucial as the implementation of regulatory frameworks.

Limitations of This Study

Our study has several limitations. We used a specific German medical board examination as a sample to represent the general distribution of medical questions. While it is acknowledged that questions evolve over time and may introduce bias, the objective of the medical board examination is to maintain a consistent level of difficulty, reflecting the minimum required knowledge to attain board approval for medical practice. The distribution of student grades has remained relatively stable over time, leading us to believe that this potential bias is minimal. In the model with translation, we used GPT to translate the questions before applying them to the model. Although we did not observe any, it is possible that translation errors occurred, potentially acting as a confounder in this study. In the context of the medical board examination, multiple-choice questions are posed to elicit clear answers that can be quantitatively assessed. By contrast, in a clinical setting, questions tend to be open-ended, which introduces a different dynamic. Nevertheless, we asked the model to justify its answers to glean insight into its computational process, thus rendering the questions more comparable to open-ended inquiries.

Conclusion

The performance of GPT models in the German medical board examination have surpassed both the passing threshold and the performance of most students. While GPT appears to possess a latent representation of uncertainty, it currently exhibits a significant degree of overconfidence. The introduced metric of CA could facilitate the appropriate measurement and fine-tuning of models to improve this aspect. However, there are numerous limitations that clinicians should be aware of. Challenges such as hallucinations, the stochastic nature of token generation, and prompt sensitivity are highlighted, indicating areas for further research and development. Further, we see the remaining open questions regarding the ethical and regulatory use of LLMs in the educational and clinical context, which need to be addressed on a policy level.

Authors' Contributions

JM participated in the conceptualization, data acquisition, data curation, formal analysis, investigation, methodology, project administration, software, validation, visualization, writing of the original draft, and review and editing of the writing, and should be considered the first author. PD, MS, BR, FPH, and HJB participated in the methodology and review editing and should be considered as second authors. LB and TS participated in the conceptualization, data acquisition, formal analysis, investigation, methodology, validation, review and editing of the writing, and should be considered last authors. Correspondence should be addressed to JM and LB.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Prompting strategies for different GPT models.
[DOCX File, 14 KB - mededu_v11i1e58375_app1.docx]

Multimedia Appendix 2
Question's difficulty and error structure of GPT model answers.
[XLSX File, 10 KB - mededu_v11i1e58375_app2.xlsx]

Multimedia Appendix 3
Correct answers of GPT models compared with required and random scores.
[XLSX File, 9 KB - mededu_v11i1e58375_app3.xlsx]

Multimedia Appendix 4
Comparison of correct answers between GPT models.
[XLSX File, 9 KB - mededu_v11i1e58375_app4.xlsx]

Multimedia Appendix 5
Supplementary analysis of GPT models answers (statistically significant results are highlighted in blue and statistically nonsignificant results are highlighted in brown).
[XLSX File, 12 KB - mededu_v11i1e58375_app5.xlsx]

Multimedia Appendix 6
Confidence of GPT models compared between correct and incorrect answers.
[XLSX File, 9 KB - mededu_v11i1e58375_app6.xlsx]

Multimedia Appendix 7
Relationship between question's difficulty, performance, and confidence of GPT model answers.
[XLSX File, 10 KB - mededu_v11i1e58375_app7.xlsx]

Multimedia Appendix 8
Comparison of GPT models justifications between correct and incorrect answers.
[XLSX File, 10 KB - mededu_v11i1e58375_app8.xlsx]

Multimedia Appendix 9
Performance, information content, confidence, and plugin usage of GPT model answers.
[XLSX File, 10 KB - mededu_v11i1e58375_app9.xlsx]

References

1. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. arXiv. Preprint posted online on Jun 12, 2017. [doi: 10.48550/arXiv.1706.03762]
2. Korngiebel DM, Mooney SD. Considering the possibilities and pitfalls of Generative Pre-trained Transformer 3 (GPT-3) in healthcare delivery. NPJ Digit Med 2021 Jun 3;4(1):93. [doi: 10.1038/s41746-021-00464-x] [Medline: 34083689]
3. Bubeck S, et al. Sparks of artificial general intelligence: early experiments with GPT-4. arXiv. Preprint posted online on Mar 22, 2023. [doi: 10.48550/arXiv.2303.12712]
4. Wolfram S. What Is ChatGPT Doing...and Why Does It Work?: Wolfram Media, Inc; 2023.
5. Traylor A, Feiman R, Pavlick E. AND does not mean OR: using formal languages to study language models' representations. Presented at: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2; Aug 1-6, 2021. [doi: 10.18653/v1/2021.acl-short.21]
6. Misra K, Rayz J, Ettinger A. COMPS: conceptual minimal pair sentences for testing robust property knowledge and its inheritance in pre-trained language models. Presented at: Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics; May 2-6, 2023; Dubrovnik, Croatia URL: https://aclanthology.org/2023.eacl-main [accessed 2025-02-28] [doi: 10.18653/v1/2023.eacl-main.213]

XSL•FO
RenderX

7. Kim N, Linzen T. COGS: a compositional generalization challenge based on semantic interpretation. Presented at: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP); Nov 16-20, 2020 URL: https://www.aclweb.org/anthology/2020.emnlp-main [accessed 2025-02-28] [doi: 10.18653/v1/2020.emnlp-main.731]
8. Ettinger A. What BERT is not: lessons from a new suite of psycholinguistic diagnostics for language models. Trans Assoc Comput Linguist 2020 Dec;8:34-48. [doi: 10.1162/tacl_a_00298]
9. Goertzel B. Generative AI vs AGI: the cognitive strengths and weaknesses of modern llms. arXiv. Preprint posted online on Sep 19, 2023. [doi: 10.48550/arXiv.2309.10371]
10. Vzorin G, Bukinich AM, Sedykh A, Vetrova I, Sergienko E. Emotional intelligence of GPT-4 large language model. PsyArXiv Preprints. Preprint posted online on Oct 20, 2023. [doi: 10.31234/osf.io/b6vys]
11. Bryant S. Assessing GPT-4's role as a co-collaborator in scientific research: a case study analyzing Einstein's special theory of relativity. Research Square. Preprint posted online on Apr 12, 2023. [doi: 10.21203/rs.3.rs-2808494/v1]
12. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? the implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 8;9:e45312. [doi: 10.2196/45312] [Medline: 36753318]
13. Nori H, King N, McKinney SM. Capabilities of GPT-4 on medical challenge problems. arXiv. Preprint posted online on Mar 20, 2023. [doi: 10.48550/arXiv.2303.13375]
14. Cosmides L, Tooby J. Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. Cognition 1996 Jan;58(1):1-73. [doi: 10.1016/0010-0277(95)00664-8]
15. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023 Mar 30;388(13):1233-1239. [doi: 10.1056/NEJMsr2214184] [Medline: 36988602]
16. Wang J, et al. Prompt engineering for healthcare: methodologies and applications. arXiv. Preprint posted online on Apr 28, 2023. [doi: 10.48550/arXiv.2304.14670]
17. Liu J, Shen D, Zhang Y, Dolan B, Carin L, Chen W. What makes good in-context examples for GPT-3? Presented at: Proceedings of Deep Learning Inside Out (DeeLIO 2022); May 27, 2022; Dublin, Ireland URL: https://aclanthology.org/2022.deelio-1 [accessed 2025-02-28] [doi: 10.18653/v1/2022.deelio-1.10]
18. Haukoos JS, Lewis RJ. Advanced statistics: bootstrapping confidence intervals for statistics with "difficult" distributions. Acad Emerg Med 2005 Apr;12(4):360-365. [doi: 10.1197/j.aem.2004.11.018] [Medline: 15805329]
19. Wang J, et al. Evaluation and analysis of hallucination in large vision-language models. arXiv. Preprint posted online on Aug 29, 2023. [doi: 10.48550/ARXIV.2308.15126]
20. Archiv medizin. IMPP. URL: https://www.impp.de/pruefungen/medizin/archiv-medizin.html [accessed 2025-02-28]
21. Belinkov Y, Glass J. Analysis methods in neural language processing: a survey. Trans Assoc Comput Linguist 2019 Apr 1;7:49-72. [doi: 10.1162/tacl_a_00254]
22. Ryan L, Maharjan J, Mataraso S, et al. Predicting pulmonary embolism among hospitalized patients with machine learning algorithms. Pulm Circ 2022 Jan;12(1):e12013. [doi: 10.1002/pul2.12013] [Medline: 35506114]
23. Dua R, Wallace GR, Chotso T, Raj VFD. Classifying pulmonary embolism cases in chest CT scans using VGG16 and xgboost. In: Lecture Notes on Data Engineering and Communications Technologies: Springer; 2023, Vol. 131:273-292. [doi: 10.1007/978-981-19-1844-5_22]
24. Ding R, Ding Y, Zheng D, et al. Machine learning-based screening of risk factors and prediction of deep vein thrombosis and pulmonary embolism after hip arthroplasty. Clin Appl Thromb Hemost 2023;29:10760296231186145. [doi: 10.1177/10760296231186145] [Medline: 37394825]
25. Wu Q, et al. AutoGen: enabling next-gen LLM applications via multi-agent conversation. arXiv. Preprint posted online on Aug 16, 2023. [doi: 10.48550/arXiv.2308.08155]
26. Verdicchio M, Perin A. When doctors and AI interact: on human responsibility for artificial risks. Philos Technol 2022;35(1):11. [doi: 10.1007/s13347-022-00506-6] [Medline: 35223383]
27. Xu J. Overtrust of robots in high-risk scenarios. Presented at: AIES '18; Feb 2-3, 2018; New Orleans, LA, United States URL: https://dl.acm.org/doi/proceedings/10.1145/3278721 [accessed 2025-02-28] [doi: 10.1145/3278721.3278786]
28. Sankararaman KA, Wang S, Fang H. BayesFormer: transformer with uncertainty estimation. arXiv. Preprint posted online on Jun 2, 2022. [doi: 10.48550/arXiv.2206.00826]
29. Jiang W, Shi H, Yu L. Forward-backward reasoning in large language models for mathematical verification. Presented at: Findings of the Association for Computational Linguistics ACL 2024; Aug 11-16, 2024; Bangkok, Thailand URL: https://aclanthology.org/2024.findings-acl [accessed 2025-02-28] [doi: 10.18653/v1/2024.findings-acl.397]
30. Yao JY, Ning KP, Liu ZH, Ning MN, Yuan L. LLM lies: hallucinations are not bugs, but features as adversarial examples. arXiv. Preprint posted online on Oct 2, 2023. [doi: 10.48550/arXiv.2310.01469]
31. Bender EM, Gebru T, Mcmillan-Major A, Shmitchell S, Shmitchell SG. On the dangers of stochastic parrots: can language models be too big? Presented at: FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency; Mar 3-10, 2021. [doi: 10.1145/3442188.3445922]
32. Xie SM, Raghunathan A, Liang P, Ma T. An explanation of in-context learning as implicit bayesian inference. arXiv. Preprint posted online on Nov 3, 2021. [doi: 10.48550/arXiv.2111.02080]

XSL•FO
RenderX

33. Zhao TZ, Wallace E, Feng S, Klein D, Singh S. Calibrate before use: improving few-shot performance of language models. arXiv. Preprint posted online on Jul 1, 2021. [doi: 10.48550/arXiv.2102.09690]
34. Zheng C, Zhou H, Meng F, Zhou J, Huang M. Large language models are not robust multiple choice selectors. arXiv. Preprint posted online on Sep 7, 2023. [doi: 10.48550/ARXIV.2309.03882]
35. Jin X, Barbieri F, Kennedy B, Mostafazadeh Davani A, Neves L, Ren X. On transferability of bias mitigation effects in language model fine-tuning. Presented at: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics; Jun 6-11, 2021. [doi: 10.18653/v1/2021.naacl-main.296]
36. Chu T, Song Z, Yang C. Fine-tune language models to approximate unbiased in-context learning. arXiv. Preprint posted online on Oct 5, 2023. [doi: 10.48550/arXiv.2310.03331]
37. Regulation of the european parliament. European Commission. URL: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF [accessed 2025-01-15]
38. The Bletchley Declaration. GOV.UK. URL: https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023 [accessed 2025-02-28]
39. ANNEX I medical device regulation. Medical Device Regulation. URL: https://www.medical-device-regulation.eu/2019/07/23/annex-i-general-safety-and-performance-requirements/ [accessed 2025-02-28]
40. Regulating the machine. POLITICO. URL: https://www.politico.eu/article/regulate-europe-race-artificial-intelligence-ai-drugs-medicines/ [accessed 2025-02-28]

Abbreviations

AI: artificial intelligence
CA: confidence accuracy
EU: European Union
GPT4P: GPT-4 + plugin
GPT4PT: GPT-4 + plugin + translation
LLM: large language model
MDR: Medical Device Regulation

Edited by B Lesselroth; submitted 14.03.24; peer-reviewed by G Gill, T Nakao; revised version received 29.07.24; accepted 23.11.24; published 21.03.25.

Please cite as:
Madrid J, Diehl P, Selig M, Rolauffs B, Hans FP, Busch HJ, Scheef T, Benning L
Performance of Plug-In Augmented ChatGPT and Its Ability to Quantify Uncertainty: Simulation Study on the German Medical Board Examination
JMIR Med Educ 2025;11:e58375
URL: https://mededu.jmir.org/2025/1/e58375
doi:10.2196/58375

© Julian Madrid, Philipp Diehl, Mischa Selig, Bernd Rolauffs, Felix Patricius Hans, Hans-Jörg Busch, Tobias Scheef, Leo Benning. Originally published in JMIR Medical Education (https://mededu.jmir.org), 21.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evolution of Learning Styles in Surgery Comparing Residents and Teachers: Cross-Sectional Study

Gabriela Gouvea Silva, MD, MSc; Carlos Dario da Silva Costa, MD, MSc; Bruno Cardoso Gonçalves, MD, MSc; Luiz Vianney Saldanha Cidrão Nunes, MD; Emerson Roberto dos Santos, MSc; Natalia Almeida de Arnaldo Rodriguez Castro, MD, MSc; Alba Regina de Abreu Lima, MSc, PhD; Vânia Maria Sabadoto Brienze, MSc, PhD; Antônio Hélio Oliani, MD, MSc, PhD; Júlio César André, MD, MSc, PhD

Center for Studies and Development of Health Education, Faculdade de Medicina de São José do Rio Preto, Avenida Brigadeiro Faria Lima, 5416, São José do Rio Preto, Brazil

Corresponding Author:
Gabriela Gouvea Silva, MD, MSc
Center for Studies and Development of Health Education, Faculdade de Medicina de São José do Rio Preto, Avenida Brigadeiro Faria Lima, 5416, São José do Rio Preto, Brazil

Abstract

Background: Studies confirm a relationship between learning style and medical career choice in the learning style patterns observed in distinct types of residency programs. Such patterns can also be applied to general surgery, from medical school to the latest stages of training. Aligning teaching strategies with the predominant learning styles in surgical residency programs has the potential to make training more effective.

Objective: This study aimed to determine the learning styles of general surgery residents and professors in a Brazilian teaching hospital and compare the results with the existing literature.

Methods: This was a cross-sectional study conducted in a teaching hospital of a public university in Brazil. Thirty-four general surgery residents of any year of training and 30 professors participated in the study. Participants completed a sociodemographic survey and David Kolb's Learning Style Inventory. This was used to classify participants into one of four distinct types of learners: accommodating, diverging, assimilating, and converging. The relationship between sociodemographic data and learning styles was analyzed using the Fisher test, adjusted using the Bonferroni method, and the effect size was measured using the Cramer V test.

Results: The learning style distribution was similar in both groups, with 43,75% diverging, 42,18% accommodating, 10.93% assimilating, and 3.12% converging styles. A significant relationship was found between sex and learning style (P=.049) and between age and learning style for professors (P=.029). The effect sizes were strong (0.46) and very strong (0.506).

Conclusions: The prevalence of learning styles among general surgery residents and professors at this Brazilian hospital differs from that observed in previous studies, with more diverging and accommodating learners and fewer converging learners, suggesting a shift in learning styles. Understanding learning styles is important for effective surgical training programs. Further research with larger and more diverse populations is needed to confirm these results and explore the factors contributing to the observed differences in learning styles.

(JMIR Med Educ 2025;11:e64767) doi:10.2196/64767

KEYWORDS

learning; general surgery; medical education; internship and residency; surgeons; Brazil

Introduction

Background

The concept of learning styles was first developed at the beginning of 1960 as a result of the interest in individual differences while learning [1]. According to Dunn [2], everyone has a unique learning style, like a signature. In this prospect, adjusting the teaching to the different learning styles may help learners and improve educational outcomes. In the current literature, there are various models to determine the learning styles. There is a long and active discussion about whether learning styles are fixed or flexible, and to what extent the context can determine it [3]. Adapting learning styles can enhance student engagement, motivation, and academic performance [4]. Furthermore, the integration of technology and personalized learning approaches has shown promise in enhancing medical education [5].

To provide a more comprehensive understanding, learning styles can be defined as the cognitive, affective, and physiological traits that serve as relatively stable indicators of how learners

XSL•FO
RenderX

perceive, interact with, and respond to the learning environment [6]. Empirical evidence supports the existence of learning styles, demonstrating that individuals exhibit consistent preferences and strengths in how they approach learning tasks. For example, some learners may excel in visual tasks, while others thrive in auditory or kinesthetic activities [7].

The Kolb Experiential Learning Theory (ELT) is a prominent framework for understanding learning styles. The ELT posits that learning is a cyclical process involving four modes: concrete experience (CE), reflective observation (RO), abstract conceptualization (AC), and active experimentation (AE) [8]. Individuals develop preferences for certain modes, leading to four distinct learning styles: converging (AC/AE), diverging (CE/RO), assimilating (AC/RO), and accommodating (CE/AE). The Kolb Learning Style Inventory (LSI) is a widely used tool for assessing these preferences. The validity of Kolb's work in the context of medical education has been demonstrated in numerous studies [9-13], which have found that medical students and professionals exhibit distinct learning style preferences that can influence their academic and professional performance.

Knowledge is the main domain of medical education, but outcomes strongly depend on other domains such as attitude, lifelong learning, and empathy; in surgery, some domains are central including resilience, craftsmanship, and decision-making, among other domains [14].

Research Gap and Problem Statement

Despite the established importance of learning styles in education, limited research has specifically examined their prevalence and impact within general surgery residency programs, particularly in diverse cultural and geographical settings. The clinical and surgical environments present unique challenges for both trainees and educators, requiring the development of complex skills and behaviors [14]. Understanding how surgical residents learn is crucial for optimizing the training process and ensuring the development of competent and well-rounded surgeons.

Moreover, current surgical trainees come from diverse educational, cultural, ethnic, and gender backgrounds, and personal factors also influence their learning characteristics [15]. Little is known about the teaching and learning preferences among surgeons and how they influence the effectiveness of training [16]. Addressing this gap in knowledge is essential for designing effective and inclusive surgical training programs that cater to the diverse needs of learners. To this end, simulation-based surgical training has emerged as a valuable tool for enhancing technical skills and improving patient outcomes [17].

Research Aims and Objectives

Little is known about the teaching and learning preferences among surgeons and how they influence the effectiveness of training [16]. Despite its relevance, studies investigating learning styles in the context of general surgery residency are scarce, especially in countries outside North America and Europe.

Therefore, to address the gaps in understanding learning styles in general surgery, particularly in diverse cultural and geographical settings, this study aims to (1) determine the learning styles of general surgery residents and professors in a Brazilian teaching hospital; (2) compare these findings with existing literature on learning styles in surgery; and (3) discuss the implications of these findings for surgical training programs.

By providing data from a Brazilian teaching hospital, we aim to contribute to a more comprehensive understanding of learning styles in surgical training and inform the development of more effective and inclusive surgical education strategies. This knowledge can inform the development of more effective and inclusive surgical education strategies, ultimately leading to better-prepared and more competent surgeons.

Methods

Study Design and Setting

This cross-sectional study was conducted in 2022 at the Hospital de Base de São José do Rio Preto, a teaching hospital affiliated with Faculdade de Medicina de São José do Rio Preto (a public university in São Paulo, Brazil).

Participants and Recruitment

The study population consisted of general surgery residents in any year of training and hospital professors. All participants were over 18 years old and signed the free and informed consent form.

Data Collection

Data collection involved two instruments: a sociodemographic survey and David Kolb's LSI. The sociodemographic survey collected information on participants' age and sex, and years of residency (for residents) or teaching experience (for professors). The LSI is a validated tool that consists of 12 questions, each with four statements that the participants ranked from 1 to 4 according to their learning preferences. The LSI tool classifies the participants into one of four types of learners based on Kolb's learning cycle: (1) accommodating (learn primarily by experience), (2) diverging (learn by RO), (3) assimilating (learn by exploring associations and interrelationships), and (4) converging (learn by doing or trying things with practical results) [18].

The LSI test was administered in a controlled environment, with a researcher present to provide instructions and clarify any doubts. Participants had 30 minutes to complete it. The sociodemographic survey was completed immediately after completing the LSI test.

Statistical Analysis

Software

Data analysis was performed using the Statistical Package for the Social Sciences (SPSS), version 26.0 (IBM Corp.).

Normality Check

Due to the relatively small sample size and the nature of the data, the Shapiro-Wilk test was used to assess the normality of continuous variables (age and years of experience).

XSL•FO
RenderX

Statistical Tests

A P value <.05 was considered statistically significant. The relationship between data was calculated using the Fisher test, adjusted by the Bonferroni method [19]. The Fisher exact test was chosen due to the small sample size and the presence of categories with expected frequencies lower than 5 [20]. The size effect was measured using Cramer V test, which indicates the grade of association between variables: the result is stronger as it approaches the value of 1 [21].

Power

The sample size was calculated using the formula for finite populations, considering a confidence level of 95%, a margin of error of 5%, and an expected prevalence of 50% for each learning style. The minimum sample size was 67 participants, and the total number of residents and professors was 80 [22].

Data Exclusion

Questionnaires that were responded to incorrectly according to Kolb's rules were discarded.

Ethical Considerations

The study was approved by the Research Ethics Committee of Faculdade de Medicina de São José do Rio Preto (approval number: 12345/2022). All participants signed the free informed consent form. Data were anonymized.

Recruitment

This study is grounded in Kolb's ELT, which posits that learning is a cyclical process involving four modes: CE, RO, AC, and AE [23]. Individuals develop preferences for certain modes, leading to four distinct learning styles:

- Converging: Individuals with this learning style excel in AC and AE. They are practical, enjoy problem-solving, and are skilled at applying theories to real-world situations.
- Diverging: Individuals with this learning style excel in concrete CE and RO. They are imaginative, enjoy brainstorming, and are skilled at generating ideas.
- Assimilating: Individuals with this learning style excel in AC and RO. They are logical, enjoy analyzing data, and are skilled at creating models and theories.
- Accommodating: Individuals with this learning style excel in concrete CE and AE. They are hands-on, enjoy taking risks, and are skilled at implementing plans and getting things done.

Our logic model is based on the premise that aligning teaching strategies with the predominant learning styles of surgical residents and professors can enhance the effectiveness of surgical training. We hypothesize that by identifying the learning styles of our participants and tailoring instructional approaches accordingly, we can improve learning outcomes and promote a more engaging and inclusive learning environment.

Table 1 provides a more detailed overview of the four learning styles, including concrete examples of learning activities and instructional approaches that are best suited for each style.

Table . Learning styles, characteristics, and instructional approaches. Source: [8].

Learning style	Characteristics	Example learning activities	Example instructional approaches
Converging	Practical, problem-solver, applies theories	Simulation-based training, case studies	Problem-based learning, hands-on workshops
Diverging	Imaginative, brainstormer, generates ideas	Group discussions, reflective writing	Mentoring, collaborative projects
Assimilating	Logical, analytical, creates models	Literature reviews, data analysis	Lectures, seminars
Accommodating	Hands-on, risk-taker, implements plans	Surgical procedures, clinical rotations	Apprenticeship, on-the-job training

All general surgery residents were invited to answer printed free and informed consent form and the LSI's test, in person. The same was done with the faculty members. The questionnaires were then collected and transformed into digital archives, processed in digital tables after codification.

Statistical Analysis

Power

The sample size was calculated using the formula for finite populations, considering a confidence level of 95%, a margin of error of 5%, and an expected prevalence of 50% for each learning style. The minimum sample size was 67 participants, and the total number of residents and professors was 80 [22].

Results

A total of 64 participants (34 residents and 30 professors) were included in this study. The sociodemographic characteristics of the participants are presented in Table 2. Among the 34 residents, 18 (52.9%) were male and 16 (47.1%) were female. Most residents (91.2%, 31/34) were under 30 years of age. Among the 30 professors, 24 (80%) were male, and 6 (20%) were female, and most of them (17/30, 56.7%) were between 40 and 70 years of age. All professors graduated from universities when traditional teaching methods (ie, primarily lecture-based instruction with limited student interaction) were used, whereas 47% of the residents graduated from universities that used active or mixed teaching methods (ie, incorporating strategies such as problem-based learning, group work, and case studies to promote student engagement).

XSL•FO
RenderX

Table . Sociodemographic characteristics of participants.

Characteristics	Residents (n=34) N (%)	Professors (n=30) N (%)
Age (years)		
<30	31 (91.2)	2 (6.7)
30 - 39	3 (8.8)	11 (36.7)
40 - 70	0 (0)	17 (56.7)
Sex		
Male	18 (52.9)	24 (80)
Female	16 (47.1)	6 (20)
Teaching method used at the University of origin		
Traditional	18 (52.9)	30 (100)
Active or mixed	16 (47.1)	0 (0)

The distribution of Kolb's learning styles is presented in Table 3 and Multimedia Appendix 1. The most prevalent learning styles were diverging (18/34) in the residents' group and accommodating (17/30) in the professors' group.

Table . Learning styles among surgery groups.

Learning styles	Residents N (%)	Professors N (%)	Total N (%)
Converging	1/34 (2.94)	1/30 (3.33)	2/64 (3.12)
Assimilating	5/34 (14.7)	2/30 (6.7)	7/64 (10.93)
Accommodating	10/34 (29.4)	17/30 (56.7)	27/64 (42.18)
Diverging	18/34 (52.9)	10/30 (33.3)	28/64 (43.75)

The relationship between sociodemographic data and learning styles was analyzed using the Fisher exact test (Table 4). A significant association was found between sex and learning style (*P*=.049; Cramer V=0.46), indicating a strong effect size. However, determining which specific categories were significantly different using the Bonferroni post-hoc test was not possible. Among professors, a significant relationship was observed between age and learning style (*P*=.029; Cramer V=0.506), suggesting a very strong effect size. However, specific age groups that differed significantly could not be identified with the Bonferroni post-hoc test, possibly due to the small sample size.

Table . Relationship between sociodemographic data and learning styles.

Variables	*P* value (Fisher exact test)	Effect sizes (Cramer V)
Sex	0.049[a]	0.46 (strong)
Age (residents)	0.999	0.12 (weak)
Age (professors)	0.029[a]	0.506 (very strong)
Teaching method used at the university of origin (residents)	0.999	0.08 (weak)

[a]Statistically significant at P<.05

Discussion

Principal Findings

Our study, utilizing Kolb's LSI, identified the distribution of four learning styles among general surgery residents and professors at a Brazilian teaching hospital. These learning styles are:

- Diverging: Learners who excel in CE and RO, are imaginative, and generate ideas effectively.
- Accommodating: Learners who excel in CE and AE, are hands-on, and enjoy implementing plans.
- Assimilating: Learners who excel in AC and RO, are logical, and create models and theories.
- Converging: Learners who excel in AC and AE, are practical, and apply theories to real-world situations.

The most prevalent learning styles were diverging (52.9%) in the residents' group and accommodating (56.7%) in the professors' group (Table 2 and Multimedia Appendix 1). A significant association was found between sex and learning style

XSL•FO
RenderX

(*P*=.049; Cramer V=0.46), indicating a strong effect size. Among professors, a significant relationship was observed between age and learning style (*P*=0.029; Cramer V=0.506), suggesting a very strong effect size.

Table 2 and Multimedia Appendix 1 show that while diverging was the most common style among residents and accommodating was most common among professors, the overall learning style distribution was relatively similar between the two groups. This convergence, where both residents and professors exhibit a blend of diverging and accommodating tendencies, can potentially facilitate both teaching and learning [16]. The shared presence of these styles suggests that both groups may value CEs and RO (diverging) as well as hands-on activities and practical application (accommodating).

This alignment can be leveraged to support instruction in different ways. For diverging learners (both residents and some professors), emphasize group discussions, brainstorming sessions, and reflective writing assignments. Encourage the sharing of diverse perspectives and the exploration of different approaches to surgical problems. In contrast, for accommodating learners (both professors and some residents), Provide opportunities for hands-on practice, simulation-based training, and real-world clinical rotations. Encourage AE and problem-solving in practical settings. By incorporating these strategies, educators can create a learning environment that caters to the predominant learning styles of both residents and professors, fostering more effective communication, engagement, and knowledge acquisition.

However, it is important to acknowledge that the similarity in distribution does not guarantee a perfect match for all individuals. The relatively lower prevalence of converging and assimilating styles in both groups suggests that those learners might require more tailored support and learning opportunities to ensure their needs are met. This underscores the importance of mapping learning styles when designing a comprehensive residency program, as it provides a basis for guiding the learning needs of all residents and professors, not just the majority.

Implications of Findings

The findings of our study have important implications for surgical education. Understanding the predominant learning styles of residents and professors can help adapt teaching strategies and curriculum design to better meet their needs. For example, incorporating more RO and practical experiences can benefit diverging and accommodating learners while also providing opportunities for AC and AE to support assimilating and converging learners.

Furthermore, with the occurrence of the pandemic, the increased distances imposed by contact restrictions have further deepened these changes. The COVID-19 pandemic has also presented unique challenges to surgical training, with restrictions on in-person learning and clinical experiences. A pan-Romanian survey by Moldovan et al [24] highlighted the impact of the pandemic on orthopedic residents, including reduced surgical volume, limited access to educational resources, and increased stress and anxiety. These challenges may have further influenced the learning styles and preferences of surgical residents and professors during this period.

Comparison With Prior Work

Few studies on learning styles in surgery were found in the literature, but we can state that our results are different from previous results.

In the 1980s, Baker III et al [25] reported a prevalence of converging (46%), followed by accommodating (26%) and assimilating (20%) styles among surgeons. In the 2000s, this pattern was confirmed by Contessa et al [26]. They argued that surgical practice requires quick decision-making and problem-resolution, justifying the converging style and its more pragmatic view. In 2007, Mammen et al [27] published similar results obtained in the US population.

After Quillin [28] reduced his working hours in general surgery residency, he showed the results collected from 1999 to 2012. At that time, the proportion of accommodating learners was higher, especially after 2003, when the workload was reduced [28].

In 2017, for the first time, diverging learners became the majority in a study with 47 surgeons in the United Kingdom [29]. In 2018, also in the United Kingdom, a study with residents in various surgical areas found that converging, followed by accommodating styles were predominant [30]. In 2020, similar results were published in Scotland by Hopkins et al [15]. The most recent publication on the topic reported a predominance of assimilating followed by converging styles in Spain [31]. Table 5 and Multimedia Appendix 2 show the existing literature.

Table . Data of learning styles in surgery through time around the world.

Author	Publication year	Country	Population (n)	Diverging	Accomodator	Assimilating	Converging
Baker III	1985	USA	Surgeons (39)	8%	26%	20%	46%
Drew	1999	UK	Basic surgical trainees (52)	3.9%	27%	9.6%	59.5%
Mammen	2007	USA	General surgery residents (91)	10.6%	14.6%	17.2%	57.8%
Brown	2018	UK	Medical students (60)	20.8%	30.2%	17%	32%
Parra	2021	Spain	Surgical residents and staff (64)	14.1%	21.9%	39.1%	25%

XSL•FO
RenderX

The results of the present study were diverging (43,8%), accommodating (42,2%), assimilating (11,0%), and converging (3,12%) styles. These results amplify the existing literature, showing an increase in diverging and a decrease in converging styles over the years. These findings indicate a shift in the learning preferences of surgical residents and professors, which may have been influenced by various factors, such as changes in surgical education, technological advancements, and sociocultural aspects.

The geographical location may be a possible explanation for our results, as previous studies were conducted in North America and Europe (Multimedia Appendix 1). Cultural differences and variations in surgical training programs across countries may have contributed to the observed differences in learning styles. Another hypothesis may be the course of time: the last two decades have seen huge technological changes, when social media, smartphones, and laptops became widely available, greatly impacting the teaching-learning process [32]. Recent studies have further explored the impact of digital technologies on medical education, highlighting both the opportunities and challenges associated with their integration [5]. Furthermore, with the occurrence of the pandemic, the increased distances imposed by contact restrictions have further deepened these changes [33].

The differing proportions of female residents (47.1%) and professors (16.0%) highlight the ongoing evolution of gender representation in surgery. The historical underrepresentation of women in surgical fields may contribute to differences in observed learning styles between residents and professors [34]. Despite this, the increasing participation of women in surgery over recent decades is a positive trend [22].

Strengths and Limitations

The population included is a small sample of a larger Brazilian surgical group. More data can be further collected to compare the country with other nations, in America, Europe or even Asia. The medical reality in Brazil is diverse and worth a broader approach.

In addition to the small sample size, our study has several other limitations that should be acknowledged. The study had a sampling bias. Our sample was drawn from a single teaching hospital in Brazil, which may not be representative of all general surgery residents and professors in Brazil or other countries. This limits the generalizability of our findings. Additionally, the voluntary nature of participation may have introduced selection bias, as those who chose to participate may differ systematically from those who did not. In addition, there was measurement bias; the Kolb LSI is a self-report instrument, which is subject to social desirability bias and response bias. Participants may have answered the questions in a way that they perceived as more favorable or aligned with societal expectations, rather than reflecting their true learning preferences. Moreover, our study did not fully explore the potential influence of various sociodemographic factors, such as cultural background, socioeconomic status, and prior educational experiences, on learning styles. However, we did not collect data on other potentially relevant sociodemographic factors such as ethnicity, social class, migration background, or detailed information about prior educational experiences. These factors may interact with learning styles in complex ways and could have influenced our results.Finally, the cross-sectional design of our study limits our ability to draw causal inferences about the relationship between learning styles and other variables. A longitudinal study would be needed to examine how learning styles evolve over time and how they impact training outcomes. These limitations should be considered when interpreting our findings.

Further research is needed to explore the underlying factors that influence these learning styles, such as personality traits, prior educational experiences, and cultural background. Understanding these factors could allow for more tailored interventions to optimize learning. Moreover, future studies should investigate the potential impact of different learning styles on surgical performance metrics, such as technical skill acquisition, error rates, and patient outcomes. This would provide valuable insights into how learning style preferences translate into real-world surgical practice.

Conclusions

This study found that diverging and accommodating learning styles were more prevalent among general surgery residents and professors in a Brazilian university hospital, differing from previous North American and European studies. The decreased prevalence of the converging style is notable and may be due to changes in surgical education, technology, and cultural differences. Understanding these learning styles can guide more effective and inclusive teaching strategies in surgical residency programs. Further research with larger, diverse populations is needed to explore the relationships between learning styles, demographics, and training outcomes.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Learning styles prevalent in surgery groups.
[PNG File, 56 KB - mededu_v11i1e64767_app1.png]

Multimedia Appendix 2
Timeline of surgical learning styles according to the existent literature.

XSL•FO
RenderX

[PNG File, 72 KB - mededu_v11i1e64767_app2.png]

References

1. Curry L. An organization of learning styles theory and constructs. 1983 URL: https://files.eric.ed.gov/fulltext/ED235185.pdf [accessed 2025-05-05]
2. Dunn R. Understanding the dunn and dunn learning styles model and the need for individual diagnosis and prescription. Journal of Reading, Writing, and Learning Disabilities International 1990 Jan;6(3):223-247 [FREE Full text] [doi: 10.1080/0748763900060303]
3. Bernacki ML, Greene MJ, Lobczowski NG. A systematic review of research on personalized learning: personalized by whom, to what, how, and for what purpose(s)? Educ Psychol Rev 2021 Dec;33(4):1675-1715. [doi: 10.1007/s10648-021-09615-8]
4. El-Sabagh HA. Adaptive e-learning environment based on learning styles and its impact on development students' engagement. Int J Educ Technol High Educ 2021 Dec;18(1):53. [doi: 10.1186/s41239-021-00289-4]
5. Chowdhury PN, Vaish A, Puri B, Vaishya R. Medical education technology: past, present and future. Apollo Medicine 2024 Dec;21(4):374-380. [doi: 10.1177/09760016241256202]
6. (U.S.) NA of SSP. student learning styles: diagnosing and prescribing programs. National Association of Secondary School Principals. 1979. URL: https://books.google.com.br/books?id=PFYmAQAAIAAJ [accessed 2025-04-30]
7. Felder RM, Silverman LK. Learning and teaching styles in engineering education. 1988. URL: https://api.semanticscholar.org/CorpusID:140475379 [accessed 2025-04-30]
8. Kolb D. Experiential learning: experience as the source of learning and development. In: J Bus Ethics 1984, Vol. 1. URL: https://www.researchgate.net/publication/235701029_Experiential_Learning_Experience_As_The_Source_Of_Learning_And_Development [accessed 2025-05-05]
9. Villanueva F. The Use of Kolb's Learning Styles Inventory (LSI) in School Settings: BU Journal of Graduate Studies in Education, Vol. 12. URL: https://files.eric.ed.gov/fulltext/EJ1263067.pdf [accessed 2025-05-05]
10. Shakeri F, Ghazanfarpour M, MalaKoti N, Soleimani Houni M, Rajabzadeh Z, Saadat S. Learning Styles of Medical Students: A Systematic Review. Med Edu Bull 2022;3(2):441-456. [doi: 10.22034/MEB.2022.328652.1050]
11. Vemuri VR, Rao KA. Assessment of Learning Styles Using Kolb's Learning Style Inventory among Medical College Students: A Cross-sectional Study. J Clin Diagn Res . [doi: 10.7860/JCDR/2024/67440.18987]
12. Cortés B, Olaya G, Olaya M, Fabricio J. Estilos de aprendizaje en estudiantes de medicina [Article in Spanish]. Universitas Médica 2018;59(2):1-10.
13. Vemuri VR, Rao KA. J Clin of Diagn Res 2024;18(2):JC01-JC04 [FREE Full text]
14. Palejwala Z, Wallman KE, Maloney S, et al. Higher operating theatre temperature during burn surgery increases physiological heat strain, subjective workload, and fatigue of surgical staff. PLoS ONE 2023;18(6):e0286746. [doi: 10.1371/journal.pone.0286746] [Medline: 37267345]
15. Hopkins L, Robinson D, James O, Egan R, Lewis W, Bailey D. Surgical learning styles; a gender gap perspective. 2020. URL: https://www.webofscience.com/wos/woscc/full-record/WOS:000595958400223 [accessed 2025-05-05]
16. Dickinson KJ, Bass BL, Graviss EA, Nguyen DT, Pei KY. How learning preferences and teaching styles influence effectiveness of surgical educators. Am J Surg 2021 Feb;221(2):256-260. [doi: 10.1016/j.amjsurg.2020.08.028] [Medline: 32921405]
17. Foppiani J, Stanek K, Alvarez AH, et al. Merits of simulation-based education: a systematic review and meta-analysis. J Plast Reconstr Aesthet Surg 2024 Mar;90:227-239. [doi: 10.1016/j.bjps.2024.01.021] [Medline: 38387420]
18. Romanelli F, Bird E, Ryan M. Learning styles: a review of theory, application, and best practices. Am J Pharm Educ 2009 Feb 19;73(1):9. [doi: 10.5688/aj730109] [Medline: 19513146]
19. Argyrous G. The chi-square test for independence. In: Statistics for Social Research: Macmillan Education UK; 1997:257-284. [doi: 10.1007/978-1-349-14777-9_16]
20. Kim HY. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. Restor Dent Endod 2017 May;42(2):152-155. [doi: 10.5395/rde.2017.42.2.152] [Medline: 28503482]
21. Dias D, Silva JS, Bernardino A. The prediction of road-accident risk through data mining: a case study from Setubal, Portugal. Informatics (MDPI) 2023;10(1):17. [doi: 10.3390/informatics10010017]
22. Xepoleas MD, Munabi NCO, Auslander A, Magee WP, Yao CA. The experiences of female surgeons around the world: a scoping review. Hum Resour Health 2020 Oct 28;18(1):80. [doi: 10.1186/s12960-020-00526-3] [Medline: 33115509]
23. Kolb AY, Kolb DA. Learning styles and learning spaces: enhancing experiential learning in higher education. In: Academy of Management Learning and Education 2005, Vol. 4:193-212. [doi: 10.5465/amle.2005.17268566]
24. Moldovan F, Gligor A, Moldovan L, Bataga T. The impact of the COVID-19 pandemic on the orthopedic residents: a pan-romanian survey. Int J Environ Res Public Health 2022 Jul 27;19(15):9176. [doi: 10.3390/ijerph19159176] [Medline: 35954536]
25. Baker JD, Reines HD, Wallace CT. Learning style analysis in surgical training. Am Surg 1985 Sep;51(9):494-496. [doi: 10.5688/aj730109] [Medline: 4037546]

XSL•FO
RenderX

26. Contessa J, Ciardiello KA, Perlman S. Surgery resident learning styles and academic achievement. Curr Surg 2005;62(3):344-347. [doi: 10.1016/j.cursur.2004.09.012] [Medline: 15890222]
27. Mammen JMV, Fischer DR, Anderson A, et al. Learning styles vary among general surgery residents: analysis of 12 years of data. J Surg Educ 2007;64(6):386-389. [doi: 10.1016/j.jsurg.2007.08.005] [Medline: 18063274]
28. Quillin RC, Pritts TA, Young GB, Edwards MJ, Davis BR. Surgical resident learning styles have changed following the implementation of the 80-hour workweek. Journal of Surgical Research 2013 Feb;179(2):230. [doi: 10.1016/j.jss.2012.10.420] [Medline: 23721929]
29. Challapalli P. Evaluation of learning styles in a surgical unit. Abtracts from the Medical Research Symposium for Students and Foundation Doctors. 2017. URL: https://www.rcpe.ac.uk/sites/default/files/medical_research_symposium_abstracts_8_feb.pdf [accessed 2025-05-01]
30. Brown C, Luton OW, Robinson D, et al. Learning style variation: a British core surgical trainee perspective. J Am Coll Surg 2018;227(4):S221. [doi: 10.1016/j.jamcollsurg.2018.07.484]
31. Martín Parra JI, Toledo Martínez E, Martínez Pérez P, et al. Analysis of learning styles in a laparoscopic technical skills course. Implications for surgical training. Cirugía Española (English Edition) 2021 Dec;99(10):730-736. [doi: 10.1016/j.cireng.2021.10.011]
32. Frenk J, Chen LC, Chandran L, et al. Challenges and opportunities for educating health professionals after the COVID-19 pandemic. Lancet 2022 Oct 29;400(10362):1539-1556. [doi: 10.1016/S0140-6736(22)02092-X] [Medline: 36522209]
33. Trott M, Driscoll R, Irlado E, Pardhan S. Changes and correlates of screen time in adults and children during the COVID-19 pandemic: a systematic review and meta-analysis. EClinicalMedicine 2022 Jun;48:101452. [doi: 10.1016/j.eclinm.2022.101452] [Medline: 35615691]
34. Motter SB, Brandão GR, Iaroseski J, et al. Women representation in academic and leadership positions in surgery in Brazil. Am J Surg 2022 Jan;223(1):71-75. [doi: 10.1016/j.amjsurg.2021.07.023] [Medline: 34315578]

Abbreviations

AC: abstract conceptualization
AE: active experimentation
CE: concrete experience
ELT: Experiential Learning Theory
LSI: learning style inventory
RO: reflective observation

Edited by B Lesselroth; submitted 25.07.24; peer-reviewed by F Moldovan, J Olivier, M Jani; revised version received 11.03.25; accepted 06.04.25; published 08.05.25.

Please cite as:
Gouvea Silva G, da Silva Costa CD, Cardoso Gonçalves B, Vianney Saldanha Cidrão Nunes L, Roberto dos Santos E, Almeida de Arnaldo Rodriguez Castro N, de Abreu Lima AR, Sabadoto Brienze VM, Oliani AH, André JC
Evolution of Learning Styles in Surgery Comparing Residents and Teachers: Cross-Sectional Study
JMIR Med Educ 2025;11:e64767
URL: https://mededu.jmir.org/2025/1/e64767
doi:10.2196/64767

© Gabriela Gouvea Silva, Carlos Dario da Silva Costa, Bruno Cardoso Gonçalves, Luiz Vianney Saldanha Cidrão Nunes, Emerson Santos, Natalia Almeida de Arnaldo Rodriguez Castro, Alba Regina Abreu de Lima, Vânia Maria Sabadoto Brienze, Antônio Hélio Oliani, Júlio César André. Originally published in JMIR Medical Education (https://mededu.jmir.org), 8.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Faculty Perceptions on the Roles of Mentoring, Advising, and Coaching in an Anesthesiology Residency Program: Mixed Methods Study

Sydney Nykiel-Bailey[1*], DO; Kathryn Burrows[2*], PhD; Bianca E Szafarowicz[1*], DO; Rachel Moquin[1*], EdD, MA

[1]Department of Anesthesiology, Washington University School of Medicine, 660 S Euclid Avenue, Saint Louis, MO, United States

[2]National Coalition of Independent Scholars, Independent Scholar 432 Division, Oregan City, OR, United States

[*]all authors contributed equally

Corresponding Author:
Sydney Nykiel-Bailey, DO
Department of Anesthesiology, Washington University School of Medicine, 660 S Euclid Avenue, Saint Louis, MO, United States

Abstract

Background: Mentoring, advising, and coaching are essential components of resident education and professional development. Despite their importance, there is limited literature exploring how anesthesiology faculty perceive these practices and their role in supporting residents.

Objective: This study aims to investigate anesthesiology faculty perspectives on the significance, implantation strategies, and challenges associated with mentorship, advising, and coaching in resident education.

Methods: A comprehensive survey was administrated to 93 anesthesiology faculty members at Washington University School of Medicine. The survey incorporated quantitative Likert-scale questions and qualitative short-answer responses to assess faculty perceptions of the value, preferred formats, essential skills, and capacity for fulfilling multiple roles in these support practices. Additional areas of focus included the impact of staffing shortages, training requirements, and the potential of these practices to enhance faculty recruitment and retention.

Results: The response rate was 44% (n=41). Mentoring was identified as the most important aspect, with 88% (n=36) of faculty respondents indicating its significance, followed by coaching, which was highlighted by 78% (n=32) of respondents. The majority felt 1 faculty member can effectively hold multiple roles for a given trainee. The respondents desired additional training for roles and found roles to be rewarding. All roles were seen as facilitating recruitment and retention. Barriers included faculty burnout; confusion between roles; time constraints; and desire for specialized training, especially in coaching skills.

Conclusions: Implementing structured mentoring, advising, and coaching can profoundly impact resident education but requires role clarity, protected time, culture change, leadership buy-in, and faculty development. Targeted training and operational investments could enable programs to actualize immense benefits from high-quality resident support modalities. Respondents emphasized that resident needs evolve over time, necessitating flexibility in appropriate faculty guidance. While coaching demands unique skills, advising hinges on expertise and mentoring depends on relationship-building. Systematic frameworks of coaching, mentoring, and advising programs could unlock immense potential. However, realizing this vision demands surmounting barriers such as burnout, productivity pressures, confusion about logistics, and culture change. Ultimately, prioritizing resident support through high-quality personalized guidance can recenter graduate medical education.

(JMIR Med Educ 2025;11:e60255) doi:10.2196/60255

KEYWORDS

coaching; faculty perceptions; mentoring; perception; medical education; anesthesia; modality; support; Washington University; university; coaching skills; training; culture change; culture; flexibility; systematic framework

Introduction

The current landscape of medical education is influenced by both medical culture and shifting demographics among learners. Factors such as medical provider burnout [1], a nationwide shortage of medical staff [2], and the evolving characteristics of different generations of learners are reshaping medical education [3]. It is imperative that the well-being and guidance of learners, both personally and professionally, are recentralized as the core of medical education. Emphasizing principles such as advising, mentoring, and coaching is crucial to support learners in their journey toward academic and personal fulfillment. These principles should be thoroughly examined

XSL•FO
RenderX

and reevaluated to empower learners to pursue paths of academic and personal success, foster self-assessment, ensure a nurturing learning environment, and encourage a commitment to lifelong learning [1,2]. The objective of this paper is to examine the attitudes and experiences of clinical-academic anesthesiology faculty with respect to their understanding and practice of mentoring, advising, and coaching. Our aim is to identify key themes that more clearly define these roles within medical education, as well as to elucidate potential barriers to their implementation and sustainability. Furthermore, we seek to understand faculty perspectives on the need for formalized educational support in these areas. We anticipate that the insights gained from this study could be broadly applicable across the graduate medical education spectrum, particularly as the focus in education increasingly shifts toward professionalism and well-being.

The education and welfare of medical residents hinge upon a multifaceted network of connections. Residents at different stages of their training will necessitate varying forms of engagement: mentoring, advising, or coaching. While these 3 avenues are distinct, they all share the common aim of nurturing education, wellness, and career progression [2,3]. Each approach serves its unique purpose and uses diverse methodologies [2]. Identifying the most suitable modality for the learner is paramount. Facilitators must adeptly involve themselves and customize sessions to ensure that expectations and objectives resonate with the learner [2].

Traditionally, mentoring has been the primary means of providing guidance [4]. It entails a sustained personal relationship between mentor and mentee, with the learner's overarching aspirations guiding the interaction. Conversations, career mapping, and counsel are derived from the mentor's experiences and expertise [2,3]. Typically, mentors possess knowledge in the pertinent field and share their insights with the learner. The mentor guides sessions, posing direct questions with long-term goals as the focal point. In residency education, mentoring often follows a structured format, though informal mentorships may naturally evolve. Institutions may request mentors to provide feedback or document these sessions for accreditation purposes [2,3].

Advising typically comprises a single, informal session focused on a specific issue or inquiry. The advisor leads the session and provides solutions or strategies based on their own experiences. The learner has the autonomy to decide whether to heed the advice. Unlike mentoring, a sustained relationship is not necessarily a prerequisite for advising, and subsequent follow-up is usually with independence and self-driven by the needs of the advisee [5]. Advisors may possess limited insight into the learner's personal or academic strengths and weaknesses, resulting in advice limited to specific scenarios [6].

Academic coaching differs from advising and mentoring in that it prioritizes the learner's agency. Coaches refrain from offering advice or engaging in decision-making. Instead, their role is to facilitate self-discovery and create a supportive atmosphere for self-assessment and future planning [2]. Coaches assist learners in identifying actions that may lead to success or failure. Unlike mentors and advisors, coaches may not necessarily possess expertise in the medical field. Coach engagement is supported by actively listening to the learner and offering questions to encourage self-awareness. Coaching fosters a consistent, enduring relationship characterized by an educational partnership between coach and learner [2].

No single form of guidance is adequate to meet the needs of today's students, and students' needs evolve as they move through residency [7]. Faculty must be facile in their ability to intuit what type of guidance is appropriate for a specific student or situation, and be able to provide that guidance or refer the student to someone who can [8]. For this reason, faculty development programs play a crucial role in supporting faculty as they rise to meet the challenges of guiding trainees, and faculty training in these support modalities may be lacking [9]. Training educators on how to target student needs by using the most effective guidance strategy will help decrease role confusion [8]. Training and developing faculty in advising, mentoring, and coaching help cultivate an ongoing culture of scholarship [10] and can help faculty navigate the competing challenges of their clinical and nonclinical roles [11]. Faculty report that lack of support from leadership and lack of proper training are barriers to their role as advisors, coaches, and mentors [11], and training and assessment tools for faculty members are crucial [7,9].

Methods

Study Design

A survey (Multimedia Appendix 1) was sent to 93 Washington University School of Medicine Anesthesiology clinical educator faculties. This target population was used as a convenience sample, representing a cohesive cohort with consistent interactions with trainees. This survey was developed based on core competencies and conceptual differentiations outlined for the roles of advisors, coaches, and mentors in medical education [5,6,8,9]. Drawing from Wolff et al [9], support modality definitions and key characteristics were designed to reflect critical distinctions regarding focus, relationship context, longevity, skill sets, and objective alignment [9]. Survey questions were formulated to assess physician perspectives across these theoretical domains for each resident support role.

A group of coaching experts within the Department of Anesthesiology was selected to create a novel survey tool. To facilitate the design and construction of the survey instrument, the research team used a modified Delphi technique, a widely recognized method for achieving consensus among experts. A subset of academic faculty was invited to participate in a pilot study aimed at testing multiple dimensions of the survey's implementation. This pilot study served several purposes: (1) to ensure the clarity and comprehensibility of the survey questions, (2) to evaluate the technical functionality of the survey platform, and (3) to assess the feasibility of applying inductive thematic analysis to the pilot data. Through iterative revisions and rounds of expert feedback, the survey underwent several modifications to enhance both face validity and content validity. The final version of the survey, which reflects the culmination of this rigorous development process, is presented in Multimedia Appendix 1. The survey is composed of 2 Likert

XSL•FO
RenderX

5-point scale quantitative items and 11 qualitative open-ended questions.

Quantitative items examined perceptions of importance and optimal configurations applying the principles of Wolff et al [9] regarding situational demands and need for role clarity [9]. Quantitative data were collected using the REDCap (Research Electronic Data Capture) Consortium platform (Vanderbilt University), a secure web-based application designed to support data capture for research studies. Faculty received the voluntary survey through department email, no incentives offered, and faculty log-in prevented duplicate entries. Data were analyzed using descriptive statistics. Qualitative questions elicited feedback on specialized skills, training interests, and implementation barriers grounded in advising, coaching, and mentoring competency frameworks [5-9]. The sequence of survey topics reflects established theory comparing and contrasting these support avenues [6-8]. An inductive qualitative analysis was conducted, using the Braun and Clarke [12] 6-phase approach to thematic analysis. This methodological framework, widely used in qualitative research, ensures both the flexibility and rigor required for the interpretative analysis of complex datasets. The 6 stages—familiarization with data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and writing up—provide a structured yet adaptable framework for data interpretation [12]. Qualitative data were collected through open-response questions included in the REDCap survey. The text from these open-response questions was analyzed using the Dedoose coding themes platform.

The process begins with open coding to identify initial patterns within the data. Codes were then further examined to uncover relationships, allowing for the grouping of related codes into broader thematic categories. Subsequently, these groups were analyzed to identify overarching themes that reflect deeper insights into the data. This iterative process was designed to ensure a comprehensive exploration of the qualitative data and enhance the interpretive depth of the analysis.

To ensure the reliability of the findings, at least two independent researchers reviewed and coded all data. The initial coding and preliminary analysis of the qualitative data was conducted by 1 author (SN-B), using Dedoose—a cloud-based software platform designed to facilitate mixed methods analysis. After the initial coding phase, 2 members of the research team engaged in collaborative discussions to reconcile coding discrepancies and synthesize their interpretations. This process of researcher triangulation not only strengthens the credibility of the findings but also helps to ensure that the emerging themes are robust and reflect the nuances present in the data [13].

The survey contained a brief textual description of the difference between the roles of mentor, advisor, and coach, respondents. Respondents were asked how important they thought each role is in graduate medical training, whether 1 individual can fulfill all 3 roles, what kind of training is needed for faculty to perform these roles, and whether resident needs for different forms of faculty relationship change over time. In addition, faculty were asked if they had ever performed any of the 3 roles. Questions were both quantitative (responses on a 5-point Likert scale) and qualitative (open-ended short responses). A total of 41 surveys were completed (44% response rate).

Ethical Considerations

This study used both quantitative and qualitative data collection and analysis. This study was approved by the Institutional Review Board at Washington University (202310164). Faculty were informed about this study via an initial email announcement, followed by 2 reminder emails. Informed consent was obtained with faculty selecting “accept” on the survey; the ability to opt out was provided. Electronic data were password protected, encrypted, and transmitted using recognized security for electronic submission. No compensation was provided.

Results

Roles

Respondents had varying opinions about the importance of mentoring, advising, and coaching in graduate medical education. Mentoring was seen as most important, with 88% of respondents indicating that they agreed or strongly agreed that it was important, and coaching was seen as less important, with only 78% of respondents indicating agreement or strong agreement that it was important (Table 1).

Table . Importance of mentoring, advising, and coaching.

	Agree or strongly agree, n (%)
Coaching	32 (78)
Advising	34 (83)
Mentoring	36 (88)

In total, 90% of respondents agreed that 1 faculty member could fulfill two or more roles for a single resident. For example, respondent 2 explained, “Faculty can possess more than one skill set and/or the relationship between a faculty and resident may benefit from a multi-faceted focus once trust has been developed.” However, others noted that there may be conflicts between roles and that the unique skills required for each role are not always possessed by the same person. Respondent 3 noted, “This works sometimes, I think, but can’t dependably work all the time. Some faculty are better at one role or another. Obviously, some coaching and advising can only be done by faculty with certain skills or areas of expertise.”

Additionally, respondents noted the role faculty mentoring and coaching play in recruitment and retention efforts for faculty and trainees. For example, respondent 3 noted, “if it were made clear that we offered thoughtful assignment of each of these roles, with examples for coaching and advising, I think that would likely be seen as a significant benefit.” Others agreed

XSL•FO
RenderX

that providing these roles to residents in a systematic way would be beneficial for recruitment, but noted barriers to implementation, as respondent 33 explained, “I think that these three roles are important to recruit residents for fellowships and faculty. Fostering a supportive environment through these roles is very important for recruitment; however, other factors such as the job market and hours worked often overshadow these aspects in recruiting.”

Training

Most respondents agreed that specialized training in all 3 roles was important, especially for coaching, which was seen as requiring a unique skill set. Formal training for all 3 roles was endorsed, especially for coaching. Respondents noted that the skills required for the roles came naturally to some faculty. For the advising role, having career experience and expertise in the graduate education process was seen as especially useful. For example, respondent 10 noted, “Knowing the residency experience well and knowing what challenges residents face. Additionally, it’s important to know career options after.” Mentoring was regarded as being based on relationship building and interpersonal skills, as well as necessitating emotional intelligence. Respondents reflected that mentorship involves skill sets not necessarily embedded in clinical training. Respondent 21 explained, “Teach the teacher/instructor courses are helpful. Being a good clinician and/or researcher do not provide us the skills of being a good teacher. A bit of more understanding, empathy, and psychological support are necessary for knowing ourselves better and using these abilities for others. Patience, more listening, time, sharing experiences, sometimes coming up with a challenging scenario to discuss, widen the horizon, show other possibilities never thought of before as options.”

Respondents indicated that they would be interested in targeted training. Coaching (63%) was the highest, however, respondents were less interested in specialized training in advising and mentoring skills (Table 2).

Table . Interest in specialized training in coaching, advising, and mentoring.

	Interest in specialized training, n (%)
Coaching	29 (63)
Advising	17 (41)
Mentoring	16 (39)

Experience

Nearly 88% of respondents had fulfilled one or more of these roles in their career, and they noted that holding all 3 roles was personally and professionally rewarding. Of the 36 faculty members who reported fulfilling these roles in the past, 15 (42%) mentioned the satisfaction of watching students progress through their training and career. Coaching was noted as being the most challenging, but also the most rewarding. For example, respondent 22 said, “Honestly, I think that serving in this role for strong residents is one of the most rewarding parts of my job. I love to see people be successful in their careers.”

Barriers

Respondents identified barriers to faculty engaging in quality mentoring, coaching, and advising, which included faculty burnout, time limitations, and confusion about roles, responsibilities, and expectations. Respondent 10 said, “The residents have so many rotations. It’s rare to have consistent clinic time to coach and mentor/advise. Coaching off hours is very time consuming.” Lack of time was mentioned by 68% of respondents, for example as respondent 29 explained “I was a terrible mentor. Never could find time to meet with my mentee.”

Respondents had mixed responses about whether the national anesthesia provider shortage had impacted their engagement with or performance of any of these roles. Respondents noted lack of time in general, and lack of protected time more specifically, as factors influencing their ability to engage in these roles, and some attributed the challenge with time to provider shortage. For example, respondent 17 said, “The shortage has decreased faculty time to provide these aspects, may be important for departments to assign a subgroup of faculty to serve these roles so time is protected.”

Table 3 presents the results of the thematic analysis, offering a detailed synthesis of the emergent themes and subthemes derived from the qualitative data. The richness of respondent narratives facilitated a comprehensive exploration, allowing for nuanced insights into the key thematic categories. These findings provide a robust framework for understanding the underlying patterns and relationships within the data, supporting the depth and validity of the analysis.

XSL•FO
RenderX

Table . Main themes and representative quotes.

Theme and subtheme		Representative quotes
Roles		
	Faculty can perform multiple roles	• "Faculty can possess more than one skill set and/or the relationship between a faculty and resident may benefit from a multi-faceted focus once trust has been developed." [Respondent 2] • "Different skill sets are needed and faculty may possess one or many of the skill sets needed." [Respondent 9] • "I believe the necessary skills can be learned and employed by a single person. It also depends upon the mentee's needs and the qualities of their relationship with the mentor/advisor/coach." [Respondent 5] • "A faculty member can take different roles throughout the 4 years that a trainee is counseled. I find that interns need mentoring and advising, as the resident progresses coaching and mentoring is important." [Respondent 16]
	Faculty cannot perform multiple roles	• "This works sometimes, I think, but can't dependably work all the time. Some faculty are better at one role or another. Obviously, some coaching and advising can only be done by faculty with certain skills or areas of expertise." [Respondent 3] • "Sometimes the line between just providing feedback for a specific case as an advisor can be hard if you are also a mentor to that person." [Respondent 37] • "Different goals and different time frames over which those goals are realized. The trainee asking advising may be frustrated by a "mentoring" approach. Some great career mentors may not have the specific subspecialty background for focused advising." [Respondent 7]
Training		

XSL•FO
RenderX

Theme and subtheme		Representative quotes
	Request for formal education or faculty development	• "I think at least some sort of education on how to be an advisor would be helpful." [Respondent 1] • "Teach the teacher/instructor courses are helpful. Being a good clinician and/or researcher do not provide us the skills of being a good teacher. A bit of more understanding, empathy, and psychological support are necessary for knowing ourselves better and using these abilities for others. Patience, more listening, time, sharing experiences, sometimes coming up with a challenging scenario to • discuss, widen the horizon, show other possibilities never thought of before as options." [Respondent 21] • "Coaching should require some training/knowledge of professional coaching, which is more structured that mentorship or career advising which can be more informal." [Respondent 4] • "Didactics/workshops/peer mentoring needed." [Respondent 31] • "Training focused to the knowledge and skillset as well as teaching techniques and current best practices." [Respondent 2] • "Structured professional coaching training." [Respondent 6]
Experiences		

XSL•FO
RenderX

Theme and subtheme		Representative quotes
	Mentor role	• "I was a terrible mentor. Never could find time to meet with my mentee." [Respondent 29] • "Mentoring has been the most rewarding, coaching second. Advising feels limited and one-directional." [Respondent 5]
	Coach role	• "All three - coaching seems to be the most challenging." [Respondent 7] • "I have played all 3 roles during my time as an educator. The coaching roles are always the most rewarding. The ability to guide residents through self-discovery is extremely rewarding. I find that coaching residents later in their training prepares them for being faculty and having a successful trajectory." [Respondent 17] • "I have been a coach and an advisor. Coaching is extremely rewarding." [Respondent 39] • "Primarily coaching, which I found rewarding when a trainee felt our interaction was beneficial through skill-based or confidence improvements." [Respondent 41]
	Advisor role	• "Advising in clear goal-directed tasks, such as a conference, abstract, paper." [Respondent 8] • "I have served as a mentor and advisor, both of which were very rewarding. I felt that it made it easier to discuss topics at work that we may otherwise would not have brought up. I also felt satisfaction getting to know the trainees better and become more a part of their lives." [Respondent 27] • "Clinical teaching while supervising trainees fulfills the "advisor" role. I was also a designated faculty mentor for a clinical fellow." [Respondent 34] • "Clinical teaching while supervising trainees fulfills the advisor role." [Respondent 33]
	Combination of roles	• "Have provided all three of these roles in different capacities. I enjoy fostering learning with the goal of being the attending I wish I had as a trainee." [Respondent 33] • "Yes, I feel that I serve as an advisor to residents and mentor to fellows." [Respondent 20] • "I would say informally on day-to-day basis interactions with residents and fellows, yes for all 3. Advisor more than mentor more than coach. It is rewarding when it seems welcomed and appreciated by the residents and fellows and I can see them grow and improve. It is frustrating when I am putting in the effort/trying to do these things and the trainees are not receptive, not appreciative, or feel as though I am being too particular or micromanaging." [Respondent 35] • "I have provided all 3. The coaching roles are always the most rewarding. The ability to guide residents through discovery is extremely rewarding." [Respondent 20]
	Recruitment role	

XSL•FO
RenderX

Theme and subtheme	Representative quotes
	• "The biggest drivers right now for recruitment are time and money. The biggest long-term satisfaction will come from deeper meaning. Using the relationships in these roles may help highlight some of these deeper meanings and may help recruit fellows and faculty if they have the sense that this is best for themselves and their families. At the same time, there has to be felt and sustained room for the individual to act on these deeper meaningful insights. Solving for individual growth requires commitment from the system as well as the individual." [Respondent 9] • "Yes. When residents can see faculty care about their education and also enjoy working here it's easier to recruit. " [Respondent 20] • "Mentorship and coaching require a relationship, that may be beneficial for recruitment." [Respondent 17] • "A structured mentor/coaching program would be very appealing to most applicants." [Respondent 31]
Barriers	

XSL•FO
RenderX

Theme and subtheme		Representative quotes
	Discrete roles	• "If role/project is not clearly defined, could cause some confusion. Time." [Respondent 1] • "Mentorship is often a friendly and personal relationship, which could make it harder to, for example, challenge the mentee in a coaching scenario. Very specific example - perhaps a mentee would feel uncomfortable doing mock oral boards with their mentor, if they're relatively advanced in training, but early in the oral boards prep process." [Respondent 3] • "Different goals and different time frames over which those goals are realized. The trainee asking for advising may be frustrated by a "mentoring" approach. Some great career mentors may not have the specific subspecialty background for focused advising." [Respondent 7]
	Time	• "Time and lack in continuous interactions with the resident." • [Respondent 18] • "Time to meet with the trainee and to establish a relationship." • [Respondent 14] • "Time and managing the balance btw one's professional responsibilities and taking on additional responsibilities that the above would entail." [Respondent 6] • "The residents have so many rotations. It's rare to have consistent clinic time to coach and mentor/advise. Coaching off hours is very time-consuming." [Respondent 10]
	Burnout	• "It would be a good recruitment tool but difficult to deliver in near future with current staffing shortages and burn-out among faculty members. In practice, it would require significant training, time, and effort to optimize and ensure an equal experience among trainees. Remuneration could increase participation but doesn't get around the issue of lack of time." [Respondent 12] • "Yes, particularly for faculty. Relatively little resources currently to develop faculty. More investment needed to reduce the chance of burnout/disengagement/attrition to other practices." [Respondent 31] • "We are all strapped for time and burnt out." [Respondent 40]
	Anesthesia shortage	• "These 3 are probably even more important for our trainees and may be beneficial to expand these past trainees and onto faculty as well. The shortage has decreased faculty time to provide these aspects, may be important for departments to assign a subgroup of faculty to serve these roles so time is protected." [Respondent 17] • "I think that with the shortages, faculty have taken on more solo assignments and have overall less contact with the residents and don't get to know them as well." [Respondent 36]

XSL•FO
RenderX

Discussion

Overview

This study explored perceptions of anesthesia faculty regarding the roles of mentoring, advising, and coaching in graduate medical education. The results highlight the perceived benefits of these practices as well as barriers to implementation. Anesthesia residency is unique in its internship, and a vast majority of education and interactions with faculty occurs at bedside in the operating room. Medical training and trainee progression differ across disciplines. This study focuses specifically on anesthesia faculty and a single institution, which overall limits generalizability.

Principal Findings

The survey results indicate that faculty view mentoring, advising, and coaching as important for resident education and development. These practices have been shown to improve resident well-being, promote career planning, facilitate reflection and self-assessment, and identify knowledge gaps [5,6]. Furthermore, implementing structured programs in these areas can aid recruitment and retention of both residents and faculty.

Of the 3 roles faculty partake in, there is a consensus on the importance of mentoring throughout training and prioritizing this role over advising and coaching. However, the data suggests a significant interest in specialized training for coaching versus roles in advising and mentoring. Investigating the differences in practice versus desire, recurrent themes of time and experience were identified. Although the roles as a mentor, advisor, and coach can overlap, a majority of the cohort indicated they prioritize mentoring given the noted constraints of time and experience.

Implications of Findings

To actualize these practices, each department must clearly define the roles of mentor, advisor, and coach. Expectations, training requirements, and time commitments should be delineated. Assignments of roles can be made between faculty and residents based on alignment of career goals, personalities, and logistics. Protected nonclinical time should be designated for these meetings separate from clinical work. Success stories and positive impacts on residents should be tracked and celebrated.

Comparison to the Literature

Recurrent themes were identified when comparing to other literature, such as the establishment of a clear definition and terms of each role. This would help faculty facilitate their approach to the learner needs [5,8]. Additional repeated themes of the overlap in roles, limitations in time, and experiences were highlighted in other studies in reference to mentorship, advising, and coaching [5,10]. Anesthesiology training presents challenges specific to the discipline, which can be generalized to medical training programs at other institutions. There has been an increased productivity within the academic institutions leading to less bedside education opportunities and difficulty establishing dedicated time for routine meetings with trainees.

Limitations

This study has several limitations. First, this study was based on a single survey with a 44% response rate, which may limit the generalizability of the findings. Nonresponders may have had different perspectives on the importance and implementation of mentoring, advising, and coaching. Second, this study was conducted at a single academic medical center, so the results may not be representative of other institutions. Additionally, this study was solely conducted with anesthesia faculty. Other specialties may not portray the same obstacles and constraints in fulfilling the roles of mentorship, advising, and coaching. The learning environment and progression through training also differ between anesthesiology and other specialties, which limits the generalizability across disciplines. The limited time and consistency with faculty may lead to less specific demands from trainees and unfulfillment from educators. Third, the survey relied on self-reported perceptions and experiences, which are subject to recall bias and social desirability bias. Fourth, this study did not explore the perspectives of residents themselves on these support modalities. Future research should examine resident experiences with and preferences for mentoring, advising, and coaching. Finally, while this study identified perceived barriers to implementing these practices, it did not evaluate specific strategies for overcoming these obstacles. Further work is needed to develop and test interventions to enhance faculty engagement in resident support roles.

Conclusion

Addressing barriers such as faculty burnout, role ambiguity, time constraints, and the need for specialized training is critical for the success of mentoring, advising, and coaching initiatives. Implementing comprehensive faculty development programs aimed at enhancing skills in these domains is essential, particularly for coaching, which requires distinct pedagogical approaches. The recruitment and retention of faculty, as well as their career longevity, may be positively influenced by the intrinsically rewarding nature of relationships with trainees.

To facilitate meaningful faculty engagement, institutional leadership must ensure protected time for participation in these activities without detriment to clinical productivity. Moreover, a cultural shift may be necessary in programs that place disproportionate emphasis on service obligations, potentially at the expense of educational and developmental support for residents. Prioritizing resident education and well-being can contribute to improved morale and overall program satisfaction.

By investing in faculty development, enhancing institutional infrastructure, and fostering a culture that values educational alliance, graduate medical education programs can realize significant benefits from high-quality mentoring, advising, and coaching relationships. Such investments are pivotal for advancing the professional development of both faculty and trainees, ultimately enhancing the overall quality of medical education.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Coaching, mentoring, and advising survey.
[DOCX File, 17 KB - mededu_v11i1e60255_app1.docx]

References

1. ăranu SM, tefăniu R, Rotaru T, et al. Factors associated with burnout in medical staff: a look back at the role of the COVID-19 pandemic. Healthcare (Basel) 2023 Sep 13;11(18):2533. [doi: 10.3390/healthcare11182533] [Medline: 37761730]
2. GlobalData Plc. The complexities of physician supply and demand: projections from 2019 to 2034. : AAMC; 2024.
3. Plochocki JH. Several ways generation Z may shape the medical school landscape. J Med Educ Curric Dev 2019 Oct 31;6:2382120519884325. [doi: 10.1177/2382120519884325] [Medline: 31701014]
4. Farkas AH, Allenbaugh J, Bonifacino E, Turner R, Corbelli JA. Mentorship of US medical students: a systematic review. J Gen Intern Med 2019 Nov;34(11):2602-2609. [doi: 10.1007/s11606-019-05256-4] [Medline: 31485967]
5. Marcdante K, Simpson D. Choosing when to advise, coach, or mentor. J Grad Med Educ 2018 Apr;10(2):227-228. [doi: 10.4300/JGME-D-18-00111.1] [Medline: 29686766]
6. Deiorio NM, Hammoud MM. Coaching in medical education: accelerating change in medical education consortium. : AMA; 2024.
7. Alisic S, Boet S, Sutherland S, Bould MD. A qualitative study exploring mentorship in anesthesiology: perspectives from both sides of the relationship. Can J Anesth/J Can Anesth 2016 Jul;63(7):851-861. [doi: 10.1007/s12630-016-0649-3]
8. Santiesteban L, Young E, Tiarks GC, et al. Defining advising, coaching, and mentoring for student development in medical education. Cureus 2022 Jul;14(7):e27356. [doi: 10.7759/cureus.27356] [Medline: 36043012]
9. Wolff M, Deiorio NM, Miller Juve A, et al. Beyond advising and mentoring: competencies for coaching in medical education. Med Teach 2021 Oct;43(10):1210-1213. [doi: 10.1080/0142159X.2021.1947479] [Medline: 34314291]
10. Reid MB, Misky GJ, Harrison RA, Sharpe B, Auerbach A, Glasheen JJ. Mentorship, productivity, and promotion among academic hospitalists. J Gen Intern Med 2012 Jan;27(1):23-27. [doi: 10.1007/s11606-011-1892-5] [Medline: 21953327]
11. Jha P, Quinn B, Durbin S, Bhandari S. Perceptions of junior faculty in general internal medicine regarding mentoring medical students and residents in scholarly projects. J Gen Intern Med 2019 Jul;34(7):1098-1099. [doi: 10.1007/s11606-019-04937-4] [Medline: 30887433]
12. Peel KL. A beginner's guide to applied educational research using thematic analysis. Pract Assess Res Eval 2020 Jan;25(1):2. [doi: 10.7275/ryr5-k983]
13. Patton M. Qualitative Research and Evaluation Methods: Sage; 2015.

Abbreviations

REDCap: Research Electronic Data Capture

Edited by B Lesselroth; submitted 06.05.24; peer-reviewed by K Kagkelaris, M Ilaghi; revised version received 23.09.24; accepted 03.12.24; published 21.01.25.

Please cite as:
Nykiel-Bailey S, Burrows K, Szafarowicz BE, Moquin R
Faculty Perceptions on the Roles of Mentoring, Advising, and Coaching in an Anesthesiology Residency Program: Mixed Methods Study
JMIR Med Educ 2025;11:e60255
URL: https://mededu.jmir.org/2025/1/e60255
doi:10.2196/60255

© Sydney Nykiel-Bailey, Kathryn Burrows, Bianca E Szafarowicz, Rachel Moquin. Originally published in JMIR Medical Education (https://mededu.jmir.org), 21.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Brief Web-Based Person-Centered Care Group Training Program for the Management of Generalized Anxiety Disorder: Feasibility Randomized Controlled Trial in Spain

Vanesa Ramos-García[1,2,3,4], BSc; Amado Rivero-Santana[1,3,4], PhD; Wenceslao Peñate-Castro[2], PhD; Yolanda Álvarez-Pérez[1,3,4], PhD; Andrea Duarte-Díaz[1,3,4], BSc; Alezandra Torres-Castaño[1,3,4], PhD; María del Mar Trujillo-Martín[1,3,4], PhD; Ana Isabel González-González[5,6], PhD; Pedro Serrano-Aguilar[3,4,7], PhD; Lilisbeth Perestelo-Pérez[3,4,7], PhD

[1]Canary Islands Health Research Institute Foundation, Santa Cruz de Tenerife, Spain

[2]Department of Clinical Psychology, Psychobiology and Methodology, University of La Laguna (ULL), Santa Cruz de Tenerife, Spain

[3]Network for Research on Chronicity, Primary Care, and Health Promotion (RICAPPS), Tenerife, Spain

[4]The Spanish Network of Agencies for Health Technology Assessment and Services of the National Health System (RedETS), Tenerife, Spain

[5]Network for Research on Chronicity, Primary Care, and Health Promotion (RICAPPS), Madrid, Spain

[6]Área de Fomento de la Innovación e Internacionalización de la Investigación Sanitaria, Subdirección General de Investigación Sanitaria y Documentación, Dirección General Investigación y Docencia, Consejería de Sanidad, Madrid, Spain

[7]Evaluation Unit (SESCS), Canary Islands Health Service (SCS), Santa Cruz de Tenerife, Spain

Corresponding Author:
Lilisbeth Perestelo-Pérez, PhD
Network for Research on Chronicity, Primary Care, and Health Promotion (RICAPPS), Tenerife, Spain

Abstract

Background: Shared decision-making (SDM) is a crucial aspect of patient-centered care. While several SDM training programs for health care professionals have been developed, evaluation of their effectiveness is scarce, especially in mental health disorders such as generalized anxiety disorder.

Objective: This study aims to assess the feasibility and impact of a brief training program on the attitudes toward SDM among primary care professionals who attend to patients with generalized anxiety disorder.

Methods: A feasibility randomized controlled trial was conducted. Health care professionals recruited in primary care centers were randomized to an intervention group (training program) or a control group (waiting list). The intervention consisted of 2 web-based sessions applied by 2 psychologists (VR and YA), based on the integrated elements of the patient-centered care model and including group dynamics and video viewing. The outcome variable was the Leeds Attitudes Towards Concordance scale, second version (LATCon II), assessed at baseline and after the second session (3 months). After the randomized controlled trial phase, the control group also received the intervention and was assessed again.

Results: Among 28 randomized participants, 5 withdrew before the baseline assessment. The intervention significantly increased their scores compared with the control group in the total scale (b=0.57; P=.018) and 2 subscales: communication or empathy (b=0.74; P=.036) and shared control (ie, patient participation in decisions: b=0.68; P=.040). The control group also showed significant pre-post changes after receiving the intervention.

Conclusions: For a future effectiveness trial, it is necessary to improve the recruitment and retention strategies. The program produced a significant improvement in participants' attitude toward the SDM model, but due to this study's limitations, mainly the small sample size, more research is warranted.

(JMIR Med Educ 2025;11:e50060) doi:10.2196/50060

KEYWORDS

person-centered care; primary care; shared decision-making; anxiety disorder; training program; SDM

Introduction

About 264 million people in the world are affected by anxiety disorders, according to the latest estimates of the World Health Organization [1]. In Spain, around 2 million people (4.1% of the population) suffer from anxiety disorders [1]. In primary care (PC) settings, the generalized anxiety disorder (GAD) is one of the most prevalent anxiety disorders [2]. GAD is characterized by a continuous state of worry and alertness most

XSL•FO
RenderX

of the time [3] and sometimes, its high comorbidity with other psychiatric and somatic disorders makes diagnosis difficult [4]. GAD has a tendency to chronicity, due to its specific characteristics, leading to the person being worried and alert most of the time [3]. Information on the causes of the disorder and the available treatments is an unmet need in this population, given that some patients with GAD are willing to have an active or collaborative role in their health care [5].

Person-centered care (PCC) is considered the gold standard for medical care in health care settings because it humanizes the person and places him or her at the center of clinical decision-making [6]. The PCC model consists of several components, one of which is shared decision-making (SDM), whose goal is to create a collaborative dialogue between patients and health care professionals, in which patients' values, preferences, and concerns about the different available treatment options are taken into account and incorporated into the decision-making process [7-9].

Patient decision aids are tools designed to facilitate SDM. Its use can help patients participate in the clinical decisions, improving the decision-making process and promoting informed decisions that are concordant with patients' values and preferences [10]. On the part of professionals, it is important to develop communication skills and empathy to help patients participate in the decisions [11-13]. Research has shown that interventions and training programs aimed to promote the PCC model may improve professionals' knowledge and the ability to communicate with patients [12,14] as well as patients' satisfaction [15]. However, there are some barriers to apply the PCC model related to time constraints, clinical uncertainty, poor expectations, patients' characteristics (eg, age, comorbidity, and attitude), lack of continuity of care, or knowledge about SDM [16-19]. Despite some SDM training programs have been developed for health care professionals, very few of them have been evaluated [20-22]. Therefore, despite the growing acceptance of interventions to implement SDM in health care settings, several gaps remain in the demand, perception, and clinical application of the PCC model [23,24]. In mental health care, and specifically in GAD, interventions to promote the SDM process are still very limited [25,26]. A recent qualitative study with patients with GAD concluded that there is scarce orientation to elicit patients' preferences and values throughout the process of care [27], emphasizing the need of interventional studies aimed at promoting SDM in the clinical encounter.

The aim of this study is to evaluate the feasibility and effect of a brief training program on the attitudes toward SDM for professionals in PC who attend patients with GAD.

Methods

Design

A feasibility randomized controlled trial (RCT) was conducted, in which participants were allocated to a PCC training program or a control group (waiting list). It was carried out in 13 PC centers in Tenerife (Canary Islands, Spain), from January 2021 to February 2022.

Ethical Considerations

The study was approved by the ethics committee of the Hospital Universitario Nuestra Señora de La Candelaria (reference: CHUNSC_2019_58). The study was not registered because participants were health professionals and not patients, the intervention was educational, and the only outcome measured was attitudinal. Participants who agreed to participate signed a web-based informed consent form.

Participants

Participants were health care professionals working in PC centers (ie, physicians and nurses) or community mental health units (ie, psychiatrists, psychologists, and nurses) for at least 1 year before the start of the study, who attend patients with GAD in the Canary Islands, Spain. There were no exclusion criteria.

Procedure, Randomization, and Allocation Concealment

The directors of the health centers were contacted and informed about the study. They were asked to invite the professionals from their centers to participate. The invitation included an infographic, graphically describing the study and a link to a web platform, where health professionals could register their willingness to participate and contact information. Then, they were contacted by telephone to provide a full explanation of the study. Those who agreed to participate signed a web-based informed consent form (reference: CHUNSC_2019_58). Participants were randomly assigned to either the intervention or control group (waiting list), using a computer-generated random number table. The randomization process was conducted by an independent researcher who was not involved in the recruitment or assignment of participants. In addition, the researcher who recruited the professionals was blinded to the group assignments in order to maintain allocation concealment. Due to the nature of the intervention, the study participants could not be blinded.

Intervention

Intervention group participants received 2 training sessions via Zoom (version 5.15.7. [21404]) based on the integrated elements of the PCC model [28]. The training was originally intended to be applied in person, in a group format, but this was not possible due to the COVID-19 pandemic, so it was finally applied on the web. Sessions were conducted by 2 researchers (VR and YA [psychologists]). The first session lasted approximately 2 hours and was focused on presenting the principal elements of intervention: (1) introduction, which included a description of common clinical relationship models (first 20 minutes); (2) basic characteristics of the basic PCC model, through group dynamics and video viewing of a role-play in the clinical practice with a patient with GAD; this included a description of the Feelings, Ideas, Function, and Expectations model [29] (60 minutes), which was developed at the University of Western Ontario and explores the patient's emotions, his or her ideas on what caused the problem, the effects of the illness on his or her functioning and relationships, and his or her expectations for the future and from medical care [29,30]; and (3) presentation of the Three-Talk Model for SDM, a multistage consultation process developed by Elwyn et al [31] (30 minutes). The

XSL•FO
RenderX

Three-Talk Model for SMD is a theoretical approach that describes collaborative deliberation. It outlines 3 broad steps that form the core elements of SDM [31]. The last 10 minutes of the session were aimed at the resolution of doubts. The detailed contents of this first SDM training session are shown in Table 1.

Table . Content of first shared decision-making (SDM) training session.

Module and content	Form of communication	Learning objectives
Introduction		
Clinical relationship models	• Lecture • Video examples • Interactive live • Feedback with group dynamic	• Be able to know the characteristics of the paternalistic, informative or contractual, interpretive or personalized, and deliberative or friendly models
Characteristics of a basic PCC[a] model		
Explore the disease	• Lecture	• Acquire skills in active listening and directed anamnesis in the use of SDM
Know the patient's perspective (beliefs, fears, expectations, repercussions, etc)	• Lecture • Video examples • Interactive live • Feedback with group dynamic	• Acquire skills in how to prepare the ground and how to explore the personal experience of the disease in terms of SDM • Be able to use the FIFE[b] model to improve the quality of communication in terms of SDM
Know the person ("moving from patient to person")	• Lecture • Video examples	• Acquire skill about how exploring the personal and social context of the disease in terms of SDM
Involve the patients in their disease	• Lecture • Video examples • Interactive live • Feedback with group dynamic	• Acquire information skills to reach agreements on problem solving, to seek shared solutions, and to involve the patient in the use of SDM
Three-Talk Model for SDM		
Team dialogue	• Lecture	• Acquire skills to establish a team dialogue based on the needs for change on beliefs and preferences
Dialogue on options	• Lecture	• Acquire skills to discuss the treatment options that exist for the disease
Dialogue on the decision	• Lecture	• Acquire skills to help the patient decide on which option to choose

[a]PCC: person-centered care.

[b]FIFE: Feelings, Ideas, Function and Expectations.

The second session was carried out 3 months later (review session), with an approximate duration of 1 hour. The structure of the session included (1) the review of the main contents of the first training module, together with comments on participants' potential and sharing their experiences applying the SDM model since then (30 minutes), and (2) the discussion on the main barriers and facilitators for patients and professionals in applying the SDM process in the clinical practice (30 minutes). Detailed content of this session is present in Table 2.

Control group participants did not receive any intervention. They were informed that they could access the training program after the feasibility RCT was completed. Participants completed the baseline and 3-month (postintervention) assessments. Subsequently, participants in the control group received the intervention and were reevaluated 3 months later (second postintervention measure).

XSL•FO
RenderX

Table . Content of second shared decision-making (SDM) training session.

Unit and content	Form of communication	Learning objectives
(1) Introduction and (2) characteristics of a basic PCC[a] model		• Review the characteristics of the paternalistic model,: informative or contractual, interpretive or personalized, and the deliberative or friendly models • Review tasks in active listening and directed anamnesis in the use of SDM: how to prepare the ground and how to explore the personal experience of the disease; how to explore the personal and social context of the disease; and how to reach agreements on problem solving, to seek shared solutions, to involve the patient-shared solutions, and to involve the patient in the use of SDM
Clinical relationship models Explore the disease; know the patient's perspective (beliefs, fears, expectations, repercussions, etc); know the person ("moving from patient to person"); and involve the patients in their disease	Lecture	
(3) Characteristics of the Three-Talk Model for SDM		• Be able to apply the principal components of the Three-Talk Model for SDM • Have knowledge about how to apply this model in clinical practice to support SDM
Fifteen characteristics total of a Three-Talk Model for SDM are described: *First step*: Take a step back, present the possibility of choice, justify the choice, personalizing preference, uncertainty, check the reaction, and postpone closure *Second step*: Check knowledge, list of options, provide decision support to the patient, and summaries *Third step*: Focus on preferences, elicit a preference, lead toward a decision, and offer review	Lecture	
(4) Barriers and enablers to apply Three-Talk Model for SDM		• Invite to participate by presenting the experience from a professional point of view in clinical practice • Openly share and discuss observations of the professional communication • Offer, explicitly and without judging, feedback on implementation
Identification of barriers from a professional point of view that can condition the application of the 3-step model for SDM	Identification of professionals' own barriers to communication with their patients	
Identification of barriers from a patient's point of view that can condition the application of the 3-step model for SDM	Identification of patients' own barriers to communication with their care team	
Identification of facilitators who may exist to carry out the 3-step model for SDM	Identify the individual facilitators in communication to implement a SDM model	

[a]PCC: person-centered care.

XSL•FO
RenderX

Measures

The outcome measure was the professionals' attitude toward PCC. It was assessed with the Leeds Attitudes Towards Concordance scale, second version (LATCon II) [32]. This self-report instrument includes 20 items with a 4-point Likert format from strongly disagree (0) to strongly agree (3). Although the original instrument includes 5 subscales, we used the 3 components identified by means of principal component analysis in the Spanish validation [33], carried out with psychiatrists and psychiatry residents. These subscales were labeled "communication/empathy" (CE, 12 items about the importance of a good communication and the consideration of patient's feelings and beliefs), "shared control" (SC, 4 items reflecting a positive attitude toward equality and SDM), and "eventual paternalistic style" (EPS, 4 items stating that sometimes a paternalistic style is necessary; these items are reverse-coded, and therefore higher scores indicate lower agreement with EPS) [33]. Scores on the total scale and the subscales are divided by the corresponding number of items, thus ranging 0 - 3. The LATCon II has shown good internal consistency in previous studies [33-35].

The following sociodemographic and professional variables were measured at baseline: age, gender, specialty (medicine or nursing), years of professional experience and work in the health care center, level of perceived workload (low, medium, and high), and previous training on PCC or SDM.

Statistical Analysis

We calculated that a mixed model with 2 repeated measures per participant (cluster) requires 38 subjects (19 in each group) in order to detect a significant moderate-to-strong between-group effect (standardized mean difference of 0.80), assuming type I and II errors of 0.05 and 0.20, respectively, and an intraclass correlation of 0.50 [35].

Descriptive statistics were calculated for continuous and categorical variables (means, SDs, and percentages). Cronbach α was calculated for the LATConII scale and its 3 subscales, as well as the correlations between the subscales (Spearman ρ). The effect of the intervention was analyzed with mixed lineal models, including fixed effects for time (pre, post), group (intervention, control) and its interaction, and the participant as a random effect (assuming an unstructured covariance matrix). Successive models were carried out adjusting for 1 covariate at a time (ie, sociodemographic and professional variables). Unstandardized β values and effect sizes (Hedges *g*) are reported.

Changes from baseline to postintervention were evaluated analyzing the effect of time in a mixed model separately for each group. The same test was used to analyze the change in the control group after receiving the intervention upon completion of the RCT. Analyses were performed with SPSS (version 25; IBM Corp) and STATA (version 17; StataCorp LLC).

Results

Thirty-four health professionals were interested in participating and were contacted by phone. After being informed in detail, 6 declined participation and 28 accepted, signing informed consent and being randomly allocated to the intervention or control group (14 each). However, 5 of them (4 in the intervention group) withdrew from the study before completing the baseline assessment (Figure 1). Table 3 shows the characteristics of the 23 participants. There were 18 women (18/23, 78.3%) and the mean age was 48.3 (range: 26 - 64) years. They had an average of 22.3 years of professional experience, and 52% (12/23) considered having a high caseload. Only 5 (21.7%) had had previous training in PCC.

XSL•FO
RenderX

Figure 1. CONSORT flow diagram. *The analysis includes 3 control participants lost at postintervention.

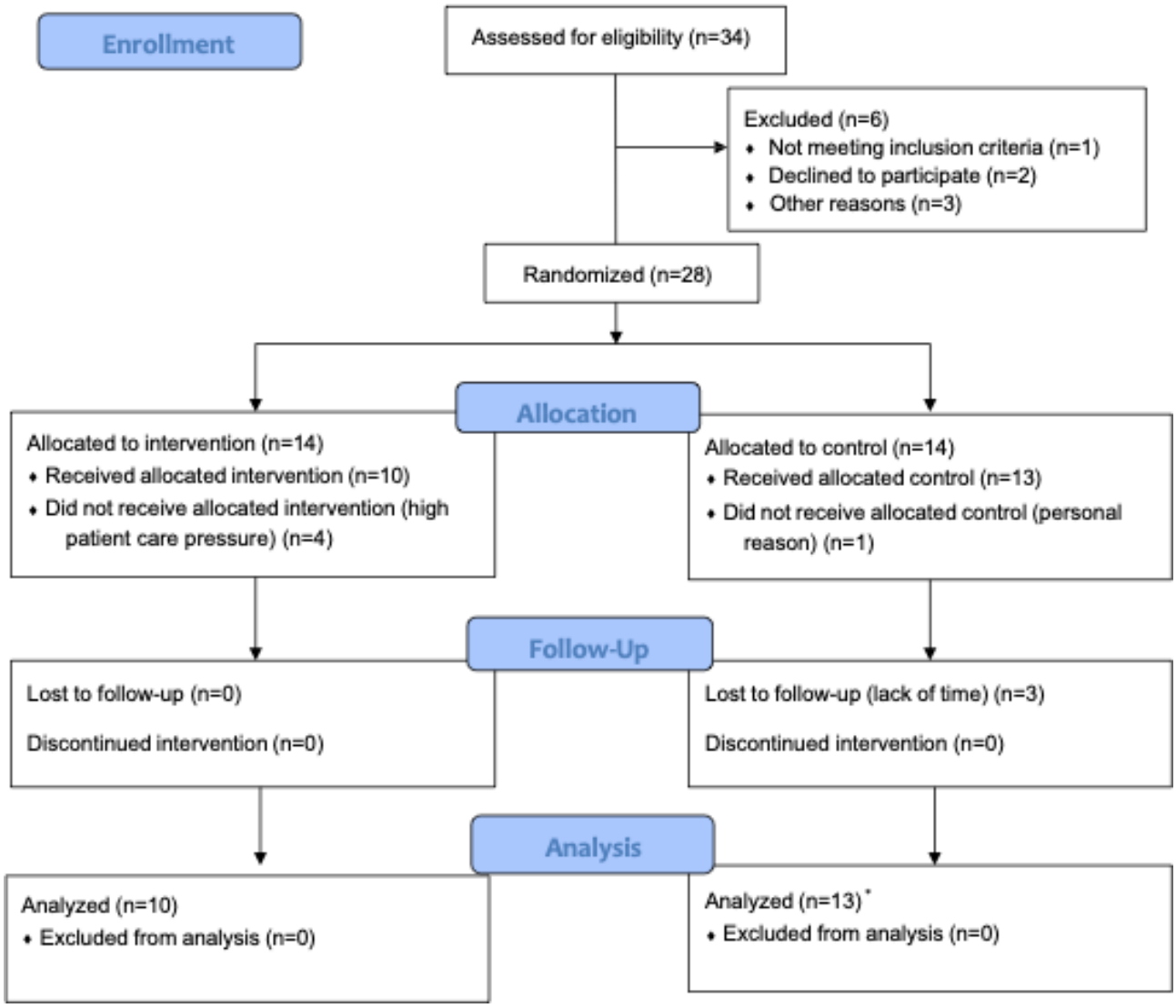


Table . Characteristics of participants.

	Intervention (n=10)	Control (n=13)	Total (N=23)
Female, n (%)	10 (100)	8 (61.54)	18 (78.26)
Age, mean (SD)	45.60 (10.82)	50.38 (9.77)	48.30 (10.24)
Specialty, n (%)			
Nursing	3 (30)	2 (15.38)	5 (21.74)
Medicine	7 (70)	11 (84.62)	18 (78.26)
Years of professional experience, mean (SD)	19.70 (9.44)	24.33 (9.42)	22.32 (9.51)
Years working in the center, mean (SD)	7.05 (7.15)	6.40 (7.02)	6.68 (6.92)
Previous training in PCC[a], n (%)	1 (10)	4 (30.77)	5 (21.74)
Self-perceived care load, n (%)			
Low-medium	6 (60)	5 (38.46)	11 (47.83)
High	4 (40)	8 (61.54)	12 (52.17)

[a]PCC: person-centered care.

XSL•FO
RenderX

At baseline, internal consistency (Cronbach α) was 0.94 for the total LATConII scale, and 0.97 (CE), 0.88 (SC), and 0.25 (EPS) for the subscales. The total mean score was 2.08 (SD 0.60), and the mean scores were 2.29 (SD 0.78), 1.77 (SD 0.85), and 1.78 (SD 0.41) for the subscales CE, SC, and EPS, respectively (Table 3). CE and SC were significantly correlated (ρ=0.49; *P*=.01), whereas EPS was not significantly associated with CE (ρ=0.11; *P*=.62) or SC (ρ=0.29; *P*=.180) (Table 4).

Table . Effect of the intervention.

Time[a]	Intervention (n=10), mean (SD)	Control (n=10), mean (SD)	Time × group interaction, *b* (*P*)[b]	Between-group effect size, Hedges *g* (95% CI)
LATCon II[c] total (range: 0 - 3)			0.57 (.018)	0.92 (0.13 to 1.71)
Pre	1.87 (0.76)	2.25 (0.40)[d]		
Post	2.27 (0.51)[e]	2.08 (0.61)		
Post2	—[f]	2.60 (0.24)[g]		
Communication/empathy (range: 0 - 3)			0.74 (.036)	0.86 (0.06 to 1.65)
Pre	1.98 (1.02)	2.52 (0.46)[d]		
Post	2.57 (0.70)[e]	2.34 (0.86)		
Post2	—	2.84 (0.20)[e]		
Shared control (range: 0 - 3)			0.68 (.040)	0.76 (0.01 to 1.52)
Pre	1.55 (1.06)	1.94 (0.63)[d]		
Post	1.80 (0.44)	1.52 (0.70)[e]		
Post2	—	2.28 (0.43)[h]		
Eventual paternalistic style (range: 0 - 3)			−0.04 (.856)	0.08 (−0.93 to 0.93)
Pre	1.83 (0.44)	1.75 (0.41)[d]		
Post	1.83 (0.57)	1.83 (0.54)		
Post2	—	2.18 (0.64)[e]		

[a]Pre-post: randomized controlled trial (intervention vs waiting list); post2: intervention period for the control group, after the randomized controlled trial.

[b]Unstandardized β coefficients (*P* value) from mixed lineal models analyzing the randomized controlled trial (pre-post), including the participant as a random effect (the analysis includes 3 control participants lost at postintervention).

[c]LATCon II: Leeds Attitudes Towards Concordance scale, second version.

[d]n=13.

[e]*P*<.05.

[f]Not applicable.

[g]*P*<.001, compared with the previous assessment (effect of time in mixed models separately by group).

[h]*P*<.01.

Three control participants were lost at postintervention (3 months), but their baseline scores were included in the mixed models on an intention-to-treat basis (postintervention scores were not imputed). The time × group interaction was statistically significant for the total scale, showing a differential increment in scores favoring the intervention (*b*=0.57; *P*=.01) (Table 4). The same occurred with the subscales CE (*b*=0.74; *P*=.036) and SC (*b*=0.68; *P*=.04). The inclusion of potential confounders in the model did not change the results (see Table S1 in Multimedia Appendix 1 for the total scale). The intervention group significantly increased their scores compared with baseline in the total scale (*b*=0.4; *P*=.033) and CE (*b*=0.58; *P*=.030), whereas the control group significantly decreased in SC (*b*=−0.43; *P*=.037) (Table 4).

After the trial was completed, the control group received the intervention and showed significant increments in the total score (*b*=0.52; *P*<.001) and the 3 subscales: CE (*b*=0.50; *P*=.020), SC (*b*=0.75; *P*=.002), and EPS (*b*=0.35; *P*=.02) (Table 4).

Discussion

Principal Findings

This study aimed to evaluate the feasibility and effect of a brief web-based training program on the attitudes toward SDM and PCC of PC professionals who treat patients with GAD. The program was initially intended to be conducted in person at the professionals' centers, but due to the pandemic context, it was shifted to a web-based format. The 2 sessions went smoothly and the professionals actively participated, asking questions

XSL•FO
RenderX

and describing their experiences related to SDM. Previous studies evaluating learning programs for health professionals or university students have not shown relevant differences between web-based and in-person formats [36-39], although in some cases better results have been observed with the face-to-face intervention [40]. Given the brevity of our program, we do not expect that there will be relevant differences between both formats.

The recruitment and retention rate were low, only 33 eligible professionals showed interest in the study (2.5 per center) during the 5-month recruitment period, and 5 declined participation when they were fully informed about the study. It is possible that direct contact with professionals, instead of the general call that was made through center directors, would have improved the recruitment rate to some extent. Among the 28 randomized participants, 5 more did not start the trial and 3 did not complete the study. The high workload, a common situation in the Spanish public health system even in a nonpandemic context, was the main reported cause of these withdrawals. On the other hand, the group format enriches the training process by enabling the interaction of professionals, but it also represents a difficulty when coordinating their schedules and availability. In summary, the participation and retention rates were not satisfactory, and for future trials it is necessary to develop more structured and intensive strategies. Theoretical frameworks as proposed by Solberg [41] that identified 7 factors that influence the recruitment of health care professionals (ie, relationships, reputation, requirements, rewards, reciprocity, resolution, and respect) could help to this aim.

Regarding effectiveness, the results showed significant moderate-to-strong effects (although with very wide confidence intervals) on the total scale and the CE and SC subscales. The pre-post change in the intervention group was greater on the former, and the similar between-group effect size was due in part to a significant decrease in SC in the control group. The EPS dimension was not affected by the intervention, but this result is unclear given the low internal consistency of this subscale (future studies should confirm the factorial structure of the instrument). After the RCT was completed, the control group received the intervention and showed significant before-after improvements of similar magnitude in the 3 dimensions. Due to the wide confidence intervals, the results should be interpreted with caution and verified in studies with greater statistical power.

Baseline scores indicated a positive attitude (values above the midpoint of the scale) for the total scale and the 3 subscales, although scores on CE and SC suggest that, comparatively, participants seemed more favorable to empathetically communicate with their patients than sharing decisions with them. This result has also been observed in several studies that applied the Patient - Practitioner Orientation Scale [42], the most frequently used instrument to assess health professionals' attitudes toward PCC, showing higher scores on the *caring* subscale of the questionnaire (ie, empathy, warmth, and treating patients as whole persons) than on the *sharing* one (ie, sharing information, decisions, and power) [43-47].

Other studies also have shown significant benefits of different training programs on professionals' and medical students' attitudes toward SDM and PCC and their intention to apply it in the future, showing high levels of satisfaction with the program [48-52]. A positive attitude toward the PCC model is an obvious requisite for the professionals' learning and demonstration of behaviors aimed at promoting SDM in consultation. Validation studies with the Patient - Practitioner Orientation Scale showed that more favorable attitudes were significantly associated with more patient - centered behaviors in consultations [53], and that concordance of patients and physicians' attitudes was associated with greater patient's satisfaction [53-55], trust, and endorsement of physicians [53], as well as fewer referrals to specialized care [56]. Nonetheless, for the implementation of SDM it is necessary to have not only a positive attitude toward PCC but also the appropriate knowledge and communication skills required by this model, for which training programs have been developed. However, the effect of interventions targeting health professionals on the actual promotion of SDM in consultation remains uncertain. The last update of a Cochrane systematic review reported a significant effect of these interventions (eg, educational meetings and materials, outreach visits, and reminders), compared with usual care when SDM in consultation was assessed by external observers, but not by patients, even when the intervention is directed to both patients and professionals [11]. Observational studies have also shown a lack of association between patients' and external observers' perception of SDM [57-59], but the causes of this discrepancy have not been investigated. Furthermore, the evidence about the effects of SDM interventions targeting health professionals on patients' cognitive, affective, behavioral, and health outcomes is also scarce [10].

Although the PCC and SDM models are a paradigm to be applied to every patient regardless of his or health problems, patients with GAD could present specific psychological characteristics that might affect the decision-making process. In experimental settings involving stimulus reinforcement, these patients have shown greater intolerance to uncertainty and impaired decision-making [55,57-59]. Nonetheless, this does not translate into a preference for a passive role in decision-making, since a recent study showed that more than 80% research participants desired to play an active or collaborative role when making decisions about treatment, although one-third of them perceived more involvement than they preferred [60]. Therefore, professionals should adapt the SDM process to the patients' preference for involvement and manage the unavoidable uncertainty about the potential adverse effects of treatment and the likelihood and intensity of symptoms' improvement.

Limitations

The study has important limitations. First, feasibility of in-person group sessions could not be evaluated due to the emergence of the COVID-19 pandemic, but that allowed us to check the web-based application of the program, which was delivered without problems. However, the recruitment and retention rates were low. The recruited sample was small and there were some relevant differences in baseline variables,

XSL•FO
RenderX

including the scores on the LATCon, and therefore a high risk of selection bias is present. The intervention group was 5 years younger and less experienced, included more nurses and less participants with prior experience on SDM training, and showed a less favorable attitude toward SDM. These characteristics suggest a greater margin for potential benefit in this group. Although the inclusion of these covariates in the model did not change the results, this analysis is strongly underpowered. Nonetheless, given the strong effects sizes obtained and the similar ones showed by the control group after receiving the training, it is reasonable to think that the intervention could produce a real improvement in attitudes, although effects sizes are probably inflated due to the mentioned confounders. The small sample size and the fact that participants were voluntary also challenges the external validity of the results, since it is probable that they were more motivated or favorable to the SDM model.

On the other side, this was a pilot study and we did not assess other professionals' outcomes (eg, knowledge of SDM, satisfaction with the program, and intention to apply SDM in the future), whether the observed effect is maintained over time or its influence on professionals' behavior in consultation as well as on patients' outcomes, which is the ultimate aim of these interventions. An RCT with an adequate sample size is warranted to confirm the results on professionals' attitude and to investigate the mentioned issues.

Acknowledgments

The authors would like to thank all those who contributed to the realization of this study: health managers, health care professionals, and all the researchers who made this study possible.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Results of the models including covariates on the LATCon total score.
[PDF File, 62 KB - mededu_v11i1e50060_app1.pdf]

Checklist 1.
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File, 961 KB - mededu_v11i1e50060_app2.pdf]

References

1. Depression and Other Common Mental Disorders: Global Health Estimates: World Health Organization; 2017.
2. Bados López A. Trastorno de ansiedad generalizada: naturaleza, evaluación y tratamiento. Generalized anxiety disorder: nature, assessment and treatment [Article in Spanish]. : University of Barcelona; 2017 URL: https://diposit.ub.edu/dspace/handle/2445/115724 [accessed 2024-12-12]
3. Diagnostic and Statistical Manual of Mental Disorders, 5th edition: American Psychiatric Association; 2013.
4. Parmentier H, García-Campayo J, Prieto R. Comprehensive review of generalized anxiety disorder in primary care in Europe. Curr Med Res Opin 2013 Apr;29(4):355-367. [doi: 10.1185/03007995.2013.770731] [Medline: 23356728]
5. Ramos-García V, Rivero-Santana A, Duarte-Díaz A, et al. Shared decision-making and information needs among people with generalized anxiety disorder. Eur J Investig Health Psychol Educ 2021 May 21;11(2):423-435. [doi: 10.3390/ejihpe11020031] [Medline: 34708821]
6. Lidsaar Powell RC, Bu SMK. Paternering with and involving patients. In: Leslie R, Martin M, DiMatteo R, editors. The Oxford Handbook of Health Communication: Oxford University Press; 2013:84-108.
7. Stiggelbout AM, Van der Weijden T, De Wit MPT, et al. Shared decision making: really putting patients at the centre of healthcare. BMJ 2012 Jan 27;344:e256. [doi: 10.1136/bmj.e256] [Medline: 22286508]
8. Makoul G, Clayman ML. An integrative model of shared decision making in medical encounters. Patient Educ Couns 2006 Mar;60(3):301-312. [doi: 10.1016/j.pec.2005.06.010] [Medline: 16051459]
9. Tilburt JC, Wynia MK, Montori VM, et al. Shared decision-making as a cost-containment strategy: US physician reactions from a cross-sectional survey. BMJ Open 2014 Jan 14;4(1):e004027. [doi: 10.1136/bmjopen-2013-004027] [Medline: 24430879]
10. Goldwag J, Marsicovetere P, Scalia P, et al. The impact of decision aids in patients with colorectal cancer: a systematic review. BMJ Open 2019 Sep 12;9(9):e028379. [doi: 10.1136/bmjopen-2018-028379] [Medline: 31515416]
11. Légaré F, Adekpedjou R, Stacey D, et al. Interventions for increasing the use of shared decision making by healthcare professionals. Cochrane Database Syst Rev 2018 Jul 19;7(7):CD006732. [doi: 10.1002/14651858.CD006732.pub4] [Medline: 30025154]

XSL•FO
RenderX

12. Geiger F, Hacke C, Potthoff J, et al. The effect of a scalable online training module for shared decision making based on flawed video examples—a randomized controlled trial. Patient Educ Couns 2021 Jul;104(7):1568-1574. [doi: 10.1016/j.pec.2020.11.033] [Medline: 33334633]
13. Tan ASL, Mazor KM, McDonald D, et al. Designing shared decision-making interventions for dissemination and sustainment: can implementation science help translate shared decision making into routine practice? MDM Policy Practice 2018 Jul;3(2):2381468318808503. [doi: 10.1177/2381468318808503]
14. Mahmood S, Hazes JMW, Veldt P, et al. The development and evaluation of personalized training in shared decision-making skills for rheumatologists. J Rheumatol 2020 Feb;47(2):290-297. [doi: 10.3899/jrheum.180780] [Medline: 30936289]
15. Ritter S, Stirnemann J, Breckwoldt J, et al. Shared decision-making training in internal medicine: a multisite intervention study. J Grad Med Educ 2019 Aug;11(4 Suppl):146-151. [doi: 10.4300/JGME-D-18-00849] [Medline: 31428272]
16. Légaré F, Witteman HO. Shared decision making: examining key elements and barriers to adoption into routine clinical practice. Health Aff (Millwood) 2013 Feb;32(2):276-284. [doi: 10.1377/hlthaff.2012.1078]
17. Hernández-Leal MJ, Pérez-Lacasta MJ, Feijoo-Cid M, Ramos-García V, Carles-Lavila M. Healthcare professionals' behaviour regarding the implementation of shared decision-making in screening programmes: a systematic review. Patient Educ Couns 2021 Aug;104(8):1933-1944. [doi: 10.1016/j.pec.2021.01.032] [Medline: 33581968]
18. Frerichs W, Hahlweg P, Müller E, Adis C, Scholl I. Shared decision-making in oncology—a qualitative analysis of healthcare providers' views on current practice. PLoS One 2016;11(3):e0149789. [doi: 10.1371/journal.pone.0149789] [Medline: 26967325]
19. Coates D, Clerke T. Training interventions to equip health care professionals with shared decision-making skills: a systematic scoping review. J Contin Educ Health Prof 2020;40(2):100-119. [doi: 10.1097/CEH.0000000000000289] [Medline: 32433322]
20. Diouf NT, Menear M, Robitaille H, Painchaud Guérard G, Légaré F. Training health professionals in shared decision making: update of an international environmental scan. Patient Educ Couns 2016 Nov;99(11):1753-1758. [doi: 10.1016/j.pec.2016.06.008] [Medline: 27353259]
21. Légaré F, Politi MC, Drolet R, et al. Training health professionals in shared decision-making: an international environmental scan. Patient Educ Couns 2012 Aug;88(2):159-169. [doi: 10.1016/j.pec.2012.01.002] [Medline: 22305195]
22. Kienlin S, Nytrøen K, Stacey D, Kasper J. Ready for shared decision making: pretesting a training module for health professionals on sharing decisions with their patients. J Eval Clin Pract 2020 Apr;26(2):610-621. [doi: 10.1111/jep.13380] [Medline: 32114700]
23. Pollard S, Bansback N, Bryan S. Physician attitudes toward shared decision making: a systematic review. Patient Educ Couns 2015 Sep;98(9):1046-1057. [doi: 10.1016/j.pec.2015.05.004] [Medline: 26138158]
24. Légaré F, Stacey D, Brière N, et al. Healthcare providers' intentions to engage in an interprofessional approach to shared decision-making in home care programs: a mixed methods study. J Interprof Care 2013 May;27(3):214-222. [doi: 10.3109/13561820.2013.763777] [Medline: 23394265]
25. Fisher A, Sharpe L, Anderson J, Manicavasagar V, Juraskova I. Development and pilot of a decision-aid for patients with bipolar II disorder and their families making decisions about treatment options to prevent relapse. PLoS One 2018;13(7):e0200490. [doi: 10.1371/journal.pone.0200490] [Medline: 29990368]
26. Tlach L, Wüsten C, Daubmann A, Liebherz S, Härter M, Dirmaier J. Information and decision-making needs among people with mental disorders: a systematic review of the literature. Health Expect 2015 Dec;18(6):1856-1872. [doi: 10.1111/hex.12251] [Medline: 25145796]
27. Hurtado MM, Villena A, Vega A, Amor G, Gómez C, Morales-Asencio JM. "I have anxiety, but I have values and preferences" experiences of users with generalized anxiety disorder: a qualitative study. Int J Ment Health Nurs 2020 Jun;29(3):521-530. [doi: 10.1111/inm.12690] [Medline: 31908140]
28. Scholl I, Zill JM, Härter M, Dirmaier J. An integrative model of patient-centeredness—a systematic review and concept analysis. PLoS One 2014;9(9):e107828. [doi: 10.1371/journal.pone.0107828] [Medline: 25229640]
29. Stewart M, Brown JB, Weston WW, McWhinney IR, McWilliam CL, Freeman TR. Patient-Centered Medicine: Transforming the Clinical Method: Sage Publisher; 1995.
30. Weston WW, Brown JB, Stewart MA. Patient-centred interviewing part I: understanding patients' experiences. Can Fam Physician 1989 Jan;35:147-151. [Medline: 21253278]
31. Elwyn G, Durand MA, Song J, et al. A three-talk model for shared decision making: multistage consultation process. BMJ 2017 Nov 6;359:j4891. [doi: 10.1136/bmj.j4891] [Medline: 29109079]
32. Knapp P, Raynor DK, Thistlethwaite JE, Jones MB. A questionnaire to measure health practitioners' attitudes to partnership in medicine taking: LATCon II. Health Expect 2009 Jun;12(2):175-186. [doi: 10.1111/j.1369-7625.2009.00545.x] [Medline: 19538648]
33. de Las Cuevas C, Rivero-Santana A, Perestelo-Perez L, et al. Mental health professionals' attitudes to partnership in medicine taking: a validation study of the Leeds Attitude to Concordance Scale II. Pharmacoepidemiol Drug Saf 2012 Feb;21(2):123-129. [doi: 10.1002/pds.2240] [Medline: 21956875]

XSL•FO
RenderX

34. He W, Bonner A, Anderson D. Translation and psychometric properties of the Chinese version of the Leeds Attitudes to Concordance II scale. BMC Med Inform Decis Mak 2015 Aug 1;15:60. [doi: 10.1186/s12911-015-0184-0] [Medline: 26232245]
35. Hsieh FY, Lavori PW, Cohen HJ, Feussner JR. An overview of variance inflation factors for sample-size calculation. Eval Health Prof 2003 Sep;26(3):239-257. [doi: 10.1177/0163278703255230]
36. Rowe L. Comparing learning outcomes and student and instructor perceptions of a simultaneous online versus in-person biochemistry laboratory course. J Chem Educ 2024 Mar 12;101(3):882-891. [doi: 10.1021/acs.jchemed.3c00571] [Medline: 38495613]
37. Holmes G, Clacy A, Hamilton A, Kõlves K. Online versus in-person gatekeeper suicide prevention training: comparison in a community sample. J Ment Health 2024 Oct;33(5):605-612. [doi: 10.1080/09638237.2024.2332811] [Medline: 38602188]
38. Iyer P, Mok V, Sehmbi AS, et al. Online versus in-person surgical near-peer teaching in undergraduate medical education during the COVID-19 pandemic: a mixed-methods study. Health Sci Rep 2024 Feb;7(2):e1889. [doi: 10.1002/hsr2.1889] [Medline: 38357488]
39. Lelutiu-Weinberger C, Clark KA, Pachankis JE. Mental health provider training to improve LGBTQ competence and reduce implicit and explicit bias: a randomized controlled trial of online and in-person delivery. Psychol Sex Orientat Gend Divers 2023 Dec;10(4):589-599. [doi: 10.1037/sgd0000560] [Medline: 38239562]
40. Bos-van den Hoek DW, van Laarhoven HWM, Ali R, et al. Blended online learning for oncologists to improve skills in shared decision making about palliative chemotherapy: a pre-posttest evaluation. Support Care Cancer 2023 Mar;31(3):184. [doi: 10.1007/s00520-023-07625-6]
41. Solberg LI. Recruiting medical groups for research: relationships, reputation, requirements, rewards, reciprocity, resolution, and respect. Implement Sci 2006 Oct 26;1:25. [doi: 10.1186/1748-5908-1-25] [Medline: 17067379]
42. Krupat E, Hiam CM, Fleming MZ, Freeman P. Patient-centeredness and its correlates among first year medical students. Int J Psychiatry Med 1999;29(3):347-356. [doi: 10.2190/DVCQ-4LC8-NT7H-KE0L] [Medline: 10642908]
43. Ishikawa H, Son D, Eto M, Kitamura K, Kiuchi T. Changes in patient-centered attitude and confidence in communicating with patients: a longitudinal study of resident physicians. BMC Med Educ 2018 Jan 25;18(1):20. [doi: 10.1186/s12909-018-1129-y] [Medline: 29370796]
44. Mudiyanse RM, Pallegama RW, Jayalath T, Dharmaratne S, Krupat E. Translation and validation of patient-practitioner orientation scale in Sri Lanka. Educ Health (Abingdon) 2015;28(1):35-40. [doi: 10.4103/1357-6283.161847] [Medline: 26261112]
45. Wang J, Zou R, Fu H, Qian H, Yan Y, Wang F. Measuring the preference towards patient-centred communication with the Chinese-revised Patient-Practitioner Orientation Scale: a cross-sectional study among physicians and patients in clinical settings in Shanghai, China. BMJ Open 2017 Sep 18;7(9):e016902. [doi: 10.1136/bmjopen-2017-016902] [Medline: 28928188]
46. Ahmad W, Ashraf H, Talat A, et al. Association of burnout with doctor-patient relationship and common stressors among postgraduate trainees and house officers in Lahore-a cross-sectional study. PeerJ 2018;6:e5519. [doi: 10.7717/peerj.5519] [Medline: 30221087]
47. Akkafi M, Sajadi HS, Sajadi ZS, Krupat E. Attitudes toward patient-centered care in the mental care services in Isfahan, Iran. Community Ment Health J 2019 Apr;55(3):548-552. [doi: 10.1007/s10597-018-0357-2] [Medline: 30535891]
48. Körner M, Ehrhardt H, Steger AK, Bengel J. Interprofessional SDM train-the-trainer program “Fit for SDM”: provider satisfaction and impact on participation. Patient Educ Couns 2012 Oct;89(1):122-128. [doi: 10.1016/j.pec.2012.04.008] [Medline: 22647558]
49. Leblanc A, Légaré F, Labrecque M, et al. Feasibility of a randomised trial of a continuing medical education program in shared DECISION-making on the use of antibiotics for acute respiratory infections in primary care: the DECISION+ pilot trial. Implement Sci 2011 Jan 18;6:5. [doi: 10.1186/1748-5908-6-5] [Medline: 21241514]
50. Volk RJ, Shokar NK, Leal VB, et al. Development and pilot testing of an online case-based approach to shared decision making skills training for clinicians. BMC Med Inform Decis Mak 2014 Nov 1;14:95. [doi: 10.1186/1472-6947-14-95] [Medline: 25361614]
51. Rusiecki J, Schell J, Rothenberger S, Merriam S, McNeil M, Spagnoletti C. An innovative shared decision-making curriculum for internal medicine residents. Acad Med 2018;93:937-942. [doi: 10.1097/ACM.0000000000001967]
52. Hoffmann TC, Bennett S, Tomsett C, Del Mar C. Brief training of student clinicians in shared decision making: a single-blind randomized controlled trial. J Gen Intern Med 2014 Jun;29(6):844-849. [doi: 10.1007/s11606-014-2765-5] [Medline: 24481686]
53. Shaw WS, Woiszwillo MJ, Krupat E. Further validation of the Patient-Practitioner Orientation Scale (PPOS) from recorded visits for back pain. Patient Educ Couns 2012 Nov;89(2):288-291. [doi: 10.1016/j.pec.2012.07.017] [Medline: 22954491]
54. Krupat E, Bell RA, Kravitz RL, Thom D, Azari R. When physicians and patients think alike: patient-centered beliefs and their impact on satisfaction and trust. J Fam Pract 2001 Dec;50(12):1057-1062. [Medline: 11742607]
55. Krupat E, Yeager CM, Putnam S. Patient role orientations, doctor-patient fit, and visit satisfaction. Psychol Health 2000 Sep;15(5):707-719. [doi: 10.1080/08870440008405481]

XSL•FO
RenderX

56. Carlsen B, Aakvik A, Norheim OF. Variation in practice: a questionnaire survey of how congruence in attitudes between doctors and patients influences referral decisions. Med Decis Making 2008;28(2):262-268. [doi: 10.1177/0272989X07311751] [Medline: 18349435]
57. Diendéré G, Farhat I, Witteman H, Ndjaboue R. Observer ratings of shared decision making do not match patient reports: an observational study in 5 family medicine practices. Med Decis Making 2021 Jan;41(1):51-59. [doi: 10.1177/0272989X20977885] [Medline: 33371802]
58. Evong Y, Chorney J, Ungar G, Hong P. Perceptions and observations of shared decision making during pediatric otolaryngology surgical consultations. J Otolaryngol Head Neck Surg 2019 Jun 17;48(1):28. [doi: 10.1186/s40463-019-0351-x] [Medline: 31208462]
59. Williams D, Edwards A, Wood F, et al. Ability of observer and self-report measures to capture shared decision-making in clinical practice in the UK: a mixed-methods study. BMJ Open 2019 Aug 18;9(8):e029485. [doi: 10.1136/bmjopen-2019-029485] [Medline: 31427333]
60. Meterko M, Wright S, Lin H, Lowy E, Cleary PD. Mortality among patients with acute myocardial infarction: the influences of patient-centered care and evidence-based medicine. Health Serv Res 2010 Oct;45(5 Pt 1):1188-1204. [doi: 10.1111/j.1475-6773.2010.01138.x] [Medline: 20662947]

Abbreviations

CE: communication/empathy
EPS: eventual paternalistic style
GAD: generalized anxiety disorder
LATConII: Leeds Attitudes Towards Concordance scale, second version
PC: primary care
PCC: person-centered care
RCT: randomized controlled trial
SC: shared control
SDM: shared decision-making

Edited by B Lesselroth; submitted 18.06.23; peer-reviewed by C Timm, J Rubel; revised version received 27.07.24; accepted 19.08.24; published 16.01.25.

Please cite as:
Ramos-García V, Rivero-Santana A, Peñate-Castro W, Álvarez-Pérez Y, Duarte-Díaz A, Torres-Castaño A, Trujillo-Martín MDM, González-González AI, Serrano-Aguilar P, Perestelo-Pérez L
A Brief Web-Based Person-Centered Care Group Training Program for the Management of Generalized Anxiety Disorder: Feasibility Randomized Controlled Trial in Spain
JMIR Med Educ 2025;11:e50060
URL: https://mededu.jmir.org/2025/1/e50060
doi:10.2196/50060

© Vanesa Ramos-García, Amado Rivero-Santana, Wenceslao Peñate-Castro, Yolanda Álvarez-Pérez, Andrea Duarte-Díaz, Alezandra Torres-Castaño, María del Mar Trujillo-Martín, Ana Isabel González-González, Pedro Serrano-Aguilar, Lilisbeth Perestelo-Pérez. Originally published in JMIR Medical Education (https://mededu.jmir.org), 16.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring Connections Between Mental Health, Burnout, and Academic Factors Among Medical Students at an Iranian University: Cross-Sectional Questionnaire Study

Elham Faghihzadeh[1], PhD; Ali Eghtesad[2], MD; Muhammad Fawad[3], PhD; Xiaolin Xu[3], PhD

[1]Department of Biostatistics and Epidemiology, Zanjan University of Medical Sciences, Zanjan, Iran

[2]School of Medicine, Zanjan University of Medical Sciences, Zanjan, Iran

[3]School of Public Health, The Second Affiliated Hospital, Zhejiang University School of Medicine, 866 Yuhangtang Road, Hangzhou, Zhejiang, China

Corresponding Author:
Muhammad Fawad, PhD
School of Public Health, The Second Affiliated Hospital, Zhejiang University School of Medicine, 866 Yuhangtang Road, Hangzhou, Zhejiang, China

Abstract

Background: Medical students face high levels of burnout and mental health issues during training. Understanding associated factors can inform supportive interventions.

Objective: This study aimed to examine burnout, psychological well-being, and related demographics among Iranian medical students.

Methods: A cross-sectional survey was conducted among 131 medical students at an Iranian University. The instruments used included the Maslach Burnout Inventory-Student Survey and the Symptom Checklist-90-Revised. Descriptive statistics, multivariate regression, and tests for group differences were used to analyze the data.

Results: The MBI-SS subscale scores indicated moderate emotional exhaustion, mean 15.00 (SD 7.08) and academic efficacy, mean 14.98 (SD 6.29), with lower cynicism, mean 10.85 (SD 5.89). The most commonly reported mental health issues were depression and obsessive-compulsive disorder. Poor psychological well-being was associated with higher overall burnout, but no significant gender differences were found. Burnout levels varied by academic year across all Maslach Burnout Inventory-Student Survey domains.

Conclusions: Despite their health education, medical students in this study reported significant burnout and mental health distress, with strong associations between the two. These issues may impact student retention and post-graduation practice plans. Supporting well-being during training is critical for positive student and physician outcomes.

(JMIR Med Educ 2025;11:e58008) doi:10.2196/58008

KEYWORDS

emotional exhaustion; exhaustion; cynicism; academic efficacy; burnout; physician burnout; mental health; mental illness; mental disease; mental disorder; medical education; medical knowledge; medical training; medical student; resident physician; resident doctor; residency; residency training

Introduction

Kary and Pines [1] initially posited the concept of academic tedium and its impact on students. They suggested that this phenomenon is not confined to a specific educational level but can manifest at various stages of schooling, including both school and university environments [1]. Based on the author's viewpoint, students might be struggling with a condition marked by a fading enthusiasm for learning, noticeable lack of motivation, and an overwhelming sense of emotional exhaustion. Later, Maslach and Jackson [2,3] specified burnout as the experience of physical and emotional drain caused by chronic stress. Burnout is the state of physical and mental fatigue caused by work, study, or any caregiving activities. It can also be known as an adverse emotional, cognitive, and physical reaction to the study, work, and life pressures. Burnout was officially classified by the World Health Organization as an occupational phenomenon in 2019 and included in the International Classification of Diseases (*ICD-11*).

Educational burnout is a type of burnout experienced during studying. To better view educational burnout, it was expanded to three factors: emotional exhaustion, cynicism, and feelings of inefficacy [4,5]. Emotional exhaustion reflects feelings of exceeding emotional resources due to academic demands. Cynicism is a negative, unresponsive, or overly snapped response to a phenomenon. Feelings of inefficacy refer to a

XSL•FO
RenderX

reduction in academic effort, leading to a sense of incompetence and reduced academic achievement. Based on findings from a systematic review published in 2021, it was determined that educational burnout affected more than 40% of students [6]. This outcome implies a heightened susceptibility to burnout among medical students on a global scale. However, few studies have examined this problem, specifically among medical students in Iran. While 16% of Iranian medical students reported burnout in one study [7], assessing prevalence rates at individual universities could further inform supportive programs. Educational burnout has an essential role in medical students' overall health and could easily impact the quality of their learning [8].

A study on 14,000 students from different countries showed that approximately 35% of the students had been diagnosed with at least one mental health disorder, such as depression or anxiety [9]. Among students, university students showed a higher likelihood of mental health disorders, and among them, medical students' issues were significant. Medical schools pose multiple demands on students. First, enrollment in medical training coincides with adolescence and early adulthood, periods already associated with vulnerability to mental health disorders [10-12]. Second, the intense nature of medical education requires students to assimilate vast amounts of health information while coping with exposure to myriad diseases [11,12]. Consequently, studies report substantial rates of depression (11% - 37%), anxiety (7.4% - 30%), and other issues in this population internationally [13-15]. Evidence suggests that positive mental health aids coping [16], yet remains understudied in Iranian cultures.

Extensive evidence demonstrates intricate connections between burnout and mental health issues among medical students. Additional studies reveal substantially higher risks of depression, anxiety, suicidal ideation, concentration deficits, and physical symptoms compared to their peers [12,17-20]. Up to half of graduating students experience burnout, linking this syndrome to exacerbated mental health decline [18]. Ultimately, these concerning rates significantly exceed general population trends, underscoring the crisis of psychological well-being in medical education. Implementing supportive interventions requires further investigating specific student populations.

The aims of this study are twofold. Primarily, we assess the prevalence of mental health issues and burnout among native Iranian medical students at Zanjan University of Medical Sciences. Additionally, we delineate connections between mental health status and burnout risk by evaluating associated academic and personal factors. By understanding these relationships, targeted interventions can eventually be developed to promote the psychological well-being of Iran's future physicians during their demanding training period.

Methods

Study Design and Participants

This cross-sectional study was conducted at Zanjan University of Medical Sciences, Zanjan, Iran, focusing on the experiences of 1500 medical students. These trainees constituted the target of our research, with their perspectives and characteristics as students comprising the central subject of investigation. Participants were recruited using a convenience sampling method. Our research team directly contacted the students, explained the study's aims, invited their voluntary participation, and emphasized the confidentiality of their responses. We then sent an electronic survey link to consenting participants. Strict data quality control measures were implemented, with incomplete questionnaire submissions excluded from the analysis to uphold the integrity of the results. Based on a previous study [21], the minimum required sample size was 120 students; however, 140 surveys were distributed, and 131 fully completed questionnaires were returned. Those with missing data or students who indicated having a diagnosed mental health issue were excluded.

Measures

Demographics

The basic sociodemographic information included age, sex, residence, history of a positive COVID-19 test, underlying diseases, diagnosis of mental health issues, and level of education. The levels of education were categorized into three sections: the initial seven semesters, referred to as preclinical, followed by a two-semester externship, and finally, a three-semester internship. At the preclinical level, students learned about basic sciences and pathophysiology; in the externship phase, they would pass a short course in each hospital unit. The residential status comprises a parental home, independent home, or a dormitory. Students were asked directly about underlying diseases, including diabetes, hypertension, and chronic disease. Additionally, they were asked whether they had been diagnosed with any mental health condition and whether they had received any treatment.

Burnout Measurement

Burnout symptoms were measured by the Persian version of the MBI-SS [3,22,23]. It comprises 15 items, which are divided into three dimensions: emotional exhaustion, cynicism, and academic efficacy . Each item has been rated on a 7-pointed Likert scale. Academic efficacy scores were reverse-coded; therefore, it was scored oppositely. A high score in three dimensions indicated greater burnout. The maximum possible scores for emotional exhaustion, cynicism, and academic efficacy were 30, 24, and 36, respectively.

Mental Health Measurement

The Symptom Checklist 90 (SCL-90), developed by Derogatis was used to assess mental health. This scale consists of 90 items, each rated on a 5-point Likert scale, effectively measuring ten primary psychological symptoms [24]. The ten psychological symptoms measured by the Symptom Checklist-90-Revised (SCL-90-R) are somatization, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, psychoticism, and sleep problems. If a person's average score for the questions related to these symptoms was greater than 2, it indicated potential psychological issues. The Global Severity Index (GSI) was calculated for our analysis, which measures the extent or depth of psychiatric disturbances. Specifically, the GSI is the average

XSL•FO
RenderX

score across all responded items and serves as an overall measure of psychiatric distress. Therefore, this study analyzed the positive rate of each subscale and the GSI. Notably, the validated Persian version of the SCL-90-R was used for this student population [25].

Statistical Analysis

We performed all statistical analyses using SPSS (version 20.0; IBM Corp) and Stata (version 12; StataCorp), and figures were drawn using R software (version 4.4.1; R Foundation for Statistical Computing) was used for visualization, including the ggplot2 package (version 3.5.1). Multivariable regression was used to assess factors associated with the three burnout subscales to examine the correlation between the response variables. The model included sociodemographic variables such as history of COVID-19, place of residence, and mental health status. The variables that were found to have a significant impact on the outcome were retained in the model.

Ethical Considerations

Ethics approval for this study was provided by the Ethics Committee of Zanjan University of Medical Sciences (IR.ZUMS.REC.1400.418). This committee approved all experimental protocols. The authors confirmed that relevant guidelines and regulations were used in all experiments. No participants were younger than 16 years. All students provided written informed consent.

Results

The study initially involved 140 students; after excluding those who dropped out due to missing answers or diagnosed mental illness, the remaining sample consisted of 131 participants. The average age of these remaining participants was approximately 24 years, with a mean of 23.95 (SD 3.69) years. Table 1 summarizes other sociodemographic characteristics of the student group. Approximately 66% were female students, while only 10% of the participants were married. An almost equal percentage of students across different academic levels completed the questionnaires. Additionally, 62% of the students had a history of a positive COVID-19 test, while 3.8% reported underlying diseases.

Table . Socio-demographic characteristics.

Sociodemographic variables	Participants (N=131), n (%)
Sex	
Male	44 (33.6)
Female	87 (66.4)
Marital status	
Single	118 (90.1)
Married	13 (9.9)
Residence	
Parental home	53 (40.5)
Own home	43 (32.8)
Dormitory	35 (26.7)
Positive COVID-19 test history	
No	50 (38.2)
Yes	81 (61.8)
Underlying diseases	
No	126 (96.2)
Yes	5 (3.8)
Academic level	
Preclinical	42 (32.1)
Externship	47 (35.9)
Internship	42 (32.1)

Table 2 shows the positive rates of SCL-90-R subscales by sex. Obsessive-compulsive disorder and depression showed the highest prevalence among symptoms. A χ^2 test examined the percentage differences between male and female students. The only symptom found to be statistically significant between the two sex was phobic anxiety. Among female students, paranoid ideation had the highest prevalence, whereas obsessive-compulsive disorder was more prevalent among male students.

XSL•FO
RenderX

Table . Comparison of SCL-90-R[a] subscales based on sex.

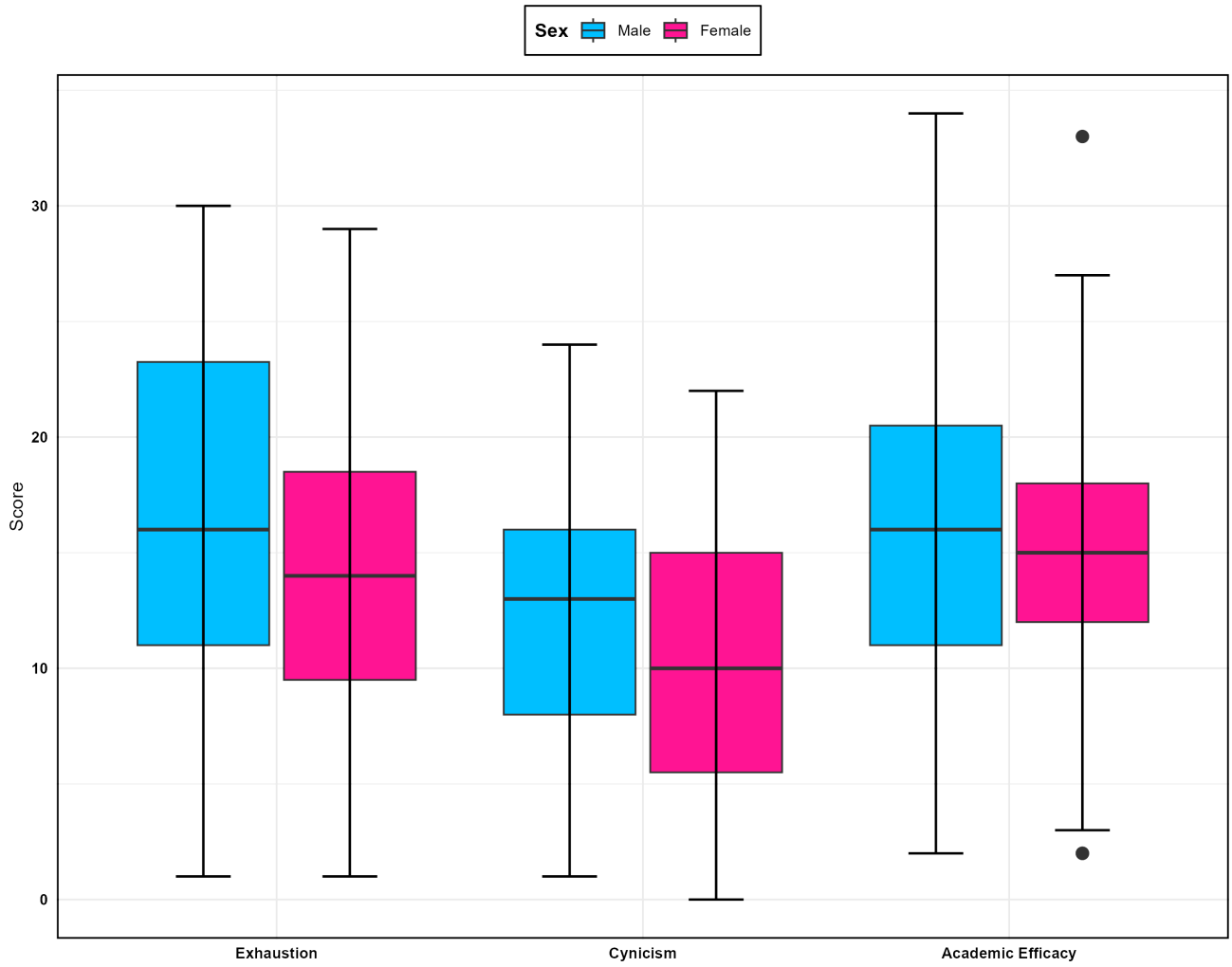
	SCL-90-R positive rates in male students n (%)	SCL-90-R positive rates in female students, n (%)	Total SCL-90-R positive rate, n (%)
Hostility	10 (22.7)	13 (14.9)	23 (17.6)
Anxiety	11 (25.0)	14 (16.1)	25 (19.1)
Obsessive-compulsive disorder	14 (31.8)	18 (20.7)	32 (24.4)
Interpersonal sensitivity	10 (22.7)	19 (21.8)	29 (22.1)
Somatization	6 (13.6)	12 (13.8)	18 (13.7)
Psychoticism	5 (11.4)	7 (8.0)	12 (9.2)
Paranoid ideation	8 (18.2)	21 (24.1)	29 (22.1)
Depression	12 (27.3)	20 (23.0)	32 (24.4)
Phobic anxiety[b]	9 (20.5)	7 (8.0)	16 (12.2)
Others	8 (18.2)	11 (12.6)	19 (14.5)

[a]SCL-90-R: Symptom Checklist-90-Revised.

[b]*P* value (χ^2 test)=.04 male versus female.

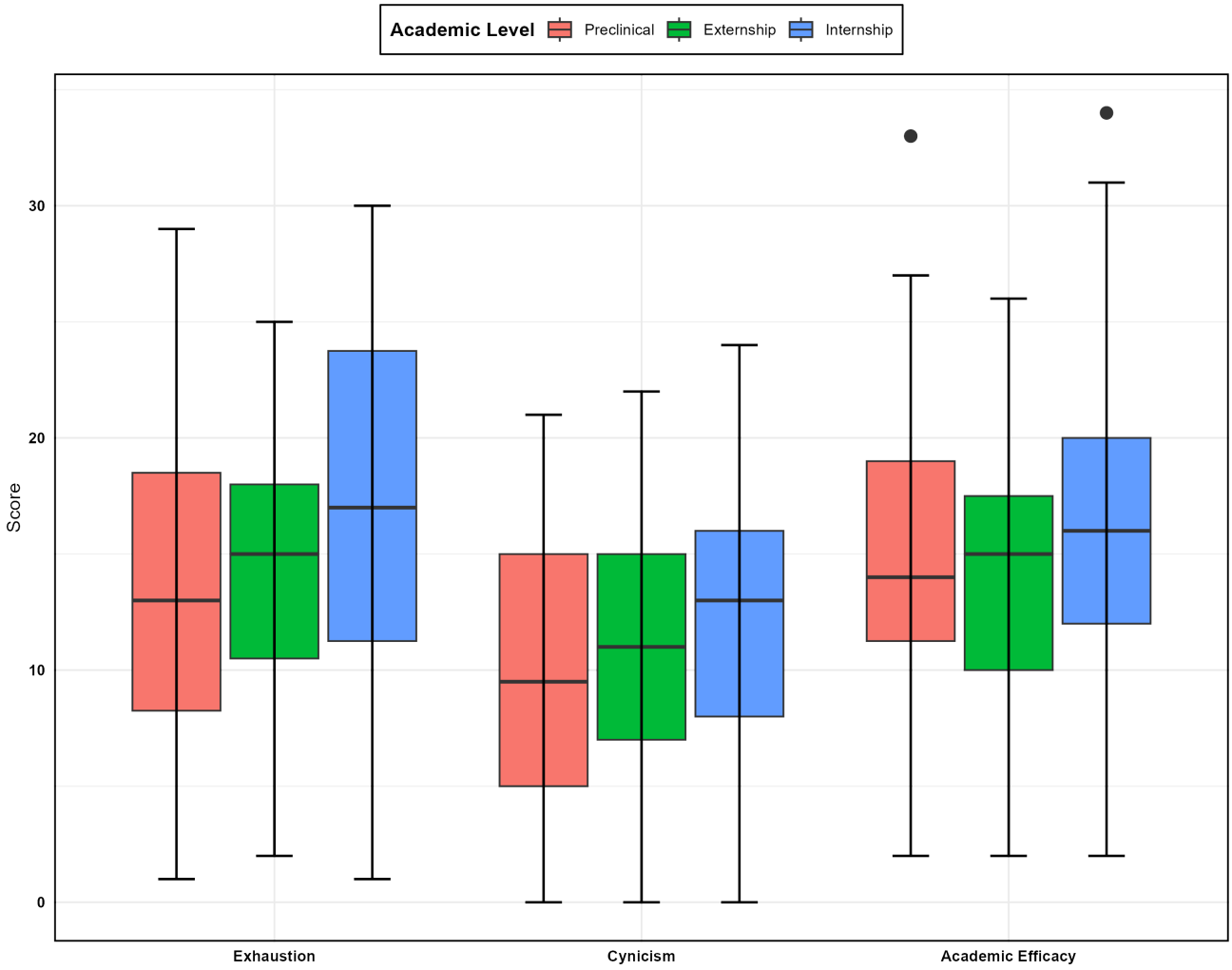
The boxplots in Figures 1 and 2 display the MBI-SS subscale scores across genders and academic levels. According to this figure, academic efficacy had the widest range of scores for male students. Additionally, female students exhibited lower mean scores compared to male students across all subscales. These boxplots indicate that interns had higher burnout scores overall. More detailed descriptive statistics can be found in Multimedia Appendix 1. Figure 3 shows a comparison of the total scores on the SCL-90-R between different levels, revealing that externs had the highest scores.

Figure 1. Comparison of Maslach Burnout Inventory-Student Survey subscale scores by sex.



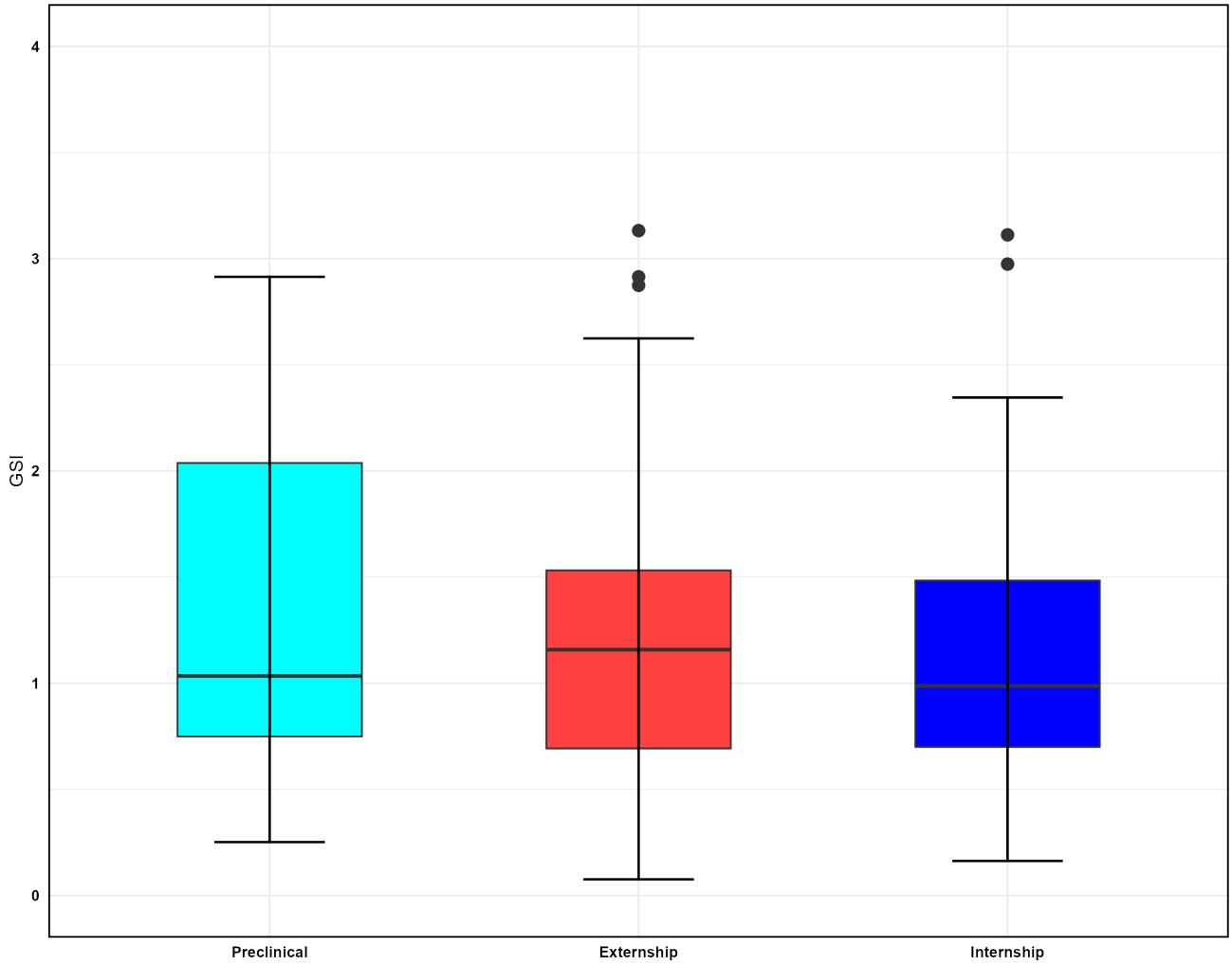
XSL•FO
RenderX

Figure 2. Comparison of Maslach Burnout Inventory-Student Survey subscale scores across academic levels.



XSL•FO
RenderX

Figure 3. Comparison of Global Severity Index scores across academic levels.



Initially, the correlation between the three dimensions of burnout (academic efficacy, cynicism, and emotional exhaustion) was examined in the modeling data. The findings revealed that academic efficacy had a significant correlation with both cynicism (r=0.41, P<0.05) and emotional exhaustion (r=0.37, P<0.05). Additionally, emotional exhaustion positively correlated with cynicism (r=0.78, P<0.05). Given these significant correlations among the burnout dimensions, a multivariable regression analysis was deemed appropriate for further modeling. Table 3 presents the results of the multivariable regression analysis. An increase of one score in GSI corresponded to an increase of 5.67, 1.71, and 4.69 scores in emotional exhaustion, cynicism, and academic efficacy, respectively. Overall, the students in the internship phase has 4.19 and 3.02 scores higher than preclinical students in emotional exhaustion and academic efficacy, respectively, whereas they had only a 0.24-difference in cynicism. Furthermore, males scored 1.53 and 0.10 points lower than female students, respectively. A comparison of β coefficients shows that GSI and internship status had significantly different effects on the three dimensions of the MBI-SS.

XSL•FO
RenderX

Table . The association of educational burnout with mental well-being, academic level, and sex.

	Emotional exhaustion				Cynicism				Academic Efficacy				Equality of β coefficients	
	β coefficient	95% CI	*F* value	*P* value	β coefficient	95% CI	*F* value	*P* value	β coefficient	95% CI	*F* value	*P* value	*F* value	*P* value
GSI[a]	5.67	4.02-7.32	6.55	<.001	1.71	1.41-2.00	5.56	<.001	4.69	3.31-6.07	4.79	<.001	18.72	<.001
Academic levels														
Preclinical	-	-	-		-	-	-		-	-	-			
Externship	1.17	−1.83 to 4.18	0.68	.44	−0.32	−0.86 to 0.21	−1.00	.23	0.56	−1.95 to 3.08	0.52	.66	0.59	.49
Internship	4.19	0.97-7.41	2.41	.01	−0.24	−0.81 to 0.33	−1.98	.40	3.02	0.33-5.71	3.10	.03	2.01	.04
Sex														
Female	-	-	-		-	-	-		-	-	-			
Male	−1.53	−4.25 to 1.20	−1.29	.27	−0.10	−0.58 to 0.38	-1.40	.69	2.34	0.06-4.61	2.02	.04	0.99	.21
R^2	0.303				0.523				0.513					

[a]GSI: Global Severity Index.

Discussion

Principal Findings

Zanjan University, a prominent institution in Iran, attracts students from various cities. Therefore, studying its students' mental and physical well-being can provide valuable insights into the overall condition of Iranian students.

Our study examined crucial mental health issues such as hostility, obsessive-compulsive disorder, and interpersonal sensitivity among medical students. In brief, our study did not detect any statistically significant differences in overall mental health scores between students across different academic level or sex. However, phobic anxiety was the only mental health issue that was significantly different between genders. Students at the externship level had higher GSI scores. This increase in mental health problems among the students is understandable, as it occurred during their clinical rotations in hospitals, where they were exposed to diverse patient cases and experienced various illnesses for the first time in their academic careers. Facing such novel and potentially challenging situations can reasonably be expected to take an emotional toll.

In previous studies, depression, stress, and anxiety are the three most prevalent mental health issues among medical students [7,11-13,15,26,27]. Cuttilan et al [13], who reviewed studies from Asia in their meta-analysis, showed that 30% of Middle Eastern students experienced depression. While our study found a slightly lower prevalence, the difference could be caused by the university environment, sample size, and social or climatic differences. Nonetheless, the rate of depression remains considerable. Aghajani Liasi et al [7], who studied the prevalence of burnout and mental health at one of Tehran's universities, reported a 37% of depression among medical students. They used the Depression, Anxiety, and Stress Scale questionnaire to survey mental health. Therefore, despite being of the same nationality, the main reason for the variation between the findings of our study compared to those reported by Aghajani Liasi et al may be due to differences in the questionnaires.

Anxiety is one of the significant issues experienced by medical students. A systematic review revealed the wide range of anxiety prevalence across different countries (15.5%-70.0%) [27]. Our study found an anxiety disorder prevalence of 19% among students at Zanjan University of Medical Sciences; this places our sampled student population in the lower quartile of anxiety rates compared to the broader range reported by medical trainees across the world.

Unlike the questionnaire used in our study (ie, SCL-90-R), many previous studies on student stress have used the DASS scale and reported high rates of stress among students. For example, a meta-analysis found that 52.7% of medical students reported significant stress during training [13]. Additionally, studies by Aghajani Liasi et al [7] and Moutinho et al [28] reported stress rates of approximately 30% and 47%, respectively within their student samples, despite the differences in study populations. While these percentages vary, these studies collectively highlight that clinically significant stress is a widespread and impactful issue for many students across educational contexts [7,28]. However, our study did not directly measure student stress, which is a limitation compared to previous existing research.

Our study showed that although there was no statistically significant difference in burnout scores between male and female students, female students reported lower burnout levels in the three burnout subscales compared to male students. Additionally, students in later years of medical education reported higher burnout levels than students in the initial phases. This indicates

XSL•FO
RenderX

that the interns about to graduate showed higher burnout levels, especially feeling emotionally drained.

The relationship between years of medical education and burnout levels is interesting. While some studies have suggested that burnout levels may increase with advancing years of medical education due to prolonged exposure to stressors, the evidence remains inconclusive [29,30]. This suggests that the intense pressures of medical school take an cumulative toll.

Prior studies have found that medical students experience some of the highest rates of burnout compared to other populations [4,6,11,20]. However, findings regarding the relationship between gender and burnout have been mixed [29,31-33]. There was more evidence suggesting that male students are more likely to face burnout than female students [6]. Therefore, it can be concluded that the relationship between gender and burnout in medical students may be influenced by various factors such as the specific population, sample sizes, and the definition of burnout used in the research.

Our study explored several potential influencing factors on the three burnout dimensions in medical students, including mental health status, gender, and academic level. These variables significantly impacted emotional exhaustion, cynicism, and academic efficacy scores. To date, no study had directly examined the linkage between mental health disorders and burnout in this population, representing a gap in understanding. However, related research by Dyrbye et al [16] showed associations between positive well-being and professionalism, which burnout may undermine. Additionally, psychologists have suggested that students with psychiatric conditions demonstrate greater emotional exhaustion [17,18,34]. Notably, in our analysis, mental health had a much more significant effect on emotional exhaustion compared to the other burnout facets. Other studies found students with higher burnout reported more suicidal thoughts and behaviors [6,17,18,26,27,34,35]. Integrating those findings with our results suggests that mental health could play an intermediary role between burnout and suicidal risks. These interrelationships between wellness, distress, and functioning highlight the need for more holistic support to promote student resilience.

Limitations

This study has some limitations. The cross-sectional design cannot determine causal relationships between variables. Additionally, the convenience sampling and voluntary participation could indicate that students with psychological issues may have been less inclined to take part or answer honestly. While different variables were recorded, others such as physical activity, social support, and economic status, should be investigated in future studies. Longitudinal follow-up studies warrant a better understanding of mental health's impact on burnout trajectories.

Another limitation was the rate of female participation compared to male participants for two reasons. According to the university's annual statistics, about 55% of students are women, increasing the female sample rate in the convenience sampling method. However, among the Iranian population, women are more interested in psychological issues and experience exhaustion about improving mental health, resulting in more women participation in our study.

An additional limitation is the potential link between financial issues, mental health, and burnout. Our survey did not include detailed questions about participants' financial situations, which could have influenced their responses to other questions. To partially address this, we included a question about place of residence, which could indirectly reflect financial circumstances and their possible effects on other survey responses.

Conclusion

The demanding nature of academic work and personal lives faced by medical students can take a severe mental toll, leading to burnout. Despite being educated on physical and psychological health, students often neglect their own well-being. This research confirms that mental health issues directly contribute to students' emotional exhaustion, cynicism, and reduced feelings of academic self-efficacy. Both burnout and psychological problems increase the risk of students dropping out or deciding against careers as general practitioners after graduation, resulting in wasted resources invested in their training. Most alarmingly, if the society cannot ensure the mental well-being of its future doctors, the overall population's health will suffer consequences. There is an undeniable connection between medical trainees' health and the communities they will serve. Fostering resilience and coping abilities in students must be a key priority, as their personal health and capacity to provide quality patient care in the future hinges on it. The findings of this study highlight the prevalence of burnout and mental health issues among medical students, underscoring the profound importance of addressing this problem for the well-being of the general population, who will rely on these future physicians for care.

Acknowledgments

The authors express their gratitude to Prof. Xiaolin Xu for his contributions as a senior author in guiding the research and to Dr. Omid Saed for his suggestion for the questionnaires, Department of Clinical Psychology, Zanjan University of Medical Sciences.

Data Availability

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: EF

XSL•FO
RenderX

Data curation: AE
Formal analysis: EF, AE
Funding acquisition: XX, MF
Methodology: EF, MF
Project administration: XX
Resources: EF
Supervision: MF, XX
Visualization: EF, MF
Writing – original draft: EF
Writing – review & editing: EF, MF, XX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Descriptive statistics of exhaustion, cynicism, academic efficacy across sex and academic level.
[DOCX File, 16 KB - mededu_v11i1e58008_app1.docx]

References

1. Kafry D, Pines A. The experience of Tedium in life and work. Human Relations 1980 Jul;33(7):477-503. [doi: 10.1177/001872678003300703]
2. Maslach C, Jackson SE. The measurement of experienced burnout. J Organ Behavior 1981 Apr;2(2):99-113. [doi: 10.1002/job.4030020205]
3. Maslach C. Burnout: a multidimensional perspective. In: Professional Burnout: Routledge; 2018:19-32. [doi: 10.4324/9781315227979-3]
4. Abreu Alves S, Sinval J, Lucas Neto L, Marôco J, Gonçalves Ferreira A, Oliveira P. Burnout and dropout intention in medical students: the protective role of academic engagement. BMC Med Educ 2022 Feb 7;22(1):83. [doi: 10.1186/s12909-021-03094-9] [Medline: 35130892]
5. Schaufeli WB, Martínez IM, Pinto AM, Salanova M, Barker AB. Burnout and engagement in university students a cross-national study. J Cross Cult Psychol 2002;33:464-481. [doi: 10.1177/0022022102033005003]
6. Rosales-Ricardo Y, Rizzo-Chunga F, Mocha-Bonilla J, Ferreira JP. Prevalence of burnout syndrome in university students: a systematic review. Salud Ment 2021 Apr 9;44(2):91-102. [doi: 10.17711/SM.0185-3325.2021.013]
7. Aghajani Liasi G, Mahdi Nejad S, Sami N, Khakpour S, Ghorbani Yekta B. The prevalence of educational burnout, depression, anxiety, and stress among medical students of the Islamic Azad University in Tehran, Iran. BMC Med Educ 2021 Sep 5;21(1):471. [doi: 10.1186/s12909-021-02874-7] [Medline: 34482821]
8. Zis P, Artemiadis A, Bargiotas P, Nteveros A, Hadjigeorgiou GM. Medical studies during the COVID-19 pandemic: the impact of digital learning on medical students' burnout and mental health. Int J Environ Res Public Health 2021 Jan 5;18(1):349. [doi: 10.3390/ijerph18010349] [Medline: 33466459]
9. Auerbach RP, Alonso J, Axinn WG, et al. Mental disorders among college students in the World Health Organization World Mental Health Surveys - CORRIGENDUM. Psychol Med 2017 Nov;47(15):2737. [doi: 10.1017/S0033291717001039] [Medline: 28462760]
10. Gerber M, Lang C, Feldmeth AK, et al. Burnout and mental health in Swiss vocational students: the moderating role of physical activity. J of Research on Adolesc 2015 Mar;25(1):63-74. [doi: 10.1111/jora.12097]
11. Wassif GO, Gamal-Eldin DA, Boulos DNK. Stress and burnout among medical students. Journal of High Institute of Public Health 2019 Dec 5;0:189-197. [doi: 10.21608/jhiph.2019.63794]
12. Rotenstein LS, Ramos MA, Torre M, et al. Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: a systematic review and meta-analysis. JAMA 2016 Dec 6;316(21):2214-2236. [doi: 10.1001/jama.2016.17324] [Medline: 27923088]
13. Cuttilan AN, Sayampanathan AA, Ho RCM. Mental health issues amongst medical students in Asia: a systematic review [2000-2015]. Ann Transl Med 2016 Feb;4(4):72. [doi: 10.3978/j.issn.2305-5839.2016.02.07] [Medline: 27004219]
14. Healy C, Ryan Á, Moran CN, Harkin DW, Doyle F, Hickey A. Medical students, mental health and the role of resilience - a cross-sectional study. Med Teach 2023 Jan;45(1):40-48. [doi: 10.1080/0142159X.2022.2128735] [Medline: 36214365]
15. Zeng W, Chen R, Wang X, Zhang Q, Deng W. Prevalence of mental health problems among medical students in China. Medicine (Abingdon) 2019;98(18):e15337. [doi: 10.1097/MD.0000000000015337] [Medline: 31045774]
16. Dyrbye LN, Harper W, Moutier C, et al. A multi-institutional study exploring the impact of positive mental health on medical students' professionalism in an era of high burnout. Acad Med 2012 Aug;87(8):1024-1031. [doi: 10.1097/ACM.0b013e31825cfa35] [Medline: 22722352]

XSL•FO
RenderX

17. Morcos G, Awan OA. Burnout in medical school: a medical student's perspective. Acad Radiol 2023 Jun;30(6):1223-1225. [doi: 10.1016/j.acra.2022.11.023] [Medline: 36586757]
18. Klein HJ, McCarthy SM. Student wellness trends and interventions in medical education: a narrative review. Humanit Soc Sci Commun 2022;9(1):92. [doi: 10.1057/s41599-022-01105-8] [Medline: 36254165]
19. Hu W, Yu Z, Liang X, Abulaiti A, Aini X, Kelimu A. A cross-sectional study on the analysis of the current situation of depression and anxiety among primary and secondary school students in Urumqi City in 2021: a case study of S district. J Affect Disord 2024 Feb 15;347:210-219. [doi: 10.1016/j.jad.2023.11.079] [Medline: 37995925]
20. Dyrbye LN, West CP, Satele D, et al. Burnout among U.S. medical students, residents, and early career physicians relative to the general U.S. population. Acad Med 2014 Mar;89(3):443-451. [doi: 10.1097/ACM.0000000000000134] [Medline: 24448053]
21. Koutsimani P, Montgomery A, Georganta K. The relationship between burnout, depression, and anxiety: a systematic review and meta-analysis. Front Psychol 2019;10:284. [doi: 10.3389/fpsyg.2019.00284] [Medline: 30918490]
22. Rostami Z, Abedi MR, Schuffli VB. Standardization of Maslash burnout inventory among female students at University of Isfahan. New Educational Approaches 2011;6:21-38 [FREE Full text]
23. Maslach C, Jackson SE, Leiter MP. Evaluating stress: a book of resources. In: Maslach Burnout Inventory, th edition: Scarecrow Education; 1997.
24. Derogatis LR, Melisaratos N. The Brief Symptom Inventory: an introductory report. Psychol Med 1983 Aug;13(3):595-605. [Medline: 6622612]
25. Akhavan Abiri F, Shir Mohammadi R. Validity and reliability of Symptom Checklist-90-revised (SCL-90-R) and Brief Symptom Inventory-53 (BSI-53). Clinical Psychology and Personality 2019;17:169-195 [FREE Full text]
26. Fares J, Al Tabosh H, Saadeddin Z, El Mouhayyar C, Aridi H. Stress, burnout and coping strategies in preclinical medical students. N Am J Med Sci 2016 Feb;8(2):75-81. [doi: 10.4103/1947-2714.177299] [Medline: 27042604]
27. Mirza AA, Baig M, Beyari GM, Halawani MA, Mirza AA. Depression and anxiety among medical students: a brief overview. Adv Med Educ Pract 2021;12:393-398. [doi: 10.2147/AMEP.S302897] [Medline: 33911913]
28. Moutinho ILD, Maddalena NDC, Roland RK, et al. Depression, stress and anxiety in medical students: a cross-sectional comparison between students from different semesters. Rev Assoc Med Bras (1992) 2017 Jan 1;63(1):21-28. [doi: 10.1590/1806-9282.63.01.21] [Medline: 28225885]
29. Almutairi H, Alsubaiei A, Abduljawad S, et al. Prevalence of burnout in medical students: a systematic review and meta-analysis. Int J Soc Psychiatry 2022 Sep;68(6):1157-1170. [doi: 10.1177/00207640221106691] [Medline: 35775726]
30. Grech M. The effect of the educational environment on the rate of burnout among postgraduate medical trainees - a narrative literature review. J Med Educ Curric Dev 2021;8:23821205211018700. [doi: 10.1177/23821205211018700] [Medline: 34104789]
31. Galán F, Sanmartín A, Polo J, Giner L. Burnout risk in medical students in Spain using the Maslach Burnout Inventory-Student Survey. Int Arch Occup Environ Health 2011 Apr;84(4):453-459. [doi: 10.1007/s00420-011-0623-x] [Medline: 21373879]
32. Backović DV, Zivojinović JI, Maksimović J, Maksimović M. Gender differences in academic stress and burnout among medical students in final years of education. Psychiatr Danub 2012 Jun;24(2):175-181 [FREE Full text] [Medline: 22706416]
33. Briggs LG, Riew GJ, Kim NH, et al. Racial and gender differences in medical student burnout: a 2021 national survey. Mayo Clin Proc 2023 May;98(5):723-735. [doi: 10.1016/j.mayocp.2022.11.003] [Medline: 37137644]
34. Ishak W, Nikravesh R, Lederer S, Perry R, Ogunyemi D, Bernstein C. Burnout in medical students: a systematic review. Clin Teach 2013 Aug;10(4):242-245. [doi: 10.1111/tct.12014] [Medline: 23834570]
35. Frajerman A, Morvan Y, Krebs MO, Gorwood P, Chaumette B. Burnout in medical students before residency: a systematic review and meta-analysis. Eur Psychiatry 2019 Jan;55:36-42. [doi: 10.1016/j.eurpsy.2018.08.006] [Medline: 30384110]

Abbreviations

GSI: Global Severity Index
MBI-SS: Maslach Burnout Inventory-Student Survey
SCL-90: Symptom Checklist-90
SCL-90-R: Symptom Checklist-90-Revised

XSL•FO
RenderX

Edited by B Lesselroth; submitted 02.03.24; peer-reviewed by A Hassan, CI Sartorão Filho, S Kale; revised version received 22.06.24; accepted 23.11.24; published 15.05.25.

Please cite as:
Faghihzadeh E, Eghtesad A, Fawad M, Xu X
Exploring Connections Between Mental Health, Burnout, and Academic Factors Among Medical Students at an Iranian University: Cross-Sectional Questionnaire Study
JMIR Med Educ 2025;11:e58008
URL: https://mededu.jmir.org/2025/1/e58008
doi:10.2196/58008

© Elham Faghihzadeh, Ali Eghtesad, Muhammad Fawad, Xiaolin Xu. Originally published in JMIR Medical Education (https://mededu.jmir.org), 15.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Barriers to and Facilitators of Implementing Team-Based Extracorporeal Membrane Oxygenation Simulation Study: Exploratory Analysis

Joan Brown[1], EdD, MBA, CCE; Sophia De-Oliveira[2], MPH; Christopher Mitchell[1,3], BSN, RN, CCRN; Rachel Carmen Cesar[4], PhD; Li Ding[4], MD; Melissa Fix[1,3], BSN, RN, CCRN-CMC; Daniel Stemen[5], MSRS, RRT-ACCS, E-AEC; Krisda Yacharn[1,3], BSN, RN, CNML; Se Fum Wong[6], MD; Anahat Dhillon[7], MD

[1]Department of Surgery, Keck School of Medicine, University of South California, Los Angeles, CA, United States

[2]Office of Performance & Transformation, Keck Hospital of USC, Los Angeles, CA, United StatesUnited States

[3]Department of Nursing, Keck Hospital of USC, Los Angeles, CA, United StatesUnited States

[4]Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, United StatesUnited States

[5]Deparment of Respiratory Therapy, Keck Hospital of USC, Los Angeles, CA, United StatesUnited States

[6]Department of Anesthesia, Kaiser Permanante, Los Angeles, CA, United States

[7]Department of Anesthesia Critical Care Medicine, Keck School of Medicine, University of South California, Los Angeles, CA, United StatesUnited States

Corresponding Author:
Joan Brown, EdD, MBA, CCE
Department of Surgery, Keck School of Medicine, University of South California, Los Angeles, CA, United States

Abstract

Introduction: Extracorporeal membrane oxygenation (ECMO) is a critical tool in the care of severe cardiorespiratory dysfunction. Simulation training for ECMO has become standard practice. Therefore, Keck Medicine of the University of California (USC) holds simulation-training sessions to reinforce and improve providers knowledge.

Objective: This study aimed to understand the impact of simulation training approaches on interprofessional collaboration. We believed simulation-based ECMO training would improve interprofessional collaboration through increased communication and enhance teamwork.

Methods: This was a single-center, mixed methods study of the Cardiac and Vascular Institute Intensive Care Unit at Keck Medicine of USC conducted from September 2021 to April 2023. Simulation training was offered for 1 hour monthly to the clinical team focused on the collaboration and decision-making needed to evaluate the initiation of ECMO therapy. Electronic surveys were distributed before, after, and 3 months post training. The survey evaluated teamwork and the effectiveness of training, and focus groups were held to understand social environment factors. Additionally, trainee and peer evaluation focus groups were held to understand socioenvironmental factors.

Results: In total, 37 trainees attended the training simulation from August 2021 to August 2022. Using 27 records for exploratory factor analysis, the standardized Cronbach α was 0.717. The survey results descriptively demonstrated a positive shift in teamwork ability. Qualitative themes identified improved confidence and decision-making.

Conclusions: The study design was flawed, indicating improvement opportunities for future research on simulation training in the clinical setting. The paper outlines what to avoid when designing and implementing studies that assess an educational intervention in a complex clinical setting. The hypothesis deserves further exploration and is supported by the results of this study.

(JMIR Med Educ 2025;11:e57424) doi:10.2196/57424

KEYWORDS

intensive care unit; ICU; teamwork in the ICU; team dynamics; collaboration; interprofessional collaboration; simulation; simulation training; ECMO; extracorporeal membrane oxygenation; life support; cardiorespiratory dysfunction; cardiorespiratory; cardiology; respiratory; heart; lungs

XSL•FO
RenderX

Introduction

Simulation training for extracorporeal membrane oxygenation (ECMO) has become standard practice for reinforcing technical skills, facilitating troubleshooting, and building teamwork [1]. ECMO is a critical tool in the care of severe cardiorespiratory dysfunction among patients of all ages [1]. Within the intensive care unit (ICU), ECMO is one of the most complicated therapies, requiring not only extensive knowledge of cardiopulmonary physiology and expertise with intricate circuit components but also skills to rapidly respond to emergent situations [2]. Therefore, high-fidelity simulation trainings are critical to practice skills and work through different emergency scenarios, such as the blood pump falling from the drive unit [3]. A randomized control study concluded that exposure to high-fidelity simulated ECMO emergencies leads to significant improvements in technical and behavioral skills among clinicians. This study demonstrated that simulation training creates a learning environment that replicates the clinical setting and fosters acquisition of cognitive, technical, and behavioral skills [4].

The Extracorporeal Life Support Organization, an international nonprofit association of health care institutions focused on ECMO research and education, recommends simulation training didactic sessions, water drills, animal sessions, and bedside training [5]. However, a randomized controlled trial published in *Critical Care Medicine* compared traditional water drill with simulation and found that simulation-based training is more effective than traditional training [6]. Water-based drills do not offer the same hands-on experience of real-time troubleshooting, and the use of animals is expensive and complex [6]. Nevertheless, traditional and simulation-based training are both beneficial to ECMO education. The benefits of simulation training on reinforcing skills have been noted in the literature [3,7-9]. Therefore, Keck Medicine of the University of California (USC) has held ECMO simulation-training sessions since 2013 for nursing education and 2021 for interprofessional simulation to reinforce and improve providers knowledge and hands-on skills in high-risk, low-frequency scenarios at no risk to patients [6].

We implemented simulation-based ECMO training to improve interprofessional collaboration through increased communication and enhanced teamwork. Moreover, the intention of the simulation training was to strengthen collaboration skills and increase confidence in providers to work through emergency scenarios. The specific aim of the study was to understand the impact of our simulation training approach on interprofessional collaboration. However, ICU staffing models impacted the ability to execute the study design as intended. This paper outlines the original study design, the challenges the research team faced during the study, and the lessons learned to ensure future studies mitigate the challenges posed by real-world ICU operations. Our primary outcome shifted to the development and validation of measurement tools and offers recommendations for the evaluation of simulation approaches in future studies.

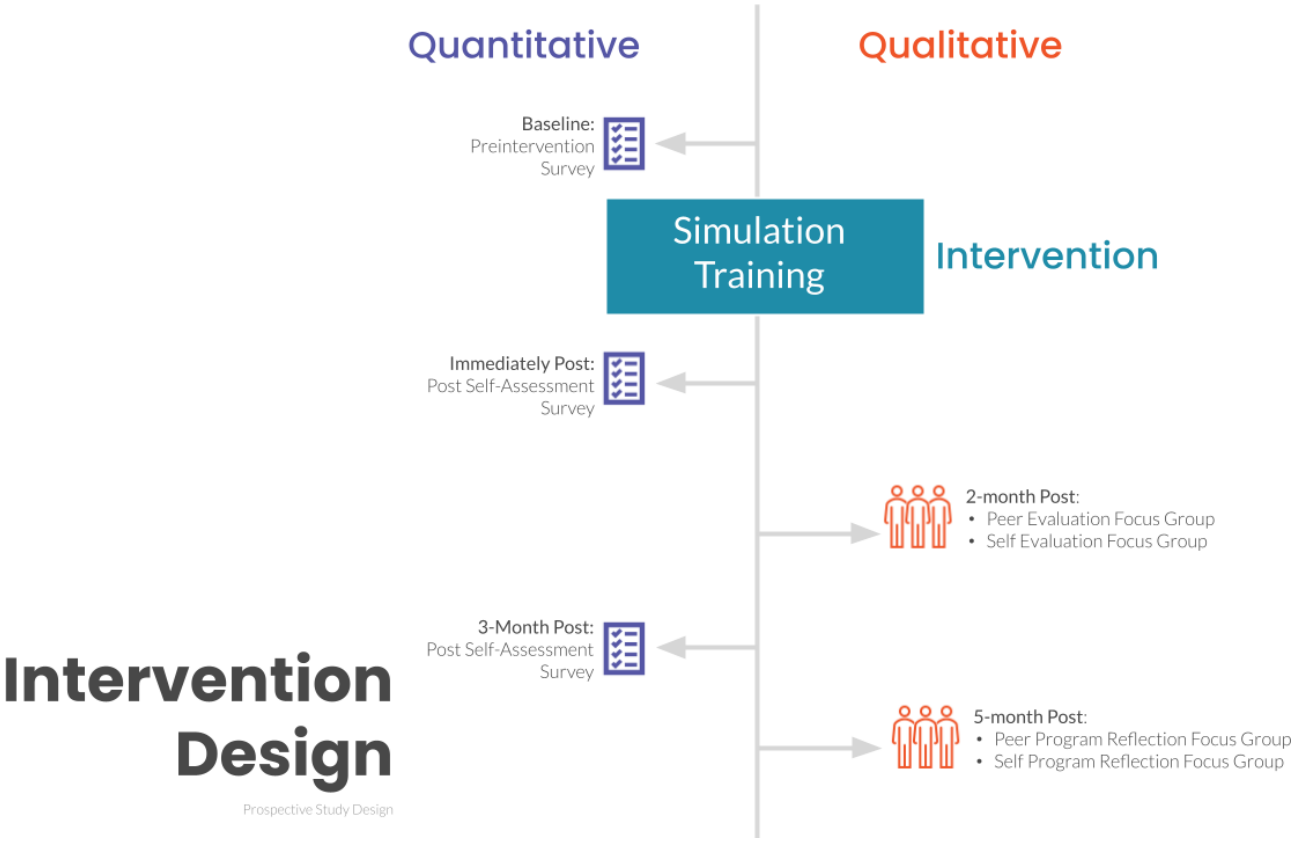
Methods

Overview

This was a single-center, mixed methods study of the Cardiac and Vascular Institute (CVI) ICU at Keck Medicine of USC conducted from September 2021 to April 2023. The study was designed to elicit quantitative feedback through an electronic survey before and post a voluntary training simulation exercise, and qualitative feedback from a participants via a series of focus groups that incorporated self- and peer evaluations (Figure 1).

XSL•FO
RenderX

Figure 1. Study design.



Participants

A census sampling strategy was used to recruit participants; in other words, all trainees that participated in the training were offered participation in the study. Participants included: (1) trainee physician fellows and residents in the CVI ICU who were offered attendance to the simulation training by their program director, and (2) peer evaluators, including the CVI ICU's Medical Director, nurse manager, nurse clinical educator, and lead respiratory therapist. Participant trainees attended a single 1-hour simulation training with roles played by clinical staff members from the CVI ICU, including an intensivist, clinical nurse educator, and respiratory therapist. To be eligible for the study, participants needed to be in a fellowship on rotation at Keck Medicine of USC. Fellows were recruited by intensivist leaders of the CVI from departments that rotated through or interacted with the CVI ICU (Pulmonary Critical Care Medicine, Cardiology, Anesthesia, Surgical Critical Care, and Cardiac Surgery). All study recruitment took place via email by the CVI Medical Director and Program Director to physicians in fellowship based on department and rotation schedule from pulmonary critical care medicine, surgical critical care medicine, cardiac surgery, and cardiology.

Simulation Training

Simulation training was designed as part of the continuing clinical education offered to the clinical team for 1-hour monthly, where participants attended a single session. Simulations were designed to focus on the interprofessional collaboration and decision-making needed to evaluate a patient for the initiation of ECMO therapy (Table S1 in Multimedia Appendix 1). Initially, low-fidelity simulations were held in a conference room using (1) a resuscitation training mannequin, (2) simulated vital signs via a hospital patient monitor connected to a rhythm simulator, (3) simulated intravenous access, (4) simulated medications, and (5) emergency equipment. In January 2022, collaboration with the Keck School of Medicine Simulation department allowed for training to be held in a simulation lab with a high-fidelity simulation mannequin and integrated simulation software LLEAP, version 8.5 from Laerdal. The availability of a higher fidelity training environment was meant to improve the training experience of the learners.

Each training session began with an orientation to the simulation environment and assigned roles. The scenario (Table S1 in Multimedia Appendix 1) was created to include relative contraindications to ECMO therapy and a potentially reversible condition that led to a cardiac arrest requiring resuscitation. Participants were assigned into roles of primary physician, code blue response provider, and cardiac surgeon prior to entering the simulation and entered the scenario when prompted by the facilitator or requested during the simulation by another participant. The patient was introduced to the learners as a 65-year-old female in-patient on a hospital cardiac telemetry unit with a past medical history of coronary artery disease, congestive heart failure, and peripheral vascular disease. The simulation began when a facilitator in the role of the patient's nurse requested help from a participant. The simulated patient was initially responsive with complaints of palpitations and shortness of breath with intermittent ventricular tachycardia displayed on the cardiac monitor. The simulated patient then became unresponsive in persistent ventricular tachycardia, and the imbedded facilitator activated the resuscitation team. When

XSL•FO
RenderX

the simulated patient's cardiac rhythm changes, the participants performed the roles of a code blue response, including coordinating the resuscitation, performing a simulated echocardiograph, and performing simulated invasive procedures including endotracheal intubation, arterial line insertion, and central line insertion. The participants collaborated to identify the candidacy of the simulated patient for ECMO therapy and proceeded to participate in a moderate-fidelity mock cannulation with ECMO training equipment. The participant in the surgeon role chose a method and site of cannulation for the simulated patient, and a practice ECMO circuit was connected to the simulator. Participants proceeded to respond and troubleshoot as the patient was set to be initially unstable during the transition to ECMO support. The simulated patient remained in ventricular tachycardia, and the participants were required to decide whether to continue attempting interventions, including, for example, chest compressions, medication, and defibrillation once the patient was placed on ECMO. The simulation ended when the patient was stabilized on ECMO and the participants decided to transfer the patient to the ICU. Areas of safety concern (Figure S1 in Multimedia Appendix 2) were emphasized in the training as points for communication to consider the decision to initiate ECMO with an unstable patient. A postsimulation debriefing session was facilitated by the simulation faculty.

Qualitative Approach

To understand the social environment factors the simulation training impacted, a total of 12 qualitative focus groups were planned (Figure 1). The 12 interviews were divided into 6 focus groups with the trainee attendees of the simulation training as a self-evaluation and 6 focus groups with colleague participants as a peer evaluation (Figure S2 in Multimedia Appendix 3). Each focus group was designed to have 2 - 4 participants. The study was designed to use the same peer evaluators for each peer evaluation focus group for the study duration. Each peer evaluation focus group was meant to target the evaluation of the individuals in the 6 simulation cohorts with a total of 4 peer participants. The focus groups were designed to be a duration of 30 minutes. Questions were developed to assess how the simulation training impacted the trainees' practice related to collaboration and teamwork (Table S3 in Multimedia Appendix 4). Questions were reviewed by the study expert in mixed methods study application.

Interviews were conducted by the simulation facilitators experienced in ECMO therapy and simulation education. Sessions were recorded using the Voice Memo application (Apple, Inc.). Once the focus groups were completed, the simulation facilitators sent the audio file to the data management author for transcription. The transcribed focus group sessions were de-identified, then uploaded and stored to a HIPAA (Health Insurance Portability and Accountability Act)-compliant Microsoft OneDrive. All audio and video files containing identifiers were deleted following transcription. Transcription documents were reviewed and coded for key themes using grounded theory methodology, an iterative process that will identify conceptual categories emerging from the comparative analyses of the data.

Quantitative Approach

An electronic survey was distributed with a QR code in person and electronically via email using Qualtrics XM software version December 2019. A total of 49 questions were posed to trainees across the pretraining, posttraining, and 3-month posttraining questionnaires (Table S2 in Multimedia Appendix 5). Teamwork-focused questions were obtained from the validated Mayo High Performance Teamwork scale (16 questions) [10]. The remaining questions regarding the effectiveness of training were devised using the Kirkpatrick Training Evaluation Framework as a basis for query design [11]. The study biostatistician performed a psychometric review to assess the validity and reliability of the survey questions. The use of existing validated tools ensured high reliability and validity of the teamwork elements of the survey tool [10]. Additionally, field tests (1 MD, 2 RNs, and 1 RT) of the survey tool showed an average survey duration of 7 minutes or less and promoted consistent comprehension of the study questions across individuals.

Exploratory factor analysis was conducted to assess questions reflecting underlying factors. The number of factors included in the final model was determined by eigenvalue and Scree plot. In the final factor pattern table, questions with a value>0.4 were considered well loaded for the factor. The Validated Mayo High Performance Teamwork scale (16 questions) was used as a sum score as recommended by the study [10]. Secondary outcomes analysis included department-level patient mortality, average device days on ECMO, decannulation percentage, and percentage of staff that had simulation training.

Ethical Considerations

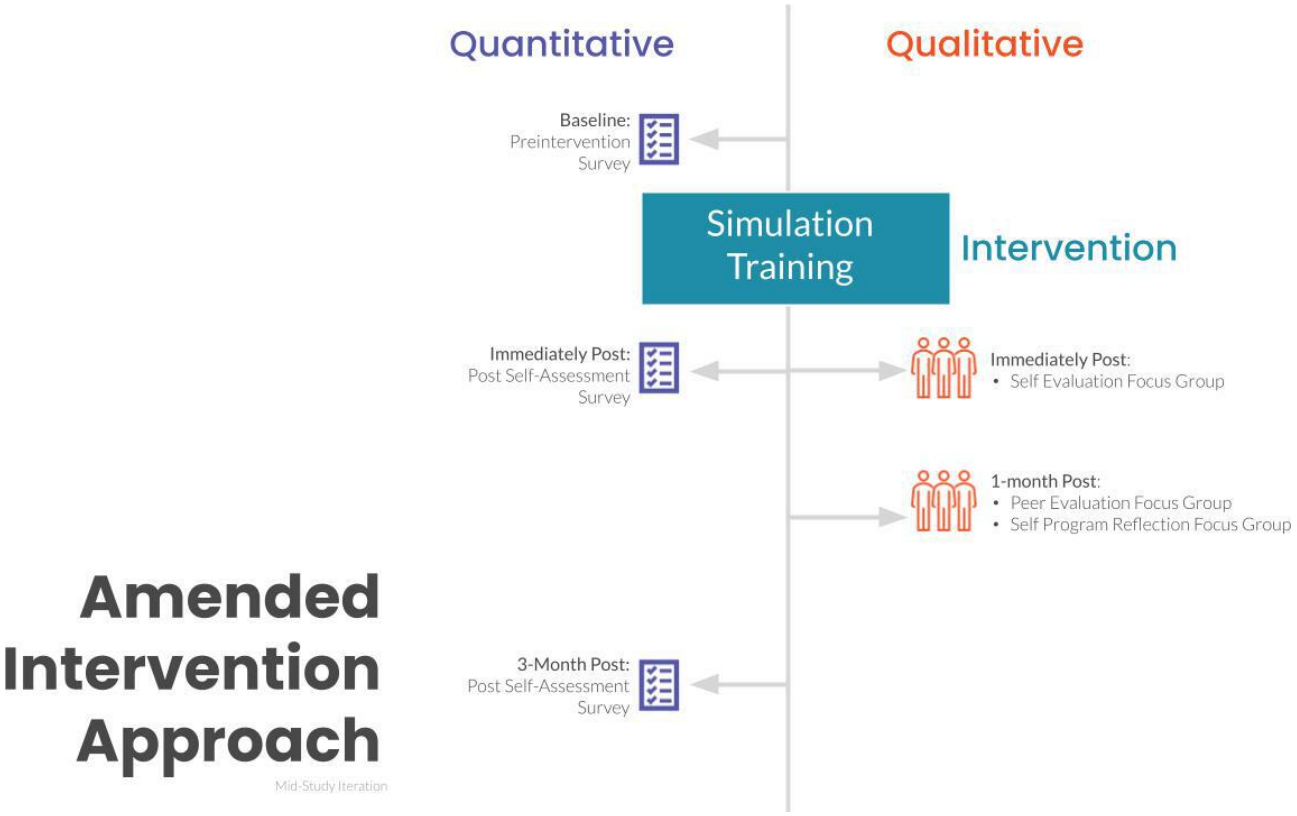
The study was approved by the University of Southern California institutional review board (UP-21-01021). Prior to participation, all study participants were required to sign an informed consent form, thereby confirming their voluntary engagement in the survey process. The study data were anonymous.

Results

A total of 37 trainees attended the training simulation from August 2021 to August 2022. However, only 7 trainees opted to participate in the qualitative portion of the study. Due to lack of participant engagement, mid-study the study design was amended to increase study participation (Figure 2). The quantitative approach remained as originally designed; however, the analysis approach was done descriptively due to the inability to compare pre- and postsurvey results on an individual basis. In other words, survey analysis aggregated all preresponses and then postresponses to compare the pregroup and postgroup responses. The qualitative approach shifted to trainee participants attending a total of 2 focus groups, an initial self-evaluation immediately following simulation and a 1-month post-program reflection if they were available (Figure S3 in Multimedia Appendix 6).

XSL•FO
RenderX

Figure 2. Amended study design.



Qualitative Approach

A total of 4 focus groups were conducted between January 2022 and August 2022 including 2 trainee self-evaluations (n=7) immediately post the training simulation and 2 peer evaluations 2-months post-training evaluation with the group of 4 peers. Focus groups for participants between August 2021 and December 2021 were not coordinated due to lack of trainee engagement in study participation. The 2 peer focus groups held highlighted an issue in trainee exposure with the peer team. Peer participants noted that they had limited clinical working exposure to the trainees being evaluated due to the nature of the fellow's rotation in the ICU. Meaning, peers did not have any recollection of working with the trainees prior to and following the simulation training to provide an appropriate evaluation of their skills in teamwork during ECMO therapy decision-making. In other words, peers remembered a trainee prior to or following the simulation, but not both. The 2 trainee self-evaluations that occurred highlighted themes that showed the simulation training benefited the trainees, including that the simulation training resulted in (1) working together as a stronger and more confident team because of simulation and (2) the creation of a space to improve communications, decision-making, and express concerns (Table S5 in Multimedia Appendix 7).

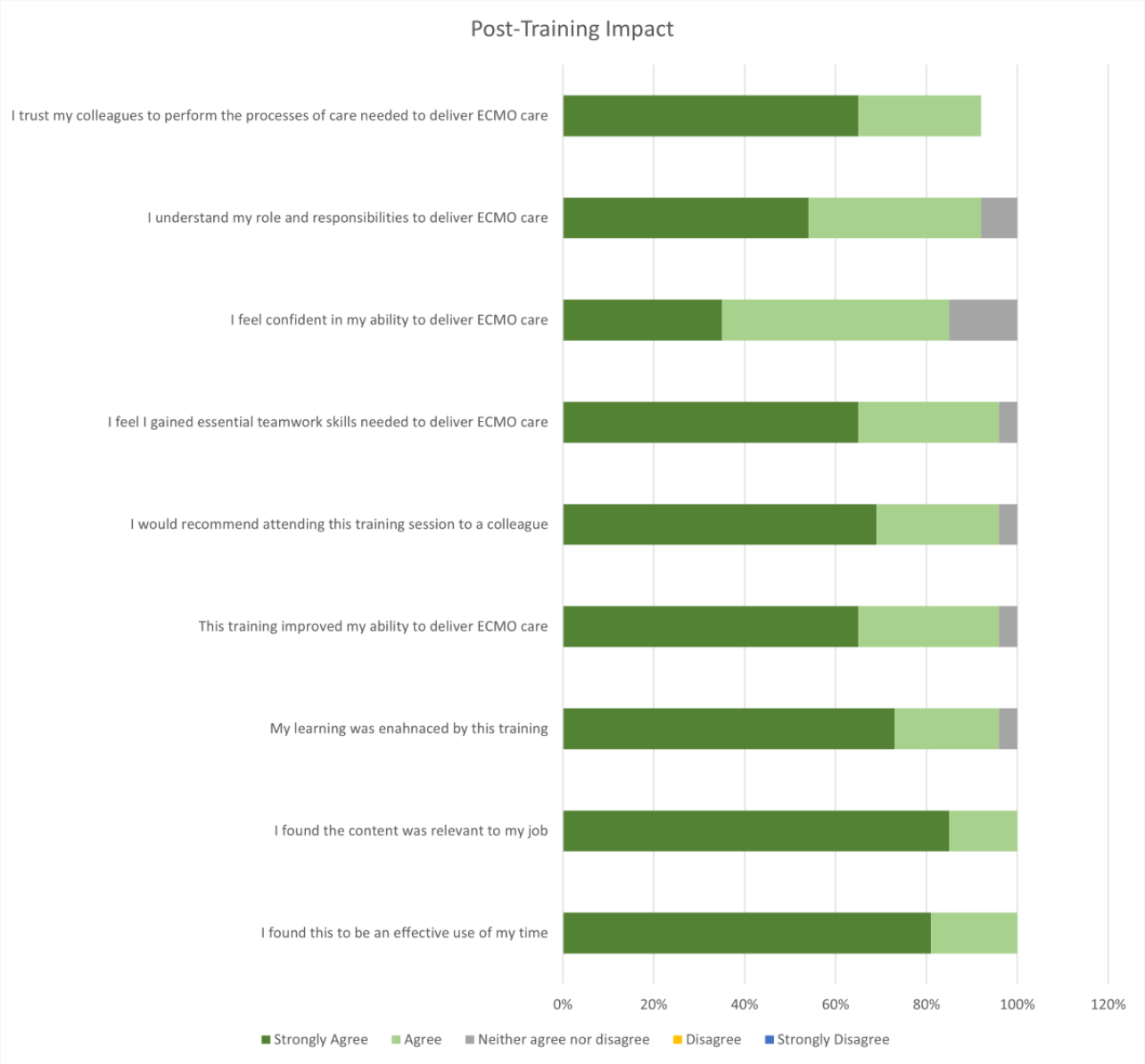
Quantitative Approach

All trainees were asked to complete the pre-, immediately post-, and 3-month surveys as part of the simulation training experience. There were a total of 37 entries recorded for the pre-survey, yielding a 100% response rate. Of those entries, 9 records were excluded due to lack of record ID or mismatching question numbers, leaving 28 entries for analysis. There was a total of 35 entries recorded for the postsurvey, yielding a 95% response rate. Following data cleaning, 9 records were excluded, leaving 26 postsurvey entries for analysis. There were a total of 2 entries for the 3-month posttraining survey, yielding a 5.4% response rate. The 3-month postsurvey results were excluded from analysis due to the low response rate. Table S4 in Multimedia Appendix 8 details the survey results of each question.

Questions posed in the post-survey focused on assessing levels 1 (reaction) and 2 (learning) of the Kirkpatrick Training Evaluation Framework demonstrated a high level of agreement for positive postsimulation training impact (Figure 3). Additionally, an increase in knowledge and understanding were noted descriptively when comparing the pre- and postresponses of the survey (Table S4 in Multimedia Appendix 8). For example, question "I understand the mechanism for activating ECMO at Keck Hospital" in the presurvey 46% of respondents agreed or strongly agreed to the statement compared to the postsimulation training survey where 100% of participants responded with a level of agreement. Levels 3 (impact) and 4 (results) of Kirkpatrick's framework were unable to be assessed due to the low response rate for the 3-month postsimulation survey.

XSL•FO
RenderX

Figure 3. Posttraining impact. ECMO: extracorporeal membrane oxygenation.



The Mayo Teamwork Scale was used to understand changes in the trainee's perspective on teamwork before and after the simulation training. The 16 focused teamwork questions demonstrated a positive shift in teamwork ability; that is, in the presurvey participants had a 71% average response on performing each question consistently, postsurvey showed an increase of the average to 95% consistently.

Exploratory Factor Analysis

An exploratory factor analysis was conducted to validate the use of the survey in evaluating the effectiveness of team training in ECMO simulation (Table 1). Entries with missing questions were excluded from the factor analysis; that is, only records with all questions answered were used. A total of 28 data points were available in the presurvey and 26 for the postsurvey. For the factor analysis, 0.4 was used as the cutoff. For the Mayo High-Performance Teamwork scale, 23 records were collected with a mean sum score of 26.87 (SD 6.41) at the presurvey and 21 records at the postsurvey with a mean sum score of 31.1 (SD 1.88). For the presurvey, we asked six 5-likelihood questions. Using 27 records for exploratory factor analysis, only 1 question, "Q5," did not reflect the underlying factor. Standardized Cronbach α was 0.686 when using all 6 questions. After excluding Q5, the standardized Cronbach α is 0.717. For the postsurvey, 26 records were used for analysis with 3 factors. Standardized Cronbach α was 0.919 when using all 15 questions. For questions from the Mayo High-Performance Teamwork scale, 21 records were used for analysis with a mean sum score of 31.1 and a SD of 1.88.

XSL•FO
RenderX

Table . Exploratory factor analysis.

		Factor 1	Factor 2	Factor 3
Presurvey				
	1. Please rate the following statements—I understand the mechanism for activating ECMO[a] at Keck Hospital	0.556	—[b]	—
	2. Please rate the following statements—I understand my role in a bedside cannulation	0.755	—	—
	3. Please rate the following statements—I feel comfortable using the ECMO equipment specific to my role	0.723	—	—
	4. Please rate the following statements—I feel comfortable using the 2-challenge rule	0.478	—	—
	5. Please rate the following statements—I feel confident to voice concerns to leadership during a critical situation	0.201	—	—
	6. Please rate the following statements—I trust my colleagues to perform the processes of care needed to deliver ECMO care	0.417	—	—
Postsurvey				
	1. Please rate the following statements—I understand the mechanism for activating ECMO at Keck Hospital	—	0.880	—
	2. Please rate the following statements—I understand my role in a bedside cannulation	—	0.602	—
	3. Please rate the following statements—I feel comfortable using the ECMO equipment specific to my role	—	—	0.412
	4. Please rate the following statements—The initiating team communicates efficiently during a bedside cannulation	0.543	0.456	—
	5. Please rate the following statements—I feel comfortable using the 2-challenge rule	—	—	0.659
	6. Please rate the following statements—I feel confident to voice concerns to leadership during a critical situation	—	0.705	—

XSL•FO
RenderX

	Factor 1	Factor 2	Factor 3
7. Please rate the following statements—I found this to be an effective use of my time	0.851	—	—
8. Please rate the following statements—I found the content was relevant to my job	0.813	—	—
9. Please rate the following statements—My learning was enhanced by this training	0.927	—	—
10. Please rate the following statements—This training improved my ability to deliver ECMO care	0.891	—	—
11. Please rate the following statements—I would recommend attending this training session to a colleague	0.844	—	—
12. Please rate the following statements—I feel I gained essential teamwork skills needed to deliver ECMO care	0.916	—	—
13. Please rate the following statements—I feel confident in my ability to deliver ECMO care	0.532	—	0.670
14. Please rate the following statements—I understand my role and responsibilities to deliver ECMO care	0.695	—	—
15. Please rate the following statements—I trust my colleagues to perform the processes of care needed to deliver ECMO care	0.788	—	—

[a]ECMO: extracorporeal membrane oxygenation.

[b]Not applicable.

Triangulation of Quantitative and Qualitative Results

Applying procedures of convergent mixed methods design, we converged the quantitative and qualitative results that were obtained separately to obtain a nuanced understanding of the core research aims. The themes identified of teamwork and improved communication in the qualitative analysis were supported by the quantitative survey results (Tables 2 and 3). Qualitative subthemes were supported by the positive shift observed descriptively from the pre- compared to the post-simulation training survey results.

XSL•FO
RenderX

Table . Triangulation of quantitative and qualitative results between frequently endorsed survey items and themes emerging from postsimulation focus groups (part).

Question		Agree level (Strongly Agree + Agree), %	Neither agree nor disagree, %	Disagree level (Strongly Disagree + Disagree),	Qualitative theme
I understand the mechanism for activating ECMO [a] at Keck Hospital					Working together as a stronger and more confident team because of simulation
	Prestimulation	46	18	36	
	Poststimulation	100	0	0	
I understand my role in a bedside cannulation					Working together as a stronger and more confident team because of simulation
	Prestimulation	23	48	30	
	Poststimulation	100	0	0	
I feel comfortable using the 2- challenge rule					Working together as a stronger and more confident team because of simulation
	Prestimulation	4	29	68	
	Poststimulation	69	15	15	
I feel confident to voice concerns to leadership during a critical situation					Working together as a stronger and more confident team because of simulation
	Prestimulation	78	11	11	
	Poststimulation	96	4	0	
I found this to be an effective use of my time (poststimulation)		100	0	0	Creating a space to improve communications, decision-making, and express concerns via simulation
I found the content was relevant to my job (poststimulation)		100	0	0	Creating a space to improve communications, decision-making, and express concerns via simulation
My learning was enhanced by this training (poststimulation)		96	4	0	Creating a space to improve communications, decision-making, and express concerns via simulation
This training improved my ability to deliver ECMO care (poststimulation)		96	4	0	Creating a space to improve communications, decision-making, and express concerns via simulation
I feel I gained essential teamwork skills needed to deliver ECMO care (poststimulation)		96	4	0	Creating a space to improve communications, decision-making, and express concerns via simulation

XSL•FO
RenderX

Question	Agree level (Strongly Agree + Agree), %	Neither agree nor disagree, %	Disagree level (Strongly Disagree + Disagree),	Qualitative theme
I feel confident in my ability to deliver ECMO care (poststimulation)	85	15	0	Creating a space to improve communications, decision-making, and express concerns via simulation
I understand my role and responsibilities to deliver ECMO care (poststimulation)	92	8	0	Creating a space to improve communications, decision-making, and express concerns via simulation
I trust my colleagues to perform the processes of care needed to deliver ECMO care (poststimulation)	92	0	0	Working together as a stronger and more confident team because of simulation

[a]ECMO: extracorporeal membrane oxygenation.

Table . Triangulation of quantitative and qualitative results triangulation between frequently endorsed survey items and themes emerging from post-simulation focus groups (part 2).

Question		% Never or rarely	% Inconsistently	% Consistently	Qualitative theme
A leader is clearly recognized by all team members					Working together as a stronger and more confident team because of simulation
	Prestimulation	0	41	59	
	Poststimulation	0	17	83	
Each team member demonstrates a clear understanding of his or her role					Working together as a stronger and more confident team because of simulation
	Prestimulation	0	37	63	
	Poststimulation	0	8	92	
The team prompts each other to attend to all significant clinical indicators throughout the procedure or intervention					Working together as a stronger and more confident team because of simulation
	Prestimulation	0	26	74	
	Poststimulation	0	8	92	
Disagreements or conflicts among team members are addressed without a loss of situation awareness					Working together as a stronger and more confident team because of simulation
	Prestimulation	0	30	70	
	Poststimulation	0	4	96	
	Poststimulation	0	4	96	
	Poststimulation	0	0	100	

Discussion

Principal Findings

This study was designed to evaluate the impact of ECMO therapy simulation training, specifically focused on enhancing teamwork and communication. The study was successful in validating the survey for future use in assessing the effectiveness of ECMO simulation training in improving teamwork and communication. However, while rigorous and well thought out in design, clear flaws were identified that need to be addressed in future attempts to study this type of simulation exercise. We outline the limitations of the study with recommendations for research with the intention to share what to avoid when

XSL•FO
RenderX

designing and implementing studies that assess a clinical education approach in a complex clinical setting. We provide a unique validated tool to assess teamwork and collaboration across clinical disciplines during ECMO therapy, where existing evidence assesses the impact of simulation approaches on knowledge.

Strengths and Limitations

First, the original focus of the study targeted physician fellows and residents from various clinical teams that practice in the CVI unit. An assumption was made in the study design that peer evaluators would have enough interaction with trainees before and after the simulation training to evaluate changes in their behavior. Due to the nature of the rotation of this participant population, the peers were unable to assess any impact. Additionally, the rotation of the trainees contributed to difficulty in follow-up for study participation in both the quantitative and qualitative aspects of the study. Only 2 responses were received for the 3-month postsimulation training survey, and the qualitative study was altered midstudy to garner more participation in study focus groups. The team was unable to obtain commitment from trainees for the 2-month and 5-month planned focus groups and amended the study for a trainee self-evaluation focus group immediately following the simulation training and 1 month post. The study team was unable to coordinate the 1-month post–focus group due to a lack of availability of the fellow and resident trainees. The lack of participation led to the inability to assess levels 3 and 4 of the Kirkpatrick Training Evaluation Framework [11]. Additionally, the lack of participation reduced the validity of the qualitative data obtained in the focus groups. To generalize the qualitative results of the study, the original target of 6 simulation cohorts with a total of 4 peer participants each would be necessary. We suggest future studies alter the study design to broaden study participants to the entire interprofessional team to ensure the target participant enrollment and focus groups are reached. Second, the trainee rotation also did not guarantee exposure of the trainees to ECMO cannulation postsimulation training to practice the technical skills gained from the simulation training. Third, quantitative results demonstrated there is merit to this training simulation approach. Where there were positive shifts from pre- compared to postsimulation training survey results. However, we were unable to calculate statistical significance in pre- and postresponses due to survey collection methods. Survey participation was anonymous and a routine part of the simulation training program. We were unable to align individual pre- and postsurvey responses to apply this statistical strategy or follow up with specific trainees that missed questions. Fourth, although the trainees had a qualitatively and quantitatively favorable response in ECMO initiation following the simulation exercise per survey results, the study did not conclusively demonstrate their ability to actively use that attained knowledge beyond the original simulation date given the lack of actual cannulations and, again, being observed by staff who could claim that teamwork was significantly improved in future interactions. Lastly, the study team anticipated a larger sample of participants, but the recruitment challenges, focus on physician fellows and residents, and staff shortages due to the impact of the COVID-19 pandemic were severe limitations of the study.

Despite these limitations, the quantitative survey results descriptively highlighted the positive impact on the trainees. Level 1 questions of the Kirkpatrick Training evaluation [11] were met with a strong level of agreement, with no level of disagreement responses (Figure 1). Additionally, each of the level 2 questions shifted to a higher level of agreement post the simulation training. The same was true for the responses to the Mayo Teamwork Scale, where each response shifted to more consistent teamwork behavior pre- and postsimulation training (Table S4 in Multimedia Appendix 8).

We know team-based interprofessional care has historically demonstrated gains in positive patient outcomes in the ICU and is seen as the solution to reduce medical errors and poor quality [12-15]. Moreover, a key component of ECMO care is interprofessional collaboration, as it requires a large and multifaceted team of providers collaborating to carry out complementary tasks to one another [16]. Simulation-based team training can cultivate and preserve interprofessional teamwork and communication [16]. However, collaboration across the care team is not a standard topic covered in clinical curriculum [13,15]. We believe our survey results support the merit of our teamwork-focused simulation training approach and its ability to foster a higher level of collaboration when the clinical team is faced with deciding to initiate ECMO therapy in the cardiac and vascular patient population. These findings highlight the importance of simulation training from other innovative ways of ECMO skills training, such as game-based mobile apps, which might not cultivate a teamwork approach to the same extent [17]. This approach could be applied to supplement the lack of practical teamwork focus in today's clinical curriculum.

Despite the identified limitations, the study underscored several positive aspects of ECMO simulation training. The quantitative survey results notably revealed a significant positive impact on the trainees. Level 1 questions of the Kirkpatrick Training evaluation demonstrated a strong level of agreement without any disagreement responses, indicating a high degree of satisfaction with the training (Figure 1). Furthermore, each of the level 2 questions exhibited a shift towards higher levels of agreement postsimulation training. Similarly, responses to the Mayo Teamwork Scale demonstrated a consistent improvement in teamwork behavior before and after simulation training (Table S4 in Multimedia Appendix 8). This reaffirms the notion that team-based interprofessional care, a cornerstone in ICU settings, can lead to enhanced patient outcomes and reduced medical errors. The study's focus on cultivating interprofessional collaboration through simulation-based training aligns with the demands of ECMO care, which relies heavily on coordinated efforts among various health care professionals. These findings highlight the effectiveness of the teamwork-focused simulation training approach in preparing clinical teams to make critical decisions regarding ECMO therapy in the cardiac and vascular patient population. Moreover, they emphasize the importance of incorporating such training into clinical curricula to ensure a holistic approach to health care education. The study's insights pave the way for future research endeavors to further explore

XSL•FO
RenderX

and refine the application of simulation training in improving teamwork and patient outcomes in complex clinical settings. By addressing the outlined recommendations and leveraging innovative approaches, such as virtual reality simulation, the medical community can continue to advance ECMO care delivery and interprofessional collaboration, ultimately enhancing patient care outcomes.

Future research may build upon the learning of this study to strengthen the understanding of a teamwork-focused simulation approach. We would encourage implementation of the following and plan for our future studies to include (1) continuing the study with the entire interprofessional team, using the survey to build on exploratory factor analysis that validated the survey questions and provide a confirmatory factor analysis to validate results; (2) emphasize established continuity with the learners and the peer evaluators in the study design to mitigate the limited interactions with the study participants outside of the actual simulation and during their clinical rotation; (3) training operational staff participating in gathering data on best practices of data collection for operations and research; (4) the team would encourage incorporating and evaluating the impact of the results on patient outcomes. Answering if patient outcomes improved with increased teamwork and collaboration of the interprofessional team. This would require a larger sample size of trainees involved in simulation training.

Conclusions

We were challenged with the reality of executing a research protocol in a highly complex health care environment, for example, clinician availability, time, response, ability for follow-up, change in protocol, data collection from clinical staff, etc. While these difficulties altered our study approach, the study team believes the design attempted in this study had merit in understanding the impact of a teamwork-focused ECMO simulation approach. We would encourage the medical community to build on the strengths of the design, fortify the weaknesses, and continue to emphasize the need for simulation training to improve ECMO care delivery and teamwork in the clinical setting. Especially as the field of simulation training continues to expand into new mediums like virtual reality [18].

Acknowledgments

Keck Medicine of USC (University of Southern California) Cardiac and Vascular Institute. Li Ding is supported by a grant (UL1TR001855) from the National Center for Advancing Translational Science (NCATS) of the US National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Training curriculum.
[DOCX File, 36 KB - mededu_v11i1e57424_app1.docx]

Multimedia Appendix 2
Extracorporeal membrane oxygenation circuit.
[DOCX File, 95 KB - mededu_v11i1e57424_app2.docx]

Multimedia Appendix 3
Study timeline.
[PDF File, 58 KB - mededu_v11i1e57424_app3.pdf]

Multimedia Appendix 4
Qualitative interview guide.
[DOCX File, 26 KB - mededu_v11i1e57424_app4.docx]

Multimedia Appendix 5
Quantitative survey tool.
[DOCX File, 39 KB - mededu_v11i1e57424_app5.docx]

Multimedia Appendix 6
Amended study timeline.
[PDF File, 58 KB - mededu_v11i1e57424_app6.pdf]

Multimedia Appendix 7
Qualitative focus group themes.

XSL•FO
RenderX

[DOCX File, 16 KB - mededu_v11i1e57424_app7.docx]

Multimedia Appendix 8
Quantitative survey results.
[DOCX File, 24 KB - mededu_v11i1e57424_app8.docx]

References

1. Betit P. Technical advances in the field of ECMO. Respir Care 2018 Sep;63(9):1162-1173. [doi: 10.4187/respcare.06320] [Medline: 30166411]
2. Banfi C, Bendjelid K, Giraud R. High-fidelity simulation for extracorporeal membrane oxygenation training, utile or futile? J Thorac Dis 2017 Nov;9(11):4283-4285. [doi: 10.21037/jtd.2017.10.54] [Medline: 29268492]
3. Fouilloux V, Gran C, Guervilly C, Breaud J, El Louali F, Rostini P. Impact of education and training course for ECMO patients based on high-fidelity simulation: a pilot study dedicated to ICU nurses. Perfusion 2019 Jan;34(1):29-34. [doi: 10.1177/0267659118789824] [Medline: 30014779]
4. Burton KS, Pendergrass TL, Byczkowski TL, et al. Impact of simulation-based extracorporeal membrane oxygenation training in the simulation laboratory and clinical environment. Simul Healthc 2011;6(5):284-291. [doi: 10.1097/SIH.0b013e31821dfcea]
5. ELSO guidelines for training and continuing education of ECMO specialists. 2010. URL: http://www.elso.org/Portals/0/IGD/Archive/FileManager/97000963d6cusersshyerdocumentselsoguidelinesfortrainingandcontinuingeducationofecmospecialists.pdf [accessed 2021-08-21]
6. Zakhary BM, Kam LM, Kaufman BS, Felner KJ. The utility of high-fidelity simulation for training critical care fellows in the management of extracorporeal membrane oxygenation emergencies: a randomized controlled trial. Crit Care Med 2017 Aug;45(8):1367-1373. [doi: 10.1097/CCM.0000000000002437] [Medline: 28422779]
7. Glass KM. Research in ECMO simulation: a review of the literature. In: Johnston LC, Su L, editors. Comprehensive Healthcare Simulation: ECMO Simulation: Springer International Publishing; 2021:147-152. [doi: 10.1007/978-3-030-53844-6_17]
8. Fehr JJ, Shepard M, McBride ME, et al. Simulation-based assessment of ECMO clinical specialists. Simul Healthcare 2016;11(3):194-199. [doi: 10.1097/SIH.0000000000000153]
9. Loeb D, Shoemaker J, Parsons A, Schumacher D, Zackoff M. How augmenting reality changes the reality of simulation: ethnographic analysis. JMIR Med Educ 2023 Jun 30;9:e45538. [doi: 10.2196/45538] [Medline: 37389920]
10. Malec JF, Torsher LC, Dunn WF, et al. The Mayo High Performance Teamwork Scale: reliability and validity for evaluating key crew resource management skills. Simul Healthcare 2007 ;2(1):4-10. [doi: 10.1097/SIH.0b013e31802b68ee]
11. Kirkpatrick JD, Kirkpatrick WK. Kirkpatrick's Four Levels of Training Evaluation: ATD Press; 2016.
12. Donovan AL, Aldrich JM, Gross AK, et al. Interprofessional care and teamwork in the ICU. Crit Care Med 2018 Jun;46(6):980-990. [doi: 10.1097/CCM.0000000000003067] [Medline: 29521716]
13. Ervin JN, Kahn JM, Cohen TR, Weingart LR. Teamwork in the intensive care unit. Am Psychol 2018;73(4):468-477. [doi: 10.1037/amp0000247] [Medline: 29792461]
14. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century: National Academies Press; 2001. [Medline: http://www.ncbi.nlm.nih.gov/books/NBK222274]
15. Manthous CA, Hollingshead AB. Team science and critical care. Am J Respir Crit Care Med 2011 Jul 1;184(1):17-25. [doi: 10.1164/rccm.201101-0185CI] [Medline: 21471081]
16. Elshenawy S, Franciscovich CD, Williams SB, French HM. Interprofessional education and ECMO simulation. In: Johnston LC, Su L, editors. Comprehensive Healthcare Simulation: ECMO Simulation: Springer International Publishing; 2021:89-98. [doi: 10.1007/978-3-030-53844-6_10]
17. Wang Z, Gu R, Wang J, et al. Effectiveness of a game-based mobile app for educating intensive critical care specialist nurses in extracorporeal membrane oxygenation pipeline preflushing: quasi-experimental trial. JMIR Serious Games 2023 Dec 7;11:e43181. [doi: 10.2196/43181] [Medline: 38062643]
18. Gupta S, Wilcocks K, Matava C, Wiegelmann J, Kaustov L, Alam F. Creating a successful virtual reality-based medical simulation environment: tutorial. JMIR Med Educ 2023 Feb 14;9:e41090. [doi: 10.2196/41090] [Medline: 36787169]

Abbreviations

CVI: Cardiac and Vascular Institute
ECMO: extracorporeal membrane oxygenation
HIPAA: Health Insurance Portability and Accountability Act
ICU: intensive care unit
USC: University of California

XSL•FO
RenderX

Edited by B Lesselroth; submitted 23.02.24; peer-reviewed by A Hutchinson, FM Carr, M Capoccia; revised version received 16.05.24; accepted 03.12.24; published 24.01.25.

Please cite as:
Brown J, De-Oliveira S, Mitchell C, Cesar RC, Ding L, Fix M, Stemen D, Yacharn K, Wong SF, Dhillon A
Barriers to and Facilitators of Implementing Team-Based Extracorporeal Membrane Oxygenation Simulation Study: Exploratory Analysis
JMIR Med Educ 2025;11:e57424
URL: https://mededu.jmir.org/2025/1/e57424
doi:10.2196/57424

© Joan Brown, Sophia De-Oliveira, Christopher Mitchell, Rachel Carmen Cesar, Li Ding, Melissa Fix, Daniel Stemen, Krisda Yacharn, Se Fum Wong, Anahat Dhillon. Originally published in JMIR Medical Education (https://mededu.jmir.org), 24.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluation of the Inverted Classroom Approach in a Case-Study Course on Antithrombotic Drug Use in a PharmD Curriculum: French Monocentric Randomized Study

Georges Jourdi[1,2], Prof Dr; Mayssa Selmi[2], PharmD; Pascale Gaussem[3,4], Prof Dr; Jennifer Truchot[5,6,7], Prof Dr Med; Isabelle Margaill[1], Prof Dr; Virginie Siguret[1,2], Prof Dr

[1]Optimisation thérapeutique en neuropharmacologie, INSERM U1144, Université Paris Cité, Paris, France
[2]Service d'Hématologie Biologique, Hôpital Lariboisière, AP-HP.Nord, Paris, France
[3]INSERM UMR-S970, Paris Cardiovascular Research Center, Université Paris Cité, Paris, France
[4]Service d'Hématologie Biologique, Hôpital Européen Georges Pompidou, AP-HP.Centre, Paris, France
[5]Emergency Department, Hôpital Cochin, AP-HP.Centre, Université Paris Cité, Paris, France
[6]iLumens, Université Paris Cité, Paris, France
[7]US UPPERS, Faculte´ de Sante´, Unite´ de Service Pour la Pe´dagogie, l'Enseignement et la Recherche, Université Paris Cité, Paris, France

Corresponding Author:
Georges Jourdi, Prof Dr
Optimisation thérapeutique en neuropharmacologie, INSERM U1144, Université Paris Cité, Paris, France

Abstract

Background: Appropriate antithrombotic drug use is crucial knowledge for pharmacy students.

Objective: We sought to compare the inverted classroom (IC) approach to a traditional question-and-answer educational approach with the aim of enhancing pharmacy students' engagement with a case-study course on antithrombotic drug use.

Methods: Third-year PharmD (Doctor of Pharmacy) students from Paris Cité University were randomly assigned to control (n=171) and IC (n=175) groups. The latter were instructed to read and prepare the preprovided course material 1 week before the in-class session to assume the instructor role on the target day, whereas students of the control group attended a traditional case-study course carried out by the same instructor. All students completed pre- and posttest multiple-choice questions surveys assessing their knowledge levels as well as stress, empathy, and satisfaction questionnaires.

Results: A significantly higher participation rate was observed in the control group (93/171, 54%) compared to the IC group (65/175, 37%; *P*=.002). Women (110/213, 52%) participated more than men (48/133, 36%; *P*=.002) whatever the group was. Students' knowledge scores from both groups had similar results with no difference neither in the prescore (1.17, SD 0.66 and 1.24, SD 0.72 of 5, respectively) nor in the short-term knowledge retention (2.45, SD 0.61 and 2.35, SD 0.73, respectively). The IC approach did not increase student stress or enhance their empathy for the instructor. It increased the preclass workload (*P*=.02) and was not well received among students.

Conclusions: This study showed that the traditional educational approach remains an efficient method for case-study courses in the early stages (ie, third-year) of the 6-year PharmD curriculum, yet dynamic methods improving the active role of students in the learning process are still needed.

(JMIR Med Educ 2025;11:e67419) doi:10.2196/67419

KEYWORDS

antithrombotic drugs; case-study course; inverted classroom; pharmacy students; traditional educational approach; medical education

Introduction

The French Regional Centers of Pharmacovigilance recently revealed that 8.5% of hospital admissions of 141 short-stay specialist medical wards randomly selected in 69 public hospitals, were related to adverse drug reactions [1]. Antithrombotic drugs were involved in 11.6% of the cases, placing them in second place right behind antineoplastic drugs. Some of the adverse drug reactions (mainly bleeding) were considered to be preventable because the drugs had not been used per the summary of product characteristics or guidelines [2]. In the French health care system, antithrombotics are mainly available in pharmacies, and for some also on websites but under the responsibility of a pharmacist. Pharmacists, particularly community-based pharmacists, are easily accessible health care professionals [3]. They are experts in drug therapy use, assessing

XSL•FO
RenderX

each patient through observation, dialogue, and consideration of clinical indicators. They are involved in monitoring the patient's compliance with treatment as well as their response to drug therapy through regular follow-up. This allows for the early detection of adverse effects or drug misuse. Therefore, pharmacist intervention should have a positive impact on the management of patients on antithrombotic therapy. All the above reasons makes the optimal use of antithrombotic drugs of utmost importance to learn for pharmacy students to prevent iatrogenesis.

In France, the PharmD (Doctor of Pharmacy) curriculum consists of a 6-year course. Three teaching models are used mainly: the lecture-based classroom, the case-study class, and the practical session. Case-study class is a hands-on approach to learning that involves presenting realistic scenarios and helps students to apply theoretical knowledge in clinic-like settings and attain a high-order cognitive level per Bloom's taxonomy [4]. The instructor asks students to participate in the case study analysis and discussion. This method favors the development of a deep understanding of the subject and avoids passive note-taking in students. However, this objective is not necessarily always reached. Since its introduction in 2000, the inverted classroom (IC) approach, switching away from the traditional educational approach for the lecture-based classroom, literally inverts the focus per Bloom's taxonomy: the bottom parts of the taxonomy (ie, understand and memorize basic concepts) are reserved for student self-instruction through readings, short recorded video, audio lectures, etc, while the class time is focused on the upper parts of the taxonomy (ie, analyze, justify a stand, and create original work). The IC approach has been increasingly studied in health professions students' education including pharmacy school [5-17]. These studies reported a positive impact of this approach on students' knowledge and skills in most of the cases in comparison to lecture-based courses [8-11,13,14,18-24]. It has been introduced into various courses including pharmacotherapy, pharmacokinetics, pharmaceutical calculations, pharmacy practice, and others [7,12,25-32]. That said, the added value of the IC approach has rarely, if at all, been tested for case-study courses in pharmacy education. Hence, we sought to conduct a monocentric study investigating the added value of an IC approach in a case-study course during the PharmD curriculum at Paris Cité University. In this IC approach, students received the course material before the in-class session and were asked to prepare it and assume the role of the instructor on the target day. We aimed to assess knowledge acquisition, preclass workload and students' self-assessment of their stress, empathy, and global satisfaction. It was hypothesized that the IC approach would lead to improved outcomes compared to the traditional question-and-answer educational approach.

Methods

Ethical Considerations

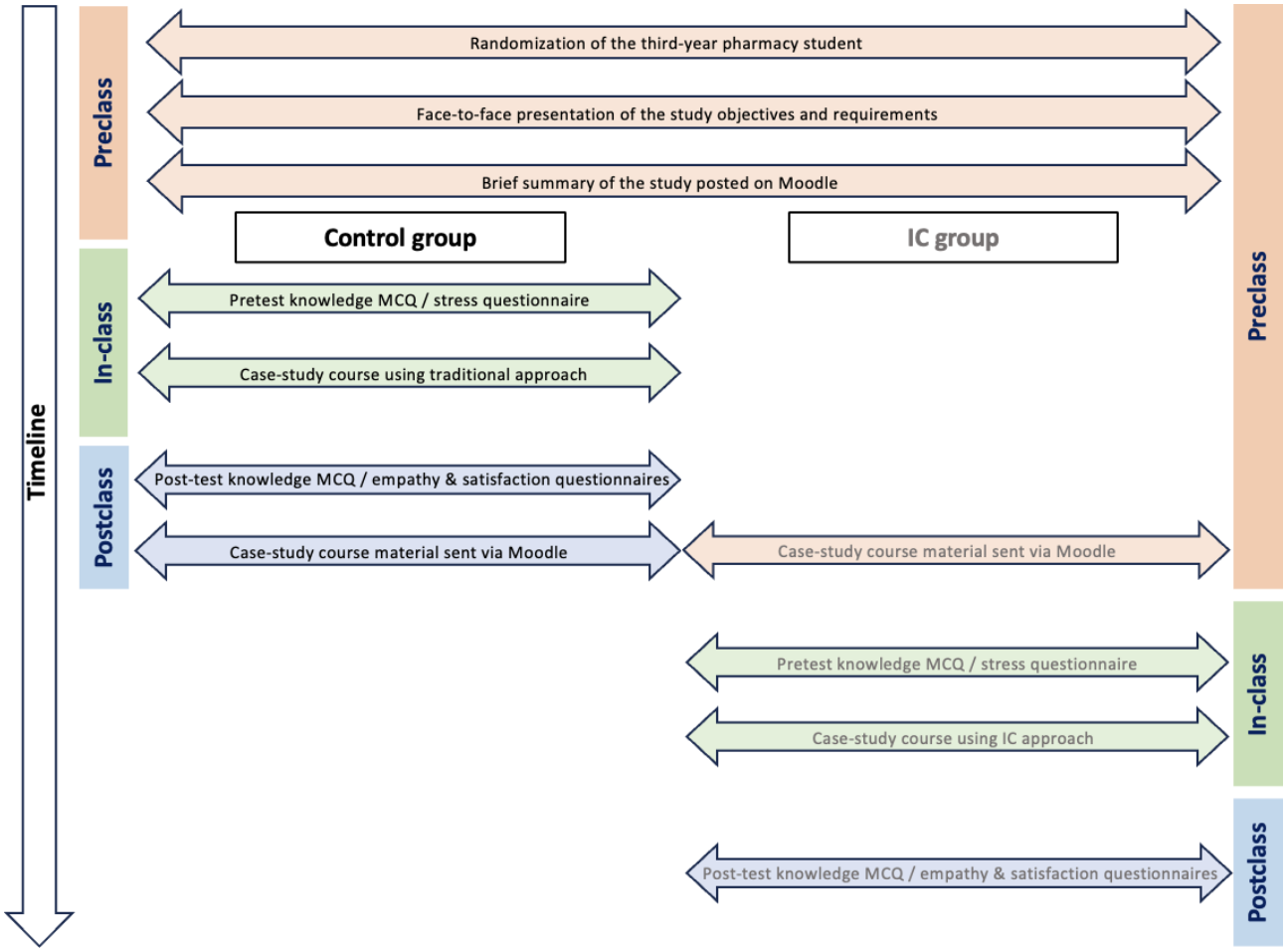
This study was approved by the Institutional Review Board and Ethics Committee of Paris Cité University (00012023‐20) and all procedures were performed per the Helsinki Declaration. Students were completely free to participate or not in this study. They were informed that neither participation nor nonparticipation in this study would influence their passing of this course or their grades. The participants were also informed about the option to withdraw from this study at any time. Informed consent was obtained from all the students who participated in this study. They were not paid for their participation. Collected data were anonymously analyzed.

Participants

We conducted this study at the faculty of Pharmacy of Paris Cité University in March 2023 (2022/2023 academic year). The participants were in the spring semester of the third year of the PharmD curriculum. At the beginning of the semester, students had basic courses on hemostasis (physiology and pathology) and thrombosis. During the month preceding the case-study course, they also had 3 lecture-based courses for a total of 4 hours on antithrombotic drugs. The recordings of the lectures were then available on a teaching platform ("Moodle"). The 90-minute case-study course on antithrombotic drug use entailed 16 groups consisting of about 20 to 22 students each, which were randomly assigned to the groups by the office of student affairs of the faculty of Pharmacy without any influence from the instructor. Therefore, a randomized assignment can be assumed. Students were highly required to respect their group assignment. Each time, 2 groups assisted simultaneously in one course, thus the course was repeated 8 times by the same instructor. Participants were assigned to 2 major groups: the control group and the IC group (Figure 1). Attendance was not mandatory.

XSL•FO
RenderX

Figure 1. Overview of this study's design. IC: inverted classroom; MCQ: multiple-choice question.



Study Design and Procedure

Three weeks before the case-study course, the instructor informed all the students about the concept of the IC approach and this study's process in addition to the importance of their engagement in the learning process. A summary of the process was also posted on Moodle. Students were informed that half of them would be assigned to the IC group where they would receive the course material for preparation 1 week before the in-class session to take the role of the instructor on the target day. The course material was delivered as a PDF document of a PowerPoint (Microsoft Corp) presentation (encompassing 63 slides) file via Moodle. The same course material was used in both approaches. The courses took place on Monday and Tuesday. One week separated the courses of the control from the IC groups.

Six cases concerning various clinically relevant scenarios were included in this course in addition to some incorporated slides to review and emphasize selected foundational concepts. They were centered on the most relevant aspects of antithrombotic drug use in different clinical settings: (1) treatment of pulmonary embolism associated with proximal deep vein thrombosis in a posttrauma patient aged 63 years; (2) prevention of thrombotic events postelective knee replacement surgery in patient aged 78 years; (3) treatment of pulmonary embolism that occurred under combined pill contraception in a woman aged 20 years; (4) prevention of thrombotic events in an acutely ill medical patient aged 80 years; (5) prevention of stroke and systemic embolism in a patient aged 82 years with atrial fibrillation and renal insufficiency; and (6) antithrombotic treatment of myocardial infarction in a patient aged 54 years. Each case was accompanied by a set of standardized questions about antithrombotic treatment decisions and adequate monitoring. Students in the control group attended a case-study course carried out by the instructor with a traditional question-and-answer approach, whereas students in the IC group took on the role of the instructor and were able to recall basic concepts or add details to the course material freely. Two to three students were randomly asked to present and discuss 1 of the 6 cases on the target day. By doing so, students were able to draw on each others' knowledge and understanding. The instructor added details, provided guidance, clarity and feedback whenever required during the progress of the students' presentation and summarized the main features at the end of each case. The classes in the control and IC groups were conducted by the same experienced instructor who is familiar with the content and organization of the case-study course to guarantee the consistency of the teaching content and objectives in the 2 educational approaches. This study's design and progress are illustrated in Figure 1.

Data Collection

On the target day, students in both control and IC groups were asked to complete a pretest (ie, at the beginning) and a posttest (at the end) survey (Multimedia Appendix 1): it consisted of

XSL•FO
RenderX

the same 5 multiple-choice questions (MCQs) to be completed within 5 minutes, then collected in an identified way (ie, including the student name, and the date and the hour of the questionnaire completion) to pair pre- and posttest scripts. Questionnaires were anonymized by a secretary and corrected afterward by an independent assistant instructor. Another 5-minute survey assessing the stress in the week preceding the in-class course (thus assessing how they were affected from the moment they knew which group they belonged to till the target day) was also completed by all the students at the beginning of the course. At the end of the course, students were asked to complete 2 additional surveys, 1 assessing their empathy for the instructor and the other their global satisfaction. The first consisted of rating 3 items by a 7-point Likert scale (1=strongly disagree, 2=disagree, 3=somewhat disagree, 4=neutral, 5=somewhat agree, 6=agree, and 7=strongly agree). The second included 3 questions regarding the preclass workload associated with the educational approach and the related students' perception, and 6 others linked to the students' satisfaction with the course objectives, course material, in-class progress, and educational approach. Surveys were built based on previously validated assessment tools [33-35]. Stress, empathy, and satisfaction surveys (Multimedia Appendices 2-4) were filled out anonymously. Students were not allowed to keep a copy of the different surveys nor to take a photo of these documents.

Statistical Analysis

The distribution of the data was evaluated using the Shapiro-Wilk test. The percentage of participation was compared between groups and sexes using a 2-way ANOVA followed by the 2-stage setup method of Benjamini, Krieger, and Yekutieli [36] for multiple comparisons. The results of the pre- and posttest MCQs were compared for statistically significant differences using the nonparametric Wilcoxon matched-pairs signed rank test. The preclass workload was compared between the 2 groups using the Mann Whitney test. The data relative to the empathy and stress self-assessment were compared between both groups using chi-square test whereas those relative to satisfaction were analyzed using the Fisher exact test. Error probability with a *P* value less than .05 was considered significant. Statistical analysis and graphical representation were performed using GraphPad Prism (version 10.0.2, GraphPad Software, Inc).

Results

Participants Characteristics

In this study, 346 third-year adult students (women: n=213, 62%; men: n=133, 38%) were randomized. As attendance is not mandatory, only 46% (n=158) attended the in-class session. All of them took part in this study. Ninety-three (women: n=70, 75%; men: n=23, 25%) were in the control group whereas 65 (women: n=40, 62% women; men: n=25, 38%) were in the IC group (Table 1). A significantly higher participation rate was observed in the control group (93/171, 54%) compared to the IC group (65/175, 37%; *P*=.002). Women (110/213, 52%) participated more than men (48/133, 36%; *P*=.002) in the case-study course no matter the allocation group. No other demographic information regarding the students was collected.

Table . Characteristics of the participants. Absolute numbers with the percentages concerning the corresponding randomized participants are reported. Women participated more than men in the case-study course whatever the group was (*P*=.002) and a significantly higher participation was observed in the control group compared to the IC[a] group (*P*=.002).

	Sex	Control group	IC group	Total
Randomized participants				
	Men	55	78	133
	Women	116	97	213
	Total	171	175	346
Effective participants, n (%)				
	Men	23 (42)	25 (32)	48 (36)
	Women	70 (60)	40 (41)	110 (52)
	Total	93 (54)	65 (37)	158 (46)

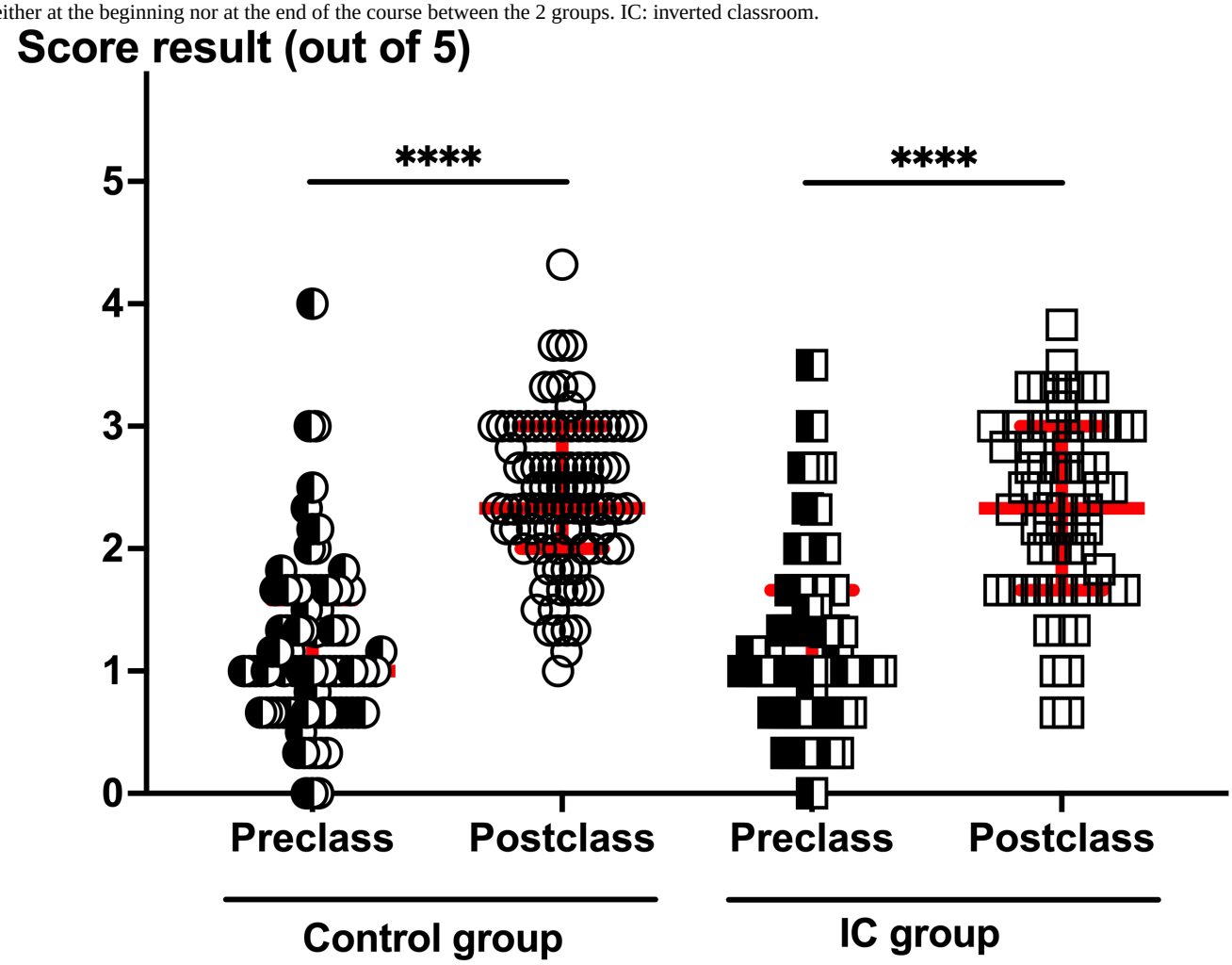
[a]IC: inverted classroom.

Pre- and Postclass Knowledge Survey

On the target day, the in-class session started for both groups with a 5 MCQ survey for 5 minutes to assess the students' readiness to discuss cases and stimulate the recall of knowledge learned before the case-study course. Questions were on the "take-home messages" related to antithrombotic drug use in real-life clinical settings. These MCQs also help students to identify their possible misconceptions at the beginning of the course. The mean and SD of the prescore (out of 5) did not differ between the control group (1.17, SD 0.66) and the IC group (1.24, SD 0.72; Figure 2). To evaluate the students' short-term knowledge retention, the same survey was completed immediately after the class. The mean score improved from the prescore (*P*<.001), in both the control group 2.45 (SD 0.61) and the IC group 2.35 (SD 0.73; Figure 2). Knowledge improvement was comparable in students from both groups.

XSL•FO
RenderX

Figure 2. Pre- and postclass knowledge assessment survey per group. Five multiple-choice questions were completed by the control (circles, n=93) and IC (squares, n=65) groups before (semiclosed symbols) and after (open symbols) the completion of the case-study course. Red bars reflect median values with IQRs. Students' scores significantly increased (*P*<.001) at the end of the course in both groups. No difference in the scores was observed neither at the beginning nor at the end of the course between the 2 groups. IC: inverted classroom.

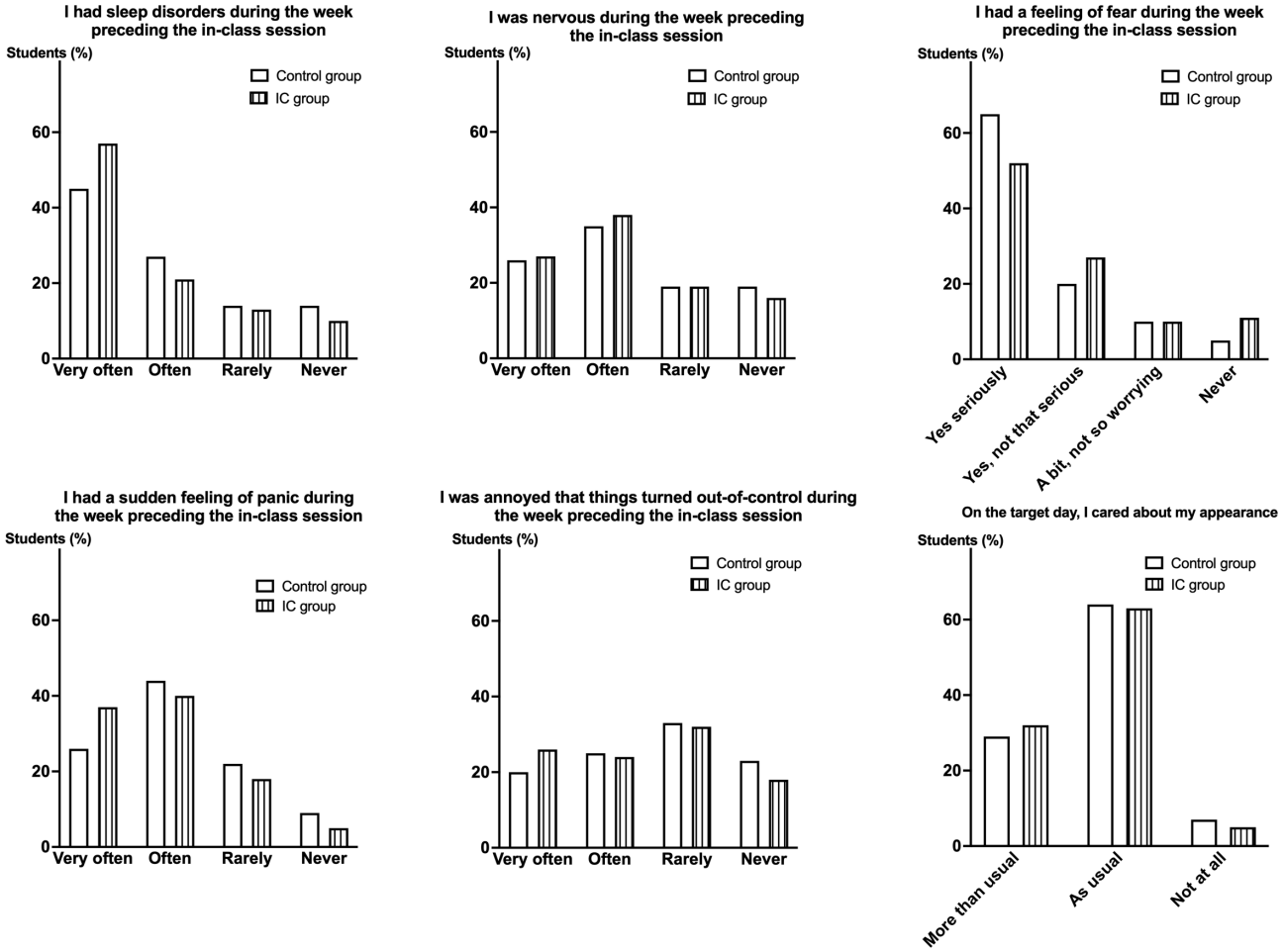


Stress Self-Assessment

Apart from the preclass knowledge assessment survey, students in both groups completed a stress survey at the beginning of the class during 5 minutes. Questions related to sleep disorders, nervousness, fear, panic, and annoyance during the week preceding the in-class course were completed (Figure 3). Although approximately 50% of the students reported sleep disorders and a feeling of fear during the week preceding the in-class session, no significant difference was observed between the 2 groups for any of the above-mentioned parameters. Likewise, students care for their appearance on the target day did not differ between the control and the IC groups.

XSL•FO
RenderX

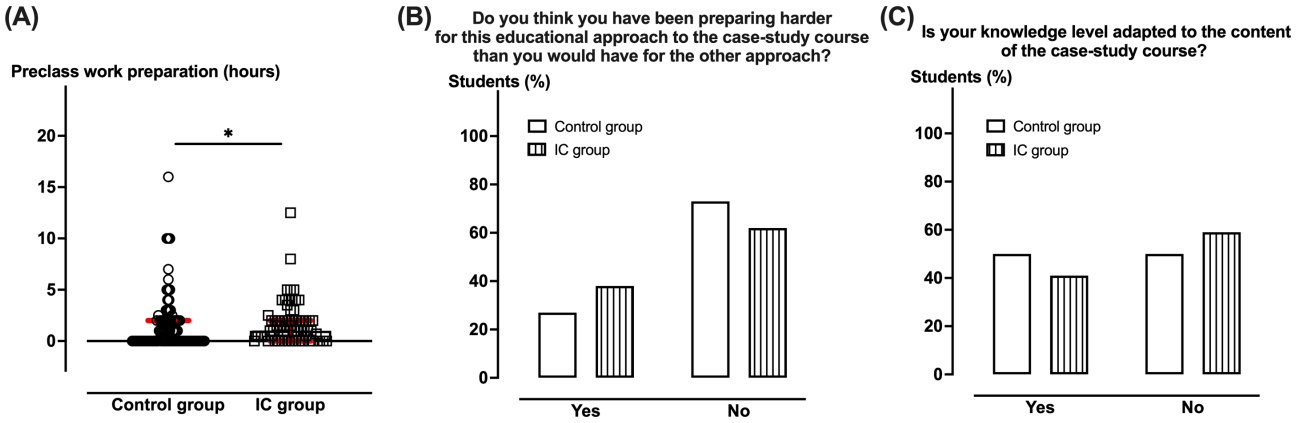
Figure 3. Stress self-assessment questionnaire. A 6-question stress self-assessment survey was completed by both the control (n=93) and IC (n=65) groups at the beginning of the case-study course. No significant difference for any of the 6 questions was observed between the 2 groups. IC: inverted classroom.



Preclass Workload

Students of both groups were asked to provide the number of hours spent preparing for the course, their perception of the preclass preparation difficulty in comparison to the other educational approach as well as the self-assessment of the required skill level. Preclass workload was estimated at 1 (IQR 0-2) hour in the IC group as a median, which was significantly higher (P=.02) than in the control group, 0 (IQR 0-2) hour (Figure 4A). While 38% of the students in the IC group considered that they had been preparing harder for the case-study course than it would have been if they were in the control group, 27% of the latter considered that as such it resulted in no significant difference between both groups (Figure 4B). Approximately half of the students in each group considered having a knowledge level adapted to the case-study course content. No difference was observed between both groups (Figure 4C).

Figure 4. Self-assessment of the preclass work per group. (**A**) Preclass preparation requirements concerning working hours, (**B**) student perception, and (**C**) knowledge level self-assessment were compared between the control and IC groups. Red bars reflect median values with IQRs. While preclass work preparation necessitated more time for the IC compared to the control groups (P=.02), no difference was observed between the 2 groups in terms of work difficulty perception and the self-assessment of the knowledge level adapted to the course. IC: inverted classroom.

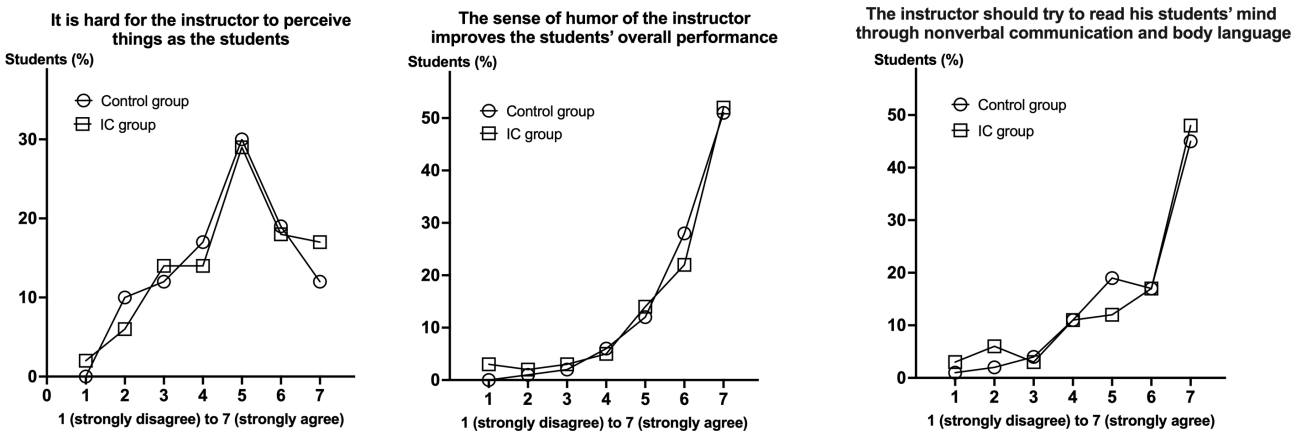


XSL•FO
RenderX

Empathy Self-Assessment

Student empathy for their instructor favors engagement and learning behavior in class, hence 3 related questions rated by a 7-point Likert scale were completed by students of both groups at the end of the course (Figure 5). Although only students in the IC group assumed the role of the instructor within the class, students' opinions were quite similar between both groups. Approximately 30% of the students considered that it is hard for the instructor to perceive things as the students while 50% confirmed the importance of the sense of humor of the instructor in enhancing students' performance and of reading the students' minds through nonverbal communication and body language. These results did not differ between both groups.

Figure 5. Empathy assessment questionnaire. The empathy of students for the instructor was assessed using 3 questions completed at the end of the case-study course. No difference was observed between the control and the IC groups for any of the 3 raised issues. IC: inverted classroom.

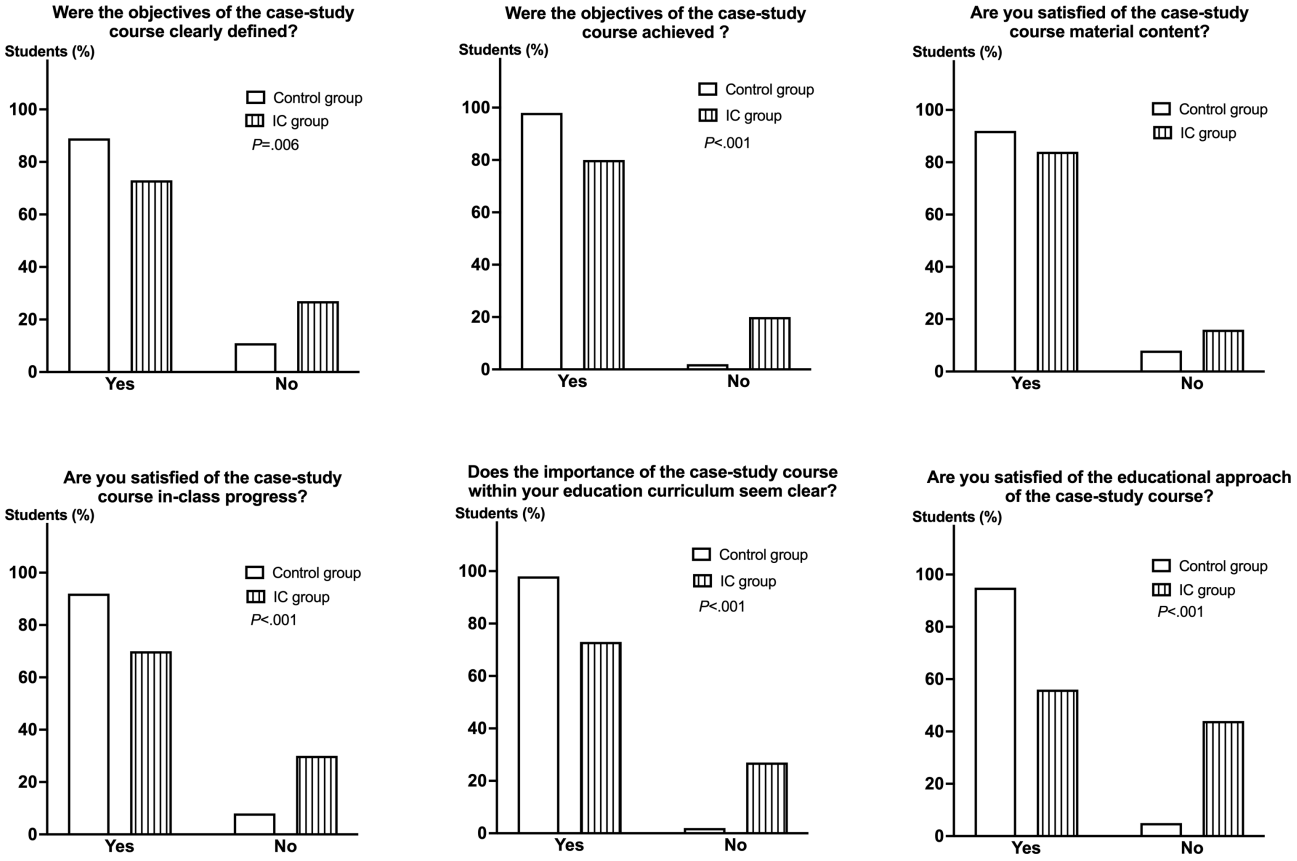


Global Satisfaction Assessment

The students' global satisfaction was evaluated at the end of the course using a 6-item questionnaire (Figure 6). While more than 80% of the students were satisfied with the content of the course material in both groups, the satisfaction survey revealed that the IC approach was not well received among the students. Indeed, 18/65 (28%) of the students found that the objectives of the case-study course not clearly defined and 13/65 (20%) considered that they were not achieved, in comparison to 10/93 (11%; P=.006) and 2/93 (2%; P<.001) in the control group, respectively. A total of 20/65 (30%) of the students in the IC group were not satisfied with the in-class progress of the course versus 7/93 (8%) in the control group (P<.001). While 91/93 (98%) of the students in the control group considered this case-study course on antithrombotic drug use important for pharmacy education, only 47/65 (72%) did so in the IC group (P<.001). When it came to the question of the overall satisfaction of the educational approach, 88/93 (95%) of the students in the control group were satisfied versus (36/65) 55% in the IC group (P<.001).

XSL•FO
RenderX

Figure 6. Satisfaction survey. A 6-item questionnaire relative to the students' satisfaction with the pedagogical strategy (ie, traditional [n=93] vs IC [n=65]) was completed at the end of the case-study course. Overall, students in the control group were significantly more satisfied than those in the IC group. IC: inverted classroom.



Discussion

Principal Findings

Education methodology is increasingly shifting from a teacher-to student-centered learning approach [37]. In this context, we sought to assess whether applying an IC approach for a case-study course has an added value for conveying optimal antithrombotic drug use skills to pharmacy students. Such an educational approach would increase the communication, critical-thinking, problem-solving, and self-learning skills of pharmacy students. Our study revealed that while the IC approach did not increase student stress, it did not enhance their short-term knowledge retention or their empathy for the instructor. It increased the preclass workload and was not well received among the students. Of note, the case-study course on antithrombotic use was the first on this topic in the PharmD curriculum.

For any educational method to be considered successful, there must be evidence that student learning is enhanced. Many small-scale studies showed enhanced student learning following flipping lecture-based pharmacy courses [5-11,24]. Here, the IC approach applied to the case-study course was not associated with an enhanced preclass knowledge level and did not improve the students' short-term knowledge retention. This was also the case in the Everly and Cochran study in which no significant differences in examination question performance between students in the lecture-based section and the flipped format section were observed [9]. That said, MCQ and examination scores are often used as a proxy measure for learning; however, the ultimate goal is for students to be able to apply classroom learning to real-life situations, which is more difficult to assess. Perhaps we need to examine other nonquantitative student characteristics outcomes following the IC approach in case-study courses (such as intellectual curiosity, personal responsibility, reasoning skills, etc) to identify pharmacy students most likely to succeed.

A recent systematic review including 45 studies with a total of 8426 students from various health professional pathways showed that implementing flipped classes may improve academic performance, and may support student satisfaction, yet the certainty of the evidence is low [37]. A second systematic review and meta-analysis including 11 randomized controlled Chinese studies enrolling 1200 participants suggested that flipped classroom pedagogy enhances students' learning enthusiasm, self-learning ability, thinking and communication skills as well as cooperative ability [38]. Although, another systematic review focusing on students in pharmacy education, incorporated 6 observational studies with 1395 participants. No overall significant difference in final academic performance between the 2 educational models was reported [39]. An important heterogeneity of student perspectives from flipped classrooms has emerged in the literature, ranging from positive [10,24,40-42] to negative [21,39,43,44] and mixed [45] perceptions. These differences are potentially due to the different contexts in which these studies were carried out as well as different student populations, backgrounds, sample sizes, and outcome measures. Apart from that, some instructors may be

XSL•FO
RenderX

more effective teachers than others regardless of the teaching modality. Research evaluating which elements contribute to the efficacy of an IC approach in pharmacy education is still needed. Moreover, it is still unclear if there is a particular area or topic that is better suited for the use of the IC approach in the PharmD curriculum. Besides, with the increased use of the IC method, it is important to consider the impact on students when this approach is incorporated into multiple concurrent courses. Determination of the ideal amount of preclass preparation time across the curriculum would provide helpful guidance to pharmacy faculties implementing such teaching methods.

One of the major limitations of the IC approach is its high dependence on the attendance to class time and on student engagement and discipline for reading and preparing the preclass material as previously emphasized [26]. As the class time attendance is not obligatory in our faculty, only 46% of the students were present in the antithrombotic drug use case-study course. The percentage of attendance was more important in the control group than in the IC group. A more elaborate strategy for students' motivation should thus be implemented to obtain a higher engagement and adherence to our case-study courses in general, and to such IC experience if it shall be repeated. Providing clear expectations to students, keeping the preparation tasks focused, and explicitly linking preparation activities to in-class active learning could be some key methods for instructors to increase the proportion of students who prepare for classes. Future research should also be devoted to assessing the potential effect of sex, gender, socioeconomic background, and age on the outcomes of such an approach in the PharmD curriculum.

Preclass preparation could be considered as a considerable "extra" work [10,29,30,46]. Indeed, students in the IC group of our study reported an increased preclass workload which might take up an amount of their spare time leading to negative feedback on this approach. It is to be mentioned that the course materials made available beforehand should not be too complex to be understood by the students on their own. We did our best to include 6 uncomplicated cases issued from real-life settings and incorporate few slides to recall knowledge learned in the three lecture-based courses during the month preceding the case-study course.

Case-study courses with an IC approach are probably important in pharmacy education as they would help students promote higher-level critical-thinking skills, foster analyzing, and improve their communication skills therefore improving their motivation and attitudes [47]. Assisting them in learning how to think and communicate like a graduated pharmacist will prepare them for their future beyond pharmacy school. Communication skills are required to ensure patient understanding and compliance [48,49]. If a patient does not understand the purpose of the antithrombotic treatment, adherence will likely be low. Communication training in pharmacy students is thus mandatory to improve their later effectiveness as future health professionals. Although third-year students were overall not satisfied with this experience, such an approach is worth being retested with "older students," from the fourth to the sixth year of the PharmD curriculum. Successful pharmacy students are expected to have the ability to manage their learning and adequately communicate their knowledge. Self-learning skills are particularly crucial to achieving effective lifelong learning in pharmacy, where scientific knowledge is continually evolving. Consequently, pharmacy students should be trained to be effective self-learners. Antithrombotic drug use assessed using a case-study course is an application-based activity, while the material taught previously is mostly a knowledge-based presentation. Therefore, third-year students may have significant difficulty in providing an application-based activity when their skill level may have been still on a much lower level of Bloom's taxonomy [4] which may explain the low postclass knowledge scores in both groups. They will complete 2 additional learning years where it is anticipated they will gain more knowledge, clinical experience, and communication skills through their traineeships in community-based and hospital pharmacies as well as clinical laboratories. That said, many students were very interested in participating in this pedagogical study and found the experience to be innovative and enjoyable.

From the instructor's perspective, the IC approach might be more challenging than a traditional session with a question-and-answer approach due to the risk that students' presentation and case discussion activities create an unsettled classroom, thus a chaotic environment in which students may feel lost, and the fear that students may be unable to deliver the course adequately. However, the IC approach makes student engagement in the course easier and empowers them as active participants in their learning in comparison to the traditional educational approach. It allows the instructor to guide students in deeper learning processes as previously shown [38,46]. A more interactive teaching strategy may be more attractive than the IC approach, such as the adventure game recently developed by Perrin et al [50]. It is a video game in which the player assumes the role of a protagonist in an interactive story driven by exploration and problem-solving tests. Briefly, the pharmacy student assumes the role of a hematology superhero named SUPER HEMO. SUPER HEMO can meet 5 unwell characters in 5 different steps. The player must answer their questions and find the best way to diagnose and cure them. At the final hematology evaluation, students who played SUPER HEMO had a slightly better (but not statistically significant) median knowledge score than those who did not [50]. The value of such an innovative strategy for case-study courses on antithrombotic drug use in the PharmD curriculum remains to be established. Team-based learning is another educational approach that provides structure, defined timeframes, and formative assessment opportunities. It was previously shown to develop students learning enthusiasm, self-study, and thinking abilities as well as communication skills [51]. Therefore, it might be considered as an alternative approach for improving case-study courses in the PharmD curriculum, and thus is worth being assessed.

One possible limitation of our study is that the long-term effect of our IC approach on knowledge retention and skill application was not assessed. The question relative to this course was deliberately not included in the final exam in order not to create any lack of equity among the students of both groups. We did not complete any postclass knowledge survey 3 to 6 months

XSL•FO
RenderX

later, nor did we assess the acquired skills through, for instance, an objective structured clinical examination. This remains to be specifically investigated. Second, we did not collect data on how many students effectively accessed the course material before the in-class session, although it might be feasible via the information technology service. As the preclass results of the students in the IC group were not better than those of the control group, we hypothesized that a lot of students had not read the course material before the target day. Noteworthily, it is hard to control whether students had effectively read and prepared the course material before the in-class session. Students might only click on the material folder without reading it or read parts of it. Also, several students might have accessed the material via 1 student user ID. We also did not ask students in the presurvey questionnaire whether they had read the assignment. However, they would, most probably, have not told the truth. Third, ideally, students in the IC group should have been asked to prepare the course material by themselves, yet this was not the case to avoid a high level of absenteeism as in-class attendance is not mandatory according to Paris Cité University policy. Despite this, a relatively small number of students effectively participated in this study. Fourth, in-class sessions with the traditional approach were completed 1 week before those with the IC approach to prevent students in the former group from having access to the material course from those of the latter group before the in-class session. Consequently, we cannot rule out the possibility that some students initially assigned to the IC group had changed their group assignment to get the case-study course at an early date. Finally, our findings cannot be generalized to other contexts, particularly to students in other year cohorts or to other specialties. This remains to be specifically investigated.

Conclusions

Our study showed that an IC approach does not appear to be suited to the case-study course on antithrombotic drug use in the third-year PharmD curriculum. While no additional gain in short-term knowledge was observed using this approach in comparison to the traditional educational approach, we perceived significantly lower student satisfaction. However, the increased instructor-student and student-student interactions are still convincing arguments to try this pedagogical approach. Hence, additional research in this field is still needed to implement innovative educational approaches aiming at improving the knowledge and skills of our future pharmacists.

Acknowledgments

The authors thank the members of the Institutional Review Board and Ethics Committee of Paris Cité University for the rapid review of the project, Farah Berrich, MSc, for her administrative support, and the 2022/2023 third-year pharmacy students of Paris Cité University for their participation in this study.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

GJ, VS, JT, and IM conceptualized this study. GJ worked on investigation. GJ and MS did the data analysis. GJ wrote the original draft. VS, MS, PG, JT, and IM reviewed and edited the writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Pre- and postclass multiple-choice questions.
[DOCX File, 16 KB - mededu_v11i1e67419_app1.docx]

Multimedia Appendix 2
Stress self-assessment survey
[DOCX File, 16 KB - mededu_v11i1e67419_app2.docx]

Multimedia Appendix 3
Empathy assessment questionnaire
[DOCX File, 14 KB - mededu_v11i1e67419_app3.docx]

Multimedia Appendix 4
Preclass workload self-assessment and satisfaction questionnaire
[DOCX File, 16 KB - mededu_v11i1e67419_app4.docx]

XSL•FO
RenderX

Checklist 1
CONSORT 2010 checklist. CONSORT: Consolidated Standards of Reporting Trials.
[PDF File, 126 KB - mededu_v11i1e67419_app5.pdf]

References

1. Laroche ML, Gautier S, Polard E, et al. Incidence and preventability of hospital admissions for adverse drug reactions in France: a prospective observational study (IATROSTAT). Br J Clin Pharmacol 2023 Jan;89(1):390-400. [doi: 10.1111/bcp.15510] [Medline: 36002314]
2. Réseau Français des Centres Régionaux de Pharmacovigilance [Website in French]. URL: https://www.rfcrpv.fr/ [accessed 2025-03-25]
3. Kehrer JP, Eberhart G, Wing M, Horon K. Pharmacy's role in a modern health continuum. Can Pharm J (Ott) 2013 Nov;146(6):321-324. [doi: 10.1177/1715163513506370] [Medline: 24228046]
4. Krathwohl DR. A revision of Bloom's taxonomy: an overview. THEORY Pract 2002 Nov 1;41(4):212-218. [doi: 10.1207/s15430421tip4104_2]
5. Roth MT, Mumper RJ, Singleton SF, et al. A renaissance in pharmacy education at the University of North Carolina at Chapel Hill. N C Med J 2014;75(1):48-52. [doi: 10.18043/ncm.75.1.48] [Medline: 24487762]
6. Stewart DW, Brown SD, Clavier CW, Wyatt J. Active-learning processes used in US pharmacy education. Am J Pharm Educ 2011 May 10;75(4):68. [doi: 10.5688/ajpe75468] [Medline: 21769144]
7. Bossaer JB, Panus P, Stewart DW, Hagemeier NE, George J. Student performance in a pharmacotherapy oncology module before and after flipping the classroom. Am J Pharm Educ 2016 Mar 25;80(2):31. [doi: 10.5688/ajpe80231] [Medline: 27073284]
8. Ferreri SP, O'Connor SK. Redesign of a large lecture course into a small-group learning course. Am J Pharm Educ 2013 Feb 12;77(1):13. [doi: 10.5688/ajpe77113] [Medline: 23459199]
9. 115th Annual Meeting of the American Associaton of Colleges of Pharmacy, Grapevine, TX, July 26-30, 2014. Am J Pharm Educ 2014 Jun;78(5):111. [doi: 10.5688/ajpe785111] [Medline: 4064488]
10. Wong TH, Ip EJ, Lopes I, Rajagopalan V. Pharmacy students' performance and perceptions in a flipped teaching pilot on cardiac arrhythmias. Am J Pharm Educ 2014 Dec 15;78(10):185. [doi: 10.5688/ajpe7810185] [Medline: 25657372]
11. Persky AM, Dupuis RE. An eight-year retrospective study in "flipped" pharmacokinetics courses. Am J Pharm Educ 2014 Dec 15;78(10):190. [doi: 10.5688/ajpe7810190] [Medline: 25657377]
12. McLaughlin JE, Gharkholonarehe N, Khanova J, Deyo ZM, Rodgers JE. The impact of blended learning on student performance in a cardiovascular pharmacotherapy course. Am J Pharm Educ 2015 Mar 25;79(2):24. [doi: 10.5688/ajpe79224] [Medline: 25861105]
13. Edginton A, Holbrook J. A blended learning approach to teaching basic pharmacokinetics and the significance of face-to-face interaction. Am J Pharm Educ 2010 Jun 15;74(5):88. [doi: 10.5688/aj740588] [Medline: 20798797]
14. Crouch MA. An advanced cardiovascular pharmacotherapy course blending online and face-to-face instruction. Am J Pharm Educ 2009 May 27;73(3):51. [doi: 10.5688/aj730351] [Medline: 19564994]
15. Wilson JA, Waghel RC, Dinkins MM. Flipped classroom versus a didactic method with active learning in a modified team-based learning self-care pharmacotherapy course. Curr Pharm Teach Learn 2019 Dec;11(12):1287-1295. [doi: 10.1016/j.cptl.2019.09.017] [Medline: 31836155]
16. Wong WJ, Lee SWH, White PJ, Efendie B, Lee RFS. Perspectives on opportunities and challenges in a predominantly flipped classroom-based pharmacy curriculum: a qualitative study. Curr Pharm Teach Learn 2023 Mar;15(3):242-251. [doi: 10.1016/j.cptl.2023.03.004] [Medline: 37055316]
17. Youhasan P, Chen Y, Lyndon M, Henning MA. Exploring the pedagogical design features of the flipped classroom in undergraduate nursing education: a systematic review. BMC Nurs 2021 Mar 22;20(1):50. [doi: 10.1186/s12912-021-00555-w] [Medline: 33752654]
18. Hess R, Hagemeier NE, Blackwelder R, Rose D, Ansari N, Branham T. Teaching communication skills to medical and pharmacy students through a blended learning course. Am J Pharm Educ 2016 May 25;80(4):64. [doi: 10.5688/ajpe80464] [Medline: 27293231]
19. Bergmann J, Sams A. Flip Your Classroom: Reach Every Student in Every Class Every Day: International Society for Technology in Education; 2012.
20. Cai L, Li YL, Hu XY, Li R. Implementation of flipped classroom combined with case-based learning: a promising and effective teaching modality in undergraduate pathology education. Medicine (Baltimore) 2022 Feb 4;101(5):e28782. [doi: 10.1097/MD.0000000000028782] [Medline: 35119043]
21. Chen KS, Monrouxe L, Lu YH, et al. Academic outcomes of flipped classroom learning: a meta-analysis. Med Educ 2018 Jun 25;52(9):910-924. [doi: 10.1111/medu.13616] [Medline: 29943399]
22. Crome M, Adam K, Flohr M, Rahman A, Staufenbiel I. Application of the inverted classroom model in the teaching module "new classification of periodontal and peri-implant diseases and conditions" during the COVID-19 pandemic. GMS J Med Educ 2021;38(5):Doc89. [doi: 10.3205/zma001485] [Medline: 34286069]

XSL•FO
RenderX

23. Malik PRV, Nakhla N. Instructor-blinded study of pharmacy student learning when a flipped online classroom was implemented during the COVID-19 pandemic. Pharmacy (Basel) 2022 May 11;10(3):53. [doi: 10.3390/pharmacy10030053] [Medline: 35645332]
24. McLaughlin JE, Roth MT, Glatt DM, et al. The flipped classroom: a course redesign to foster learning and engagement in a health professions school. Acad Med 2014 Feb;89(2):236-243. [doi: 10.1097/ACM.0000000000000086] [Medline: 24270916]
25. McCabe C, Smith MG, Ferreri SP. Comparison of flipped model to traditional classroom learning in a professional pharmacy course. Educ Sci 2017;7(3):73. [doi: 10.3390/educsci7030073]
26. White PJ, Naidu S, Yuriev E, Short JL, McLaughlin JE, Larson IC. Student engagement with a flipped classroom teaching design affects pharmacology examination performance in a manner dependent on question type. Am J Pharm Educ 2017 Nov;81(9):5931. [doi: 10.5688/ajpe5931] [Medline: 29302082]
27. Saseen JJ, Linnebur SA, Borgelt LM, Trujillo J, Fish DN, Mueller S. A pharmacotherapy capstone course to target student learning and programmatic curricular assessment. Am J Pharm Educ 2017 Apr;81(3):45. [doi: 10.5688/ajpe81345] [Medline: 28496265]
28. Koo CL, Demps EL, Farris C, Bowman JD, Panahi L, Boyle P. Impact of flipped classroom design on student performance and perceptions in a pharmacotherapy course. Am J Pharm Educ 2016 Mar 25;80(2):33. [doi: 10.5688/ajpe80233] [Medline: 27073286]
29. Gloudeman MW, Shah-Manek B, Wong TH, Vo C, Ip EJ. Use of condensed videos in a flipped classroom for pharmaceutical calculations: student perceptions and academic performance. Curr Pharm Teach Learn 2018 Feb;10(2):206-210. [doi: 10.1016/j.cptl.2017.10.001] [Medline: 29706277]
30. Cotta KI, Shah S, Almgren MM, Macías-Moriarity LZ, Mody V. Effectiveness of flipped classroom instructional model in teaching pharmaceutical calculations. Curr Pharm Teach Learn 2016 Sep;8(5):646-653. [doi: 10.1016/j.cptl.2016.06.011]
31. Patanwala AE, Erstad BL, Murphy JE. Student use of flipped classroom videos in a therapeutics course. Curr Pharm Teach Learn 2017;9(1):50-54. [doi: 10.1016/j.cptl.2016.08.043] [Medline: 29180154]
32. See S, Conry JM. Flip my class! A faculty development demonstration of a flipped-classroom. Curr Pharm Teach Learn 2014 Jul;6(4):585-588. [doi: 10.1016/j.cptl.2014.03.003]
33. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. J Psychosom Res 2002 Feb;52(2):69-77. [doi: 10.1016/S0022-3999(01)00296-3]
34. Cohen S, Sherrod DR, Clark MS. Social skills and the stress-protective role of social support. J Pers Soc Psychol 1986 May;50(5):963-973. [doi: 10.1037//0022-3514.50.5.963] [Medline: 3486973]
35. Mirani SH, Shaikh NA, Tahir A. Assessment of clinical empathy among medical students using the Jefferson scale of empathy-student version. Cureus 2019 Feb 28;11(2):e4160. [doi: 10.7759/cureus.4160] [Medline: 31058043]
36. Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false discovery rate. Biometrika 2006 Sep 1;93(3):491-507 [FREE Full text] [doi: 10.1093/biomet/93.3.491]
37. Naing C, Whittaker MA, Aung HH, Chellappan DK, Riegelman A. The effects of flipped classrooms to improve learning outcomes in undergraduate health professional education: a systematic review. Campbell Syst Rev 2023 Sep;19(3):e1339. [doi: 10.1002/cl2.1339] [Medline: 37425620]
38. Peng W, Xiong Y, Wei J, et al. Flipped classroom improves student learning outcome in Chinese pharmacy education: a systematic review and meta-analysis. Front Pharmacol 2022;13:936899. [doi: 10.3389/fphar.2022.936899] [Medline: 36110553]
39. Gillette C, Rudolph M, Kimble C, Rockich-Winston N, Smith L, Broedel-Zaugg K. A meta-analysis of outcomes comparing flipped classroom and lecture. Am J Pharm Educ 2018 Jun;82(5):6898. [doi: 10.5688/ajpe6898] [Medline: 30013248]
40. White PJ, Larson I, Styles K, et al. Adopting an active learning approach to teaching in a research-intensive higher education context transformed staff teaching attitudes and behaviours. HERD 2016 May 3;35(3):619-633. [doi: 10.1080/07294360.2015.1107887]
41. He Y, Lu J, Huang H, et al. The effects of flipped classrooms on undergraduate pharmaceutical marketing learning: a clustered randomized controlled study. PLoS ONE 2019;14(4):e0214624. [doi: 10.1371/journal.pone.0214624] [Medline: 30969976]
42. Anderson HG Jr, Frazier L, Anderson SL, et al. Comparison of pharmaceutical calculations learning outcomes achieved within a traditional lecture or flipped classroom andragogy. Am J Pharm Educ 2017 May;81(4):70. [doi: 10.5688/ajpe81470] [Medline: 28630511]
43. Missildine K, Fountain R, Summers L, Gosselin K. Flipping the classroom to improve student performance and satisfaction. J Nurs Educ 2013 Oct;52(10):597-599. [doi: 10.3928/01484834-20130919-03] [Medline: 24044386]
44. Kugler AJ, Gogineni HP, Garavalia LS. Learning outcomes and student preferences with flipped vs lecture/case teaching model in a block curriculum. Am J Pharm Educ 2019 Oct;83(8):7044. [doi: 10.5688/ajpe7044] [Medline: 31831896]
45. McLaughlin JE, White PJ, Khanova J, Yuriev E. Flipped classroom implementation: a case report of two higher education institutions in the United States and Australia. Comput Sch 2016 Jan 2;33(1):24-37. [doi: 10.1080/07380569.2016.1137734]
46. Akçayır G, Akçayır M. The flipped classroom: a review of its advantages and challenges. Comput Educ 2018 Nov;126:334-345. [doi: 10.1016/j.compedu.2018.07.021]

XSL•FO
RenderX

47. Freeman S, O'Connor E, Parks JW, et al. Prescribed active learning increases performance in introductory biology. CBE Life Sci Educ 2007;6(2):132-139. [doi: 10.1187/cbe.06-09-0194] [Medline: 17548875]
48. Ranjan P, Kumari A, Chakrawarty A. How can doctors improve their communication skills? J Clin Diagn Res 2015 Mar;9(3):JE01-JE04. [doi: 10.7860/JCDR/2015/12072.5712] [Medline: 25954636]
49. Maguire P, Pitceathly C. Key communication skills and how to acquire them. BMJ 2002 Sep 28;325(7366):697-700. [doi: 10.1136/bmj.325.7366.697] [Medline: 12351365]
50. Perrin J, Meeus A, Broseus J, et al. A serious game about hematology for health care workers (SUPER HEMO): development and validation study. JMIR Serious Games 2023 Feb 13;11:e40350. [doi: 10.2196/40350] [Medline: 36780215]
51. Lang B, Zhang L, Lin Y, Han L, Zhang C, Liu Y. Team-based learning pedagogy enhances the quality of Chinese pharmacy education: a systematic review and meta-analysis. BMC Med Educ 2019 Jul 29;19(1):286. [doi: 10.1186/s12909-019-1724-6] [Medline: 31357986]

Abbreviations

IC: inverted classroom
MCQ: multiple-choice question
PharmD: Doctor of Pharmacy

Edited by TDA Cardoso; submitted 11.10.24; peer-reviewed by D Schwarz, MM Vasquez, S Arsić; revised version received 25.02.25; accepted 12.03.25; published 10.04.25.

Please cite as:
Jourdi G, Selmi M, Gaussem P, Truchot J, Margaill I, Siguret V
Evaluation of the Inverted Classroom Approach in a Case-Study Course on Antithrombotic Drug Use in a PharmD Curriculum: French Monocentric Randomized Study
JMIR Med Educ 2025;11:e67419
URL: https://mededu.jmir.org/2025/1/e67419
doi:10.2196/67419

© Georges Jourdi, Mayssa Selmi, Pascale Gaussem, Jennifer Truchot, Isabelle Margaill, Virginie Siguret. Originally published in JMIR Medical Education (https://mededu.jmir.org), 10.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Preclinical Medical Students' Perspectives and Experiences With Structured Web-Based English for Medical Purposes Courses: Cross-Sectional Study

Radhakrishnan Muthukumar[1], MBBS, MClinEmbryol, PhD; Isaraporn Thepwongsa[2], MD, MFM, PhD; Poompong Sripa[3], MD; Bangonsri Jindawong[2], MSc, PhD; Kamonwan Jenwitheesuk[1], MD; Surapol Virasiri[1], MD

[1]Academic Affairs, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand

[2]Department of Community, Family and Occupational Medicine, Faculty of Medicine, Khon Kaen University, 123 Mittraparb Road, Khon Kaen, Thailand

[3]Inverkeithing Medical Group, Inverkeithing, United Kingdom

Corresponding Author:
Isaraporn Thepwongsa, MD, MFM, PhD
Department of Community, Family and Occupational Medicine, Faculty of Medicine, Khon Kaen University, 123 Mittraparb Road, Khon Kaen, Thailand

Abstract

Background: English for medical purposes (EMP) is essential for medical students as it serves as a foundational language for medical communication and education. However, students often undervalue its importance within the medical curriculum. Given their demanding schedules and workload, educational methods for EMP must align with their needs. Structured web-based learning offers flexibility and convenience, yet limited research has explored its exclusive application for EMP in undergraduate medical education.

Objective: This study aimed to evaluate medical students' perspectives on structured web-based EMP courses and assess their impact on medical English proficiency using objective and subjective measures.

Methods: Structured web-based EMP courses were developed based on evidence-based guidelines, addressing barriers to web-based learning during development and implementation. A cross-sectional study was conducted with 535 medical students who completed these courses. Data were collected via questionnaires, the learning management system, and the Khon Kaen University Medical English Test (KKUMET), which assessed proficiency in listening, reading, writing, and speaking. Data were analyzed using descriptive statistics.

Results: Of the 535 students, 452 (84.5%) completed the survey. Participants reported confidence in reading (mean 4.11, SD 0.87), vocabulary (mean 4.04, SD 0.84), and listening skills (mean 4, SD 0.89), but lower confidence in writing skills (mean 3.46, SD 1.07). The KKUMET results showed statistically significant improvements in all 4 language skills after course completion ($P<.001$). The top-rated benefits of the courses were convenience (mean 4.77, SD 0.59), sufficient instruction (mean 4.5, SD 0.85), and clear content (mean 4.41, SD 0.80).

Conclusions: Structured web-based EMP courses are relevant and well received by medical students. These courses significantly improve students' medical English proficiency, as evidenced by both subjective feedback and objective measures. Medical educators should consider integrating structured web-based EMP programs to better support students' language proficiency in medical contexts.

(JMIR Med Educ 2025;11:e65779) doi:10.2196/65779

KEYWORDS

English for medical purposes; online course; online learning; online education; medical students; medical school; online; online learners; perspectives; English; English language; medical research; educational method; lesson; course; instructional designs; English for medical professional; EMP; barriers; web-based

Introduction

Background

English is the international language of medicine [1,2]. It is the dominant language used in medical research and publications [1,3]. A growing number of medical journals worldwide have been published in English [4]. Almost 90% of the world's publications indexed in the MEDLINE database are published in the English language [3,5]. In addition, considering their high impact factors and citation scores, almost all quality research articles have been published in English-medium journals [1].

XSL•FO
RenderX

The language of medicine has shifted from Latin and Greek to English, influencing the development of modern medical terminology [3]. The use of English in a medical community involves communication with colleagues and staff, reading medical journals, giving conference presentations, or pursuing postgraduate education in English-speaking countries [1,3]. Therefore, the English language and literacy are essential in medical curricula [2,4,6].

General English focuses on overall language competency, emphasizing foundational grammar, vocabulary, and communication skills applicable to a wide range of everyday and academic contexts. English for medical purposes (EMP), on the other hand, is a specialized subset of English tailored to the medical field. It emphasizes medical terminology, reading and comprehension of medical texts and research articles, professional communication skills that are required for clinical and academic purposes, such as writing case reports, interacting with colleagues, and participating in medical discussions [7].

Approximately half of medical schools in the world use English as a medium of instruction [8]. The current status of English use with other languages in medicine is that "English is one lingua franca in medicine but speaks many tongues" [5]. This means that although medical teaching and patient interaction are in the local language, medical professionals still use medical languages with specialized vocabulary in their professional communication [3]. For students or professionals whose English is not their first language, using medical English for their professional-related activities is much more challenging and demanding than it is for those whose English is their first language [3].

The assessment of learning needs is considered fundamental to EMP curriculum design [2,4,7,9] and should be identified early [10]. The learner-centered approach is used to obtain needs from the learner's needs assessment. The learner-centered approach empowers learners to decide what, how, and where to learn, while teachers act as facilitators [11,12]. Learning needs assessments help define program goals and specific teaching objectives [7,13], which leads to the development of lesson plans, materials, tests, assignments, and activities. Many studies have examined the need for EMP [2,9,13-15]. Medical students need to read medical literature, listen to lectures, and write clinical reports and short essays, whereas practitioners need patient interaction and conference participation [1,7]. The development of EMP courses addresses specific language challenges that hinder medical students' learning, particularly in contexts where English is partially used as the medium of instruction, while local languages dominate communication between instructors, students, and patients [16].

Teaching an EMP course differs from teaching general English [7,14]. Instead of learning grammar and fundamental structures, the goal of learning English at this level is to apply the language to medical studies [7]. Previous studies have described the development of EMP courses, including detailed information about curriculum development processes based on students' learning needs [2,16-18], as well as EMP courses developed for international medical graduates [18] and undergraduate medical students [2,16,17]. However, these EMP courses were taught in medical curricula in countries where English was the official language or medium of instruction. In other countries, the EMP course design has not yet been clearly explored. As a result, the requirements of EMP for medical students must be investigated [19], and EMP instruction must be tailored to the intended audience needs. In addition, many studies have focused on preparing students with linguistic tasks for their future careers [9,15]. Less is focused on the design of EMP courses aiming to improve academic performance during undergraduate studies, preclinical years in particular, where basic science knowledge is most of the content rather than clinical application.

Educational methods for EMP should also be matched with learners' needs and preferences. Online and web-based educational methods have been adopted in medical education for the past decade [20]. Online teaching was the primary modality of instruction offered to undergraduate medical students during the COVID-19 pandemic [21]. The web-based education methods have been tested for their effectiveness [22-24]. The advantages of web-based educational methods include improving medical students' knowledge and skills [24]. However, prior meta-analyses have demonstrated that web-based learning is just as effective as the traditional learning methods [23,24]. There are few studies on the exclusive use of web-based learning techniques in EMP. One study explored students' readiness for internet-based learning in EMP [25] or internet- and computer-based as part of blended learning for general English [26,27]; however, none has solely used web-based educational methods in EMP courses for medical students.

Problem Statement

Despite the growing adoption of web-based learning methods in medical education [28], limited research has explored their exclusive use in teaching EMP, especially for preclinical medical students [24]. Web-based education, proven to be as effective as traditional methods [22,23,29,30], offers unique advantages such as flexibility and accessibility [24]. However, few studies have investigated the development and evaluation of structured web-based EMP courses tailored to students in non–English-speaking environments.

Thus, this study developed an exclusive structured web-based learning program for EMP and evaluated its effectiveness. Grounded in constructivist principles, the study emphasizes active, self-directed learning. The structured web-based EMP courses were designed to enable students to construct knowledge through interactive activities, scenario-based learning, and multimedia resources. These approaches foster engagement and empower learners to take ownership of their educational journey, ensuring the content is relevant and applicable to their academic and professional needs. The guiding research question for this study was whether a structured web-based learning program exclusively for EMP would be accepted, considered relevant, and meet the satisfaction of preclinical medical students. We hypothesized that learning EMP through structured web-based courses would be both relevant and well-received by preclinical medical students. This investigation formed the basis for evaluating the course. To comprehensively evaluate the course, the research is divided into 2 parts. The first part, the present study, focuses on assessing the relevance and acceptance of

XSL•FO
RenderX

structured web-based EMP courses based on students' perceptions of the web-based learning mode. The second part examines the effectiveness of the structured web-based EMP courses in improving students' medical English proficiency. To evaluate the first part, a cross-sectional descriptive study was conducted to assess the relevance and acceptance of the structured web-based EMP courses among participating medical students. Additionally, feedback from course instructors, gathered during the course design and development phases, was incorporated to capture their perceptions of the web-based EMP program. The findings from the second part are discussed in a separate article; however, this study included significant findings on the program's effects on students' medical English proficiency. The findings are intended for medical educators, curriculum designers, and policy makers in medical education, particularly those serving non–English-speaking regions.

Our Medical School Curriculum

Our medical school recruits students directly from high school through the Thailand University Central Admission System (TCAS). TCAS, implemented in 2018, is an admission framework consisting of 5 rounds held annually. Medical schools in Thailand often adopt unique criteria and processes within TCAS, reflecting their specific admission requirements and competitive nature [31]. With 288 students enrolled each year, our medical school is one of the regional institutions in northeastern Thailand. The undergraduate medical curriculum spans 6 years, including the first 3 years focused on medical sciences and the last 3 years on clinical practice [32]. Our 2019 updated medical curriculum states that medical students must earn 259 credits over 6 years of medical school. In the first year, premedical education courses (the majority are general education and general principles for medical sciences) account for 38.5 credits. In the second and third years, preclinical education courses comprise 76.5 credits. Clinical rotations, accounting for 48 credits, are conducted in the sixth year, while clinical education courses make up 96 credits across the fourth and fifth years [33]. The majority of the lecture slides, suggested texts, teaching and learning materials, and exam questions are in English, even though the instruction at our school is in Thai.

It is mandatory for Thais to learn English from primary school till higher education [34,35]. In Thailand, the English component taught in high school is of a very basic level, the general English proficiency is low and unsatisfactory [36,37], and by itself is not sufficient for medical professional courses. English proficiency is considered as a part of the entry criteria to the medical schools. Despite English proficiency being a part of the weighted formula for entering medical schools, at present, there is no specific cutoff score or a minimal standardized English language proficiency requirement for entering a medical school in Thailand [38]. English as a foreign language, therefore, is a requirement for students at our medical school [33]. Once students are admitted, they begin enrolling in English courses starting in their first year of study. During medical school, students are required to complete 6 English courses. Four general English courses are taken during the first 3 years of medical school and are taught by nonmedical staff from the university's language institute. They focus on foundational grammar, vocabulary, and communication skills, providing a general linguistic foundation for academic and social contexts. Two EMP courses are introduced in the second and third years of medical school. The content focuses on medical terminology and language skills necessary for academic and clinical tasks, such as reading medical literature, writing reports, and engaging in clinical communication. The curriculum shifts from general language acquisition in the initial years to specialized language skills tailored to medical contexts as students advance in their studies.

Annual feedback from our students consistently indicates that they perceive limited relevance of general English to their medical studies. This perception may stem from several factors. Despite years of English education, many students struggle with basic speaking skills due to insufficient vocabulary and grammar knowledge [39]. Additionally, a lack of intrinsic motivation often leads students to view general English as a mandatory requirement rather than a valuable skill for personal or professional development [40]. The culture within medical programs often emphasizes the need for English proficiency tailored to medical contexts, which shapes students' perceptions and priorities [41]. Therefore, they could not match learning general English in the medical curriculum.

The findings from our students' needs assessments revealed that our medical students needed to improve their English proficiency and wanted the school to organize a test for their English proficiency [42]. They preferred a self-directed web-based learning method and teachers who were both English language experts and medical professionals [42]. Based on the needs assessments, students value medical English because it directly aligns with their academic and professional goals. Unlike general English, medical English equips them with the specific skills needed to read medical literature, participate in clinical discussions, and engage effectively within global medical communities [41,43]. Needs assessments and feedback consistently emphasize students' preference for EMP courses over general English, highlighting the practical benefits they associate with EMP in their medical studies and future careers. Based on learning needs assessments, 2 additional EMP courses were developed while 2 general English courses were removed from the medical curriculum. The newly developed EMP courses were English for Medical Purposes I for the second-year medical students and English for Medical Purposes II for the third-year medical students, which were launched at the same time in the academic year 2021.

Proficiency in medical English is considered to enhance students' academic performance, evidenced by a positive correlation between English proficiency and academic success among medical students [44,45], and to support their development as independent learners [15]. However, evaluating the extent to which students achieve independent learning is beyond the scope of this study.

XSL•FO
RenderX

Methods

Study Objectives

This study aims to evaluate the relevance, acceptance, and satisfaction of structured web-based EMP courses among preclinical medical students.

Study Design

A cross-sectional study was conducted to explore medical students' opinions on learning EMP in a structured web-based course, and qualitative insights from instructors' feedback were obtained.

Participants

The participants were second-year medical students enrolled in the English for Medical Purposes I course and third-year medical students enrolled in the English for Medical Purposes II course at the time of the academic year 2021.

The WinPepi, a statistical software package, was used to calculate sample size based on the proportion of medical students who acknowledged web-based learning [46]. Assuming a proportion of 0.73, a population size of 535, a design effect of 1, and a significance level of 0.05, a total number of 121 was sufficient. However, all 535 students were included in this study to avoid a potential source of selection bias. It is important to clarify that this power analysis was performed for the broader research project, which consists of 2 parts. The first part, detailed in this manuscript, explores medical students' perceptions and experiences with web-based EMP learning. The second part examines the effects of the structured web-based EMP courses on changes in students' English language proficiency after course completion, which is reported in a separate article.

Development of Structured Web-Based EMP Courses

Based on our medical students' needs, EMP course objectives were set accordingly. The main objective of the EMP course was to enhance students' medical English proficiency in all 4 core English language skills (reading, writing, listening, and speaking) through structured, targeted, and interactive web-based practice.

An overview of the four main English language skills is as follows. (1) Reading: Defined as the comprehension of medical literature and academic texts relevant to the students' year of medical study. Instructional strategies include guided analysis of medical articles, scenario-based reading exercises, and formative assessments such as quizzes. (2) Writing: Focuses on effective medical note-taking and medical essays, emphasizing medical content, structure, vocabulary, readability, precision, and clarity. Instructional strategies involve structured writing guides, assignments, peer reviews, and scenario-based tasks such as summarizing patient cases. (3) Listening: Aimed at developing comprehension of medical conversations from case studies and lectures. Instructional methods include exposure to scenario-based audio materials, repeated listening tasks, and interactive multimedia resources. (4) Speaking: Enhances verbal communication in medical contexts, including explaining medical terms, conducting history taking, discussing treatment plans, and educating patients. Instructional strategies feature medical speaking guides, public speaking guides, role-play, video-recorded practice sessions, and feedback loops. Performance improvement for each skill is measured through targeted assessments conducted before and after the course. Standardized testing for reading and listening comprehension, comparison of baseline and final writing samples to assess coherence and technical accuracy, and evaluations of verbal communication in simulated scenarios are used to document and demonstrate progress effectively. These details ensure that the EMP course objectives are transparent and tied directly to improving students' proficiency in medical English through targeted and practical methods.

To develop an effective structured web-based course, we followed the steps of Hays and Veitch's recommendations [47] and applied Cook and Dupras's guide [48]. Key principles of Hays and Veitch emphasize the importance of conducting a thorough needs assessment to identify gaps in knowledge, skills, and practices while ensuring relevance to participants' professional roles and daily practices. They recommend interactive methods such as case studies, group discussions, and workshops to foster engagement and promote practical learning. Flexible delivery formats accommodate diverse learning preferences and schedules, while regular program evaluations ensure continuous improvement, sustainability, and adaptability to emerging needs and advancements in medical education. Cook and Dupras provide a guide specifically for developing web-based learning programs. Their framework focuses on conducting needs assessments, defining clear and measurable learning objectives, and designing content tailored to learners' needs. They highlight the importance of interactive and multimedia-rich content to enhance engagement, as well as ensuring usability and accessibility for diverse learners. Evaluation methods, including pre- and postassessments, combined with continuous feedback and iterative improvements, are essential to measure learning outcomes and maintain program quality. Together, these frameworks provide a robust foundation for creating effective, learner-centered programs that meet educational goals and align with the needs of preclinical medical students. Barriers to web-based learning were considered [49]. More details regarding the development of EMP courses are provided in Multimedia Appendix 1.

The EMP courses were uploaded to a customized web-based learning management system (LMS) developed by our medical school [29]. Each EMP course ran throughout the semester for approximately 48 weeks, and the students' learning schedule was approximately 3 hours per week. However, students can manage their time to study as scheduled or at any time that suits them (see more details on developing the web-based EMP course in Multimedia Appendix 1). Data from the LMS monitoring system were used to report student enrollment and completion of the EMP courses.

Data Sources, Questionnaire, and Assessment

To evaluate the course, questionnaires were used as tools to gather student perspectives on various aspects of the courses, including their confidence in medical English skills, satisfaction with the content, and perceived benefits of web-based learning. These tools provided quantitative data on student experiences

XSL•FO
RenderX

and attitudes. Additionally, feedback from instructors on course design and delivery was integrated into the evaluation process to ensure alignment with educational objectives and identify areas for improvement. Together, these methods offered both quantitative and qualitative insights into the evaluation of the web-based EMP courses.

An online administered questionnaire was developed based on the literature on the use and the evaluation of web-based learning [46,50,51]. The following constructs of web-based learning effectiveness and attitudes toward the structured web-based EMP course were included in the questionnaire: individual learners, confidence in medical English skills, perceived factors influencing web-based participation, perceived satisfaction with content and instructional designs, perceived ease of use, perceived advantages and barriers to web-based learning, and perceived outcomes and benefits of web-based EMP courses. The questionnaire included a 5-point Likert scale ranging from disagree to agree, which was used to assess learners' agreement with each item. For learners' confidence in their medical English skills, the constructs included a 5-point Likert scale ranging from not confident to strongly confident. For the factors influencing web-based learning participation, the constructs included a 5-point Likert scale, ranging from no influence to influence. For learners' perceived benefits of web-based EMP courses, the constructs included a 5-point Likert scale, ranging from not beneficial to strongly beneficial.

To ensure the face and content validity of the questionnaire, each item was evaluated thoroughly by 3 separate experts in the field of medical education at our school. Thirty students participated in a pilot test of the questionnaire, which led to revisions. The Cronbach α coefficient of the attitude and experience part of the questionnaire was 0.93.

The details of the course instructions, content, and materials are provided in Multimedia Appendix 1. The course media included PDF files, audio, videos, and external links for the downloads. The second-year medical students completed a 45-hour web-based asynchronous English for Medical Purposes I course. The third-year medical students completed a 45-hour web-based asynchronous English for Medical Purposes II course. The students would be considered to have completed the module if that module was accessed at least 50% of the total learning time for that module because the LMS has a double speed-up function for playing video or audio clips.

After participants completed the web-based EMP courses, they were invited to complete the questionnaire. A link to the online questionnaire on Google Forms was added at the end of each course. The students completed the questionnaire between December 2021 and April 2022. The questionnaire data were transferred from Google Sheets, and data from the LMS monitoring system were exported to Microsoft Excel. The data were then gathered and checked for completion before being transferred to IBM SPSS for Windows.

Of note, the summative assessment for medical English proficiency was conducted using the Khon Kaen University Medical English Test (KKUMET), which examined proficiency in listening, reading, writing, and speaking. Each component was assessed before and after completing the EMP courses. The baseline EMP proficiency of the participating students, as assessed before enrollment in the 2 EMP courses, revealed that more than two-thirds of the students were at beginner or intermediate levels. The instructional goal of these EMP courses was to enhance their proficiency to intermediate and advanced levels. This study focuses on highlighting final outcomes; detailed results of the EMP proficiency tests conducted before and after these web-based EMP courses are reported in a separate article.

Statistical Analysis

Data from questionnaires were analyzed using IBM SPSS for Windows version 26.0. A pairwise deletion strategy was applied to handle the missing data. The demographic data were described using descriptive statistics. The participants' responses on a 5-point Likert scale to a 42-item questionnaire on their web-based learning experiences were dichotomized by calculating the mean scores and SD. Mean scores of 3.5 and above were considered agreed upon.

Ethical Considerations

This study was approved by the Human Research Ethics Committee of Khon Kaen University (project number: HE631465). Students were recruited through the researcher's assistant, who invited volunteers to participate. Before completing the questionnaires, participants were informed that participation was voluntary and that they could drop out of the study at any time. They were informed that their opinions were important for enhancing the medical English courses and were therefore encouraged to express them. All students voluntarily agreed to participate without receiving compensation. The participants' privacy and identity were protected, and confidentiality was assured in that no identifying information was asked. The study objectives were explained to the participants, and the study was conducted according to the academic ethical code.

Results

Participating Medical Student Demographics

A total of 535 medical students completed the web-based EMP courses, 452 of whom started and returned the completed questionnaires (response rate: 84.5%). Table 1 presents the participants' demographic data. The numbers of male and female participants were relatively similar, as were the numbers of medical students each year.

XSL•FO
RenderX

Table . Demographic data of the participants.

Demographics	Values
Age (n=451), years	
Range	18 - 24
Mean (SD)	20.37 (0.74)
Sex (n=452), n (%)	
Male	219 (48.5)
Female	222 (49.1)
Prefer not to say	11 (2.4)
Year of study (n=450)	
2	231 (51.3)
3	219 (48.7)

Influence of Web-Based Learning Characteristics

Respondents rated the degree of influence of the characteristics of web-based learning on their participation in web-based EMP courses (Table 2). Convenience, flexibility, and accessibility were rated the highest, while facilitator interactions received lower ratings.

Table . Rating of the influence of web-based learning characteristics on participation in web-based EMP[a] courses (n=452).

Characteristics of web-based learning	The degree of influence on adoption of web-based EMP courses, mean[b] (SD)
The convenience of completion the courses at any time or place	4.82 (0.49)
Flexibility to complete and save small sections at a time	4.80 (0.56)
Easy to access the course content	4.60 (0.75)
Easy to use/complete the course	4.58 (0.71)
Access to other useful links and resources	4.53 (0.84)
Instant access to feedback and the right answers when completing quizzes	4.50 (0.85)
The quality of content	4.46 (0.75)
Accessibility to technical support if difficulties are encountered	4.35 (0.94)
The use of case-based information and discussion	3.99 (1.08)
Facilitator's regular input/participation	3.77 (1.17)
The opportunity to communicate/interact with the facilitator	3.75 (1.14)

[a]EMP: English for medical purposes.

[b]Mean was calculated using a 5-point Likert scale ranging from 1 (strongly no influence), 2 (no influence), 3 (neutral), 4 (influence), and 5 (strongly influence).

Confidence in Medical English Skills

Respondents' confidence in their medical English skills is shown in Table 3. The participating medical students reported feeling confident about medical English reading, vocabulary, and listening skills but were not sure about their writing skills.

XSL•FO
RenderX

Table . The participant's confidence in their medical English skills after completing the medical English courses (n=452).

Confidence in medical English skills after completing the modules	Values, mean[a] (SD)
Medical English reading skill	4.11 (0.87)
The use of medical English vocabulary	4.04 (0.84)
Medical English listening skill	4.00 (0.89)
Applying professional-specific knowledge and skills in English	3.92 (0.91)
Medical English speaking skill	3.50 (1.05)
Medical English writing skill	3.46 (1.07)

[a]Mean was calculated using a 5-point Likert scale ranging from 1 (strongly not confident), 2 (not confident), 3 (not sure), 4 (confident), and 5 (strongly confident).

Perceived Advantages, Barriers To, and Attitudes Toward Instructional Designs of Web-Based EMP Courses

The participants identified useful and beneficial aspects of web-based EMP courses. The top 3 highly rated participants agreed on the advantages of convenience, sufficient instructions, and clear and easy-to-understand content (Table 4). For the instructional designs of the web-based EMP courses, the participants agreed on the clarity and appropriateness of the overall design of the courses, including clear objectives and content, appropriate content, arrangement, instruction, media, delivery method, course assessment, and grading (Table 5).

Table . The participants' web-based learning experience of the medical English courses (n=452).

	Values, mean[a] (SD)
Advantages of the web-based medical English modules	
I was able to learn at any place	4.77 (0.59)
Overall, this web-based course provided me with adequate instruction	4.50 (0.85)
The content was easy to understand and clear	4.41 (0.80)
Overall, the course contents covered its objectives	4.30 (0.87)
I felt more comfortable learning in this web-based course than in the face-to-face session	4.19 (1.13)
Overall, the instruction I obtained from this web-based program was motivating	4.16 (0.99)
I knew how to contact the facilitator and the facilitator responses promptly to my questions	4.04 (1.02)
The web-based program fulfilled my learning needs to improve my medical English skills	3.93 (1.04)
If I had technical problems during participating in this program, I received adequate help with technical problems (n=66)[b]	3.30 (1.05)
Difficulties in accessing and completing the course	
There was too much basic, well-known information in the course	3.06 (1.14)
The module took too long to complete	3.03 (1.28)
I spent more time in access to a computer to access this web-based program	2.31 (1.46)
The internet connection was very slow	2.25 (1.42)
I spent more time in downloading external links	2.08 (1.35)
The course was not useful for me because I do not have adequate computer skills to complete the course	1.95 (1.41)

[a]Mean was calculated using a 5-point Likert scale ranging from 1 (Strongly disagree), 2 (disagree), 3 (neutral), 4 (agree), and 5 (strongly agree).

[b]When the participants were asked if they had technical problems during participating in this program, of the 452 participants, 66 (14.6%) had, while 386 (85.4%) had not.

XSL•FO
RenderX

Table . The participants' rating on the web-based medical English course contents and instructional designs (n=452).

The web-based course contents and instructional designs	Values, mean[a] (SD)
The course had clear learning objectives	4.48 (0.79)
The contents were complete, appropriate, and relevant to the objectives	4.41 (0.82)
The order of contents was arranged properly	4.41 (0.82)
The learning media (eg, audios, videos, and PDF files) were appropriate	4.39 (0.90)
Overall, the instructional design of the program was appropriate	4.38 (0.85)
The web-based teaching was appropriate	4.37 (0.86)
The course's assessments (formative assessments on listening, reading, writing, and speaking) were appropriate	3.97 (1.05)
Grading criteria were appropriate	3.74 (1.07)

[a]Mean was calculated using a 5-point Likert scale ranging from 1 (Strongly disagree), 2 (disagree), 3 (neutral), 4 (agree), and 5 (strongly agree).

Perceived Learning Outcomes

The participants agreed that their 4 main skills improved. They also believed that all types of teaching media and lectures benefited them (Table 6).

Table . The participants' perceived outcomes to and benefits of the web-based medical English courses after the course completion (n=452).

	Values, mean (SD)
The participants' perceived outcomes after the course completion[a]	
I improved my listening skills	4.37 (0.85)
I improved my reading skills	4.23 (0.90)
I gained knowledge on how to practice my medical English skills	4.21 (0.89)
I improved my writing skills	3.74 (1.15)
I improved my speaking skills	3.68 (1.21)
Benefits of the web-based medical English courses after the course completion[b]	
The audios to practice listening skills	4.45 (0.83)
The media for medical terms to learn medical terms and practice reading, speaking, and writing skills	4.36 (0.83)
The videos to practice reading, listening, and speaking skills	4.35 (0.91)
The lectures on listening and reading	4.34 (0.84)
The course was organized in the modules	4.28 (0.92)
The recommended reading articles to practice reading and writing skills	4.07 (0.99)
The lectures on scientific writing	4.07 (1.00)
The lectures on speaking	4.02 (1.04)

[a]Mean was calculated using a 5-point Likert scale ranging from 1 (strongly disagree), 2 (disagree), 3 (neutral), 4 (agree), and 5 (strongly agree).

[b]Mean was calculated using a 5-point Likert scale ranging from 1 (strongly not beneficial), 2 (not beneficial), 3 (neutral), 4 (beneficial), and 5 (strongly beneficial).

Learning Behaviors and Engagement Patterns (Data From the LMS Monitoring System)

The second-year students (n=263) completed the learning module ranging from 0 to 7 modules (mean 6.54, SD 1.41). The total learning time ranged from 172 to 24,439 minutes, with a mean of 925.02 (SD 1759.80) minutes. Regarding students' learning behaviors, 153 (58.2%) out of 263 students spent less than 10 hours on the course and often logged in to the course during the afternoon (1069/2993 logins, 35.72%) and evening hours (1005/2993 logins, 33.58%).

The third-year students (n=272) completed the learning module ranging from 0 to 7 modules (mean 6.61, SD 1.41). The total learning time ranged from 0 to 14,100 minutes, with a mean of 768.59 (SD 930.56) minutes. Regarding students' learning behaviors, 136 (50%) students spent less than 10 hours on the course and often logged in to the course during the evening

XSL•FO
RenderX

(1000/2962 logins, 33.76%) and morning hours (655/2962 logins, 22.11%).

Improvements in Medical English Proficiency

The summative assessment of medical English proficiency was conducted using KKUMET, which evaluates listening, reading, writing, and speaking skills. Each component was assessed before and after the EMP courses. The results showed statistically significant improvements in students' scores across all 4 skills, as well as in the total test scores ($P<.001$), underscoring the effectiveness of the courses. Additional details about the summative testing results are presented in a separate article.

Instructor Feedback on Course Design and Implementation

During the development and implementation of the structured web-based EMP courses, all course instructors actively participated in providing feedback. This feedback was gathered through regular meetings, written evaluations, and iterative reviews of course content and delivery methods. Instructors emphasized the importance of aligning the courses with medical students' academic and professional needs. Their insights directly influenced several key aspects of the course design.

First, scenario-based learning was integrated to make the content more applicable to clinical practice. Case-based scenarios were incorporated into reading and listening exercises to enhance the relevance and practicality of the material. Second, interactive assessments were adapted based on instructor feedback to emphasize practical application. These included the use of peer reviews for writing assignments and role-play activities for speaking exercises, allowing students to practice and refine their skills in realistic scenarios. Third, instructors highlighted the need for multimedia resources to create a more engaging learning experience. This led to the inclusion of audio clips, videos, and interactive tasks designed to cater to various learning preferences. Finally, adaptive scheduling was implemented to accommodate the heavy workloads of medical students. The courses were designed to be asynchronous, enabling students to learn at their own pace and manage their schedules effectively.

Instructors reported that the structured design of the courses, combined with the iterative feedback process, resulted in a cohesive program that effectively addressed students' learning needs. Their ongoing involvement ensured that the courses remained both relevant and practical for medical students.

Discussion

Principal Findings

This study explored the perspectives of preclinical medical students on structured web-based EMP courses and evaluated their proficiency improvements. A total of 535 students enrolled in and completed the courses by the published due dates, of which 452 (84.5%) students completed the questionnaire. The high response rate can be attributed to several factors. The survey's integration at the end of the web-based EMP courses ensured participants encountered it immediately after course completion, when their experiences were fresh and engagement levels were high [52]. The web-based format allowed flexibility, enabling participants to complete the survey at their convenience [53]. Clear instructions, the assurance of anonymity, and the perceived relevance of the survey further encouraged participation [54]. Additionally, students' satisfaction with the course content and instructional design likely motivated them to provide feedback [52].

The summative assessment, conducted using KKUMET, revealed statistically significant improvements across all 4 English language skills (listening, reading, writing, and speaking). Participants reported high confidence in reading, vocabulary, and listening skills but expressed lower confidence in writing and speaking skills. Students rated convenience, clarity of content, and sufficient instruction as the top benefits of the courses. These results affirm the relevance, acceptance, and satisfaction of structured web-based EMP courses for medical students, aligning with the primary objectives of improving proficiency and fostering recognition of EMP's importance in supporting their academic learning.

Evaluation of Course Effectiveness

The primary goal of the EMP courses was to improve students' proficiency in all 4 English language skills and introduce relevant medical terminology tailored to their academic content. This goal was achieved, as evidenced by measurable improvements in KKUMET scores and student-reported confidence levels.

We evaluated the effectiveness of web-based EMP courses, finding significant improvements in students' medical English proficiency across all skills, as measured by KKUMET scores, and increased confidence in reading, vocabulary, and listening. By comparing subjective learner feedback with objective results, we confirmed that reported confidence aligned with measurable gains, minimizing potential overestimation from cognitive biases like the Dunning-Kruger effect [55,56]. These findings underscore the importance of validating self-reported outcomes with objective data to ensure reliable assessments of learning impact [48,57], with future research recommended to explore the influence of cognitive biases and interventions to enhance learning outcomes [55,58].

Furthermore, the courses aimed to highlight the importance of EMP in supporting learning across other subjects, particularly through contextual and content-based design [7,15]. Students expressed initial difficulty in understanding the importance of general English and its relevance within the medical curriculum. However, they highly valued the structured web-based EMP courses and expressed satisfaction with the content, particularly due to its alignment with their academic and professional needs. This alignment was achieved by designing and developing courses that were self-directed and created by medical professionals, which students recognized as enhancing their engagement and acceptance of the educational approach. While students did not explicitly indicate a need for increased interaction with faculty facilitators, they expressed a preference for instructors who were both English language experts and knowledgeable about medical content. This preference suggests an implicit desire for contextualized guidance, highlighting the

XSL•FO
RenderX

value students place on instructors with dual expertise. The structured web-based EMP courses effectively met these needs, motivating them to integrate EMP learning into their study schedules, underscoring the importance of tailoring course design to align with student preferences.

Comparison With Previous Studies

The findings of this study align with existing research on the effectiveness of web-based learning in medical education while addressing gaps specific to EMP. Prior studies have emphasized the effectiveness of web-based learning in enhancing knowledge and skills in medical education [23,24]. Similarly, this study demonstrated statistically significant improvements in all 4 core English language skills (listening, reading, writing, and speaking) among medical students enrolled in structured web-based EMP courses. These findings confirm that well-designed web-based educational modalities can be as effective as traditional methods.

Students greatly appreciated the flexibility and accessibility offered by web-based courses, a perspective consistently supported in previous research [25,26,29,30]. Students appreciated the ability to learn at their own pace, access course materials conveniently, and engage with multimedia content tailored to their needs. The use of diverse teaching media, such as audio, video, and interactive exercises, aligns with widely recognized recommendations for enhancing learning engagement and retention [59,60]. This highlights the role of self-directed learning in increasing engagement and satisfaction.

Unlike previous studies focusing on blended learning or face-to-face instruction [25-27], this study uniquely explored the exclusive use of structured web-based courses for EMP. The findings underscore the potential of this approach to address the challenges faced by medical students in non–English-speaking regions, thereby filling a critical gap in the literature.

This study also found that students reported greater confidence in receptive skills (reading and listening) than in productive skills (writing and speaking), a pattern consistent with findings from previous research [7,15,17]. This underscores the need for targeted instructional strategies to support the development of productive language skills in EMP courses [2,17].

By situating the findings within the broader body of research, this study contributes to the evolving understanding of effective teaching strategies for EMP and web-based learning in medical education.

Implications of Findings for Practice

The findings of this study have important implications for curriculum development and teaching strategies in medical schools, particularly in non–English-speaking contexts. Structured web-based EMP courses significantly improved students' medical English proficiency, demonstrating their potential to meet academic and professional needs effectively. These results suggest that similar approaches could be adopted by other medical schools to enhance student engagement and learning outcomes.

Medical schools can leverage the flexibility and accessibility of web-based learning to design EMP courses that accommodate students' demanding academic schedules. By focusing on targeted skill development and incorporating medical terminology into course content, institutions can create tailored programs that address specific language needs [14,19]. The self-paced nature of web-based learning further enables students to manage their time effectively, aligning with their individual schedules and learning preferences. This adaptability can increase motivation and reduce barriers to learning, particularly in resource-constrained settings.

Integrating EMP with other areas of medical education, such as reading medical literature and writing clinical reports, can help students perceive medical English as an essential part of their academic and professional journey rather than as a standalone requirement. Additionally, structured web-based courses offer scalable and accessible solutions for institutions with large student cohorts, ensuring consistent, high-quality content delivery while reducing the resource burden on faculty and support staff.

Despite the benefits of self-directed learning, low engagement with "consultant hours" highlights the need for integrating opportunities for active faculty-student interaction within course designs. Addressing this issue could involve developing mechanisms that encourage and normalize faculty interaction, which may be especially beneficial in contexts where cultural preferences for independence or heavy academic workloads limit voluntary engagement with support services [23,24,26]. Such measures could enhance overall student support and satisfaction.

These insights underscore the value of structured web-based EMP courses as a model for improving language proficiency and supporting broader academic goals in medical education.

Strengths and Limitations

This study offers a comprehensive evaluation of structured web-based EMP courses, combining subjective learner feedback with objective proficiency measures. A key strength lies in the design of the courses, informed by needs assessments and evidence-based guidelines [47,48]. Moreover, the inclusion of summative assessments provides robust evidence of the courses' effectiveness. This study had a large sample size and response rate, which ensures robust findings. However, the study's scope was limited to a single medical school, potentially affecting generalizability [11,57].

Future research should focus on strategies to enhance productive skills (writing and speaking) while continuing to explore the impact of cognitive biases, such as the Dunning-Kruger effect, on the gap between perceived and actual proficiency [56]. Large-scale, multi-institutional studies are warranted to validate these findings and provide broader recommendations for integrating EMP into medical curricula. Additionally, research should investigate the long-term impact and scalability of such courses across diverse educational settings, as well as the development of adaptive learning technologies to customize course content based on students' baseline proficiency levels, effectively addressing specific skill gaps.

XSL•FO
RenderX

Conclusions

Structured web-based EMP courses are highly relevant, widely accepted, and well-received by medical students, demonstrating significant improvements in their medical English proficiency, particularly in reading, vocabulary, and listening skills, as evidenced by both subjective feedback and objective measures. The flexibility, accessibility, and practicality of structured web-based learning make it an effective approach to address the unique challenges faced by medical students with demanding schedules. By tailoring course content to meet students' academic and professional needs and incorporating engaging instructional designs, these courses provide a scalable and sustainable solution for medical education in non–English-speaking regions. Future developments in EMP course design should focus on enhancing productive language skills, such as writing and speaking, while maintaining the balance between self-directed learning and faculty support, integrating these courses into medical curricula as an essential component to equip students with the language skills necessary for academic success and global medical practice.

Data Availability

The data from this research project are available upon reasonable request.

Conflicts of Interest

The English for Medical Purposes I and II online courses are copyrighted to Khon Kaen University. RM, IT, and KJ have patents for the English for Medical Purposes I and II online courses.

Multimedia Appendix 1
Development of structured web-based English for medical purpose courses.
[DOCX File, 41 KB - mededu_v11i1e65779_app1.docx]

References

1. Outemzabet B, Sarnou H. Exploring the significance of English-based communication for a community of medical academics in a public university teaching hospital in Algeria. Engl Specif Purp 2023 Apr;70:116-130. [doi: 10.1016/j.esp.2022.12.001]
2. Antic Z, Milosavljevic N. Some suggestions for modelling a contemporary medical English course design based on need analysis. Lingua 2016 Dec;184:69-78. [doi: 10.1016/j.lingua.2016.06.002]
3. Džuganová B. Medical language – a unique linguistic phenomenon. JAHR Eur J Bioeth 2019 Jul 29;10(1):129-145. [doi: 10.21860/j.10.1.7]
4. Gotti M, Salager-Meyer F. Teaching medical discourse in higher education: an introduction. Lang Learn High Educ ;6(1):1-13. [doi: 10.1515/cercles-2016-5001]
5. Baethge C. The languages of medicine. Dtsch Arztebl Int 2008 Jan;105(3):37-40. [doi: 10.3238/arztebl.2008.0037] [Medline: 19633751]
6. Tsai PH. A pedagogical dialogue between English for general purposes and English for medical purposes: marching from short story reading and art practice to the writing of a history of present illness. Taiwan J TESOL 2022;19(2):65-106. [doi: 10.30397/TJTESOL.202210_19(2).0003]
7. Lodhi MA, Shamim M, Robab M, Shahzad S, Ashraf A. English for doctors: an ESP approach to needs analysis and course design for medical students. Int J Engl Linguist 2018;8(5):205. [doi: 10.5539/ijel.v8n5p205]
8. Skelton J, Richards C. Communication for medicine: state-of-the-art. ESP Today 2021;9(1):9-29. [doi: 10.18485/esptoday.2021.9.1.1]
9. Alqurashi F. English for medical purposes for Saudi medical and health professionals. Adv Lang Lit Stud 2016;7(6):243-252. [doi: 10.7575/aiac.alls.v.7n.6p.243]
10. Maher J. English for medical purposes. Lang Teach 1986 Apr;19(2):112-145. [doi: 10.1017/S0261444800012003]
11. Liang H. A critical discourse analysis of medical English course syllabuses. J Lang Tech Res 2023;14(4):865-870. [doi: 10.17507/jltr.1404.02]
12. Antić Z. Teacher education in English for special purposes. Acta Fac Med Naiss 2016;33(3):211-215. [doi: 10.1515/afmnai-2016-0022]
13. Antic Z. Forward in teaching English for medical purposes. Med Biol 2007;14(3):141-147 [FREE Full text]
14. Khalid A. Needs assessment in ESP: a review. Stud Lit Lang 2016;12(6):38-46. [doi: 10.3968/8161]
15. Antić Z. Towards uniformity in English for medical purposes: evaluation and design. Srp Arh Celok Lek 2009;137(7-8):454-457. [Medline: 19764605]
16. Shi L, Corcos R, Storey A. Using student performance data to develop an English course for clinical training. Engl Specif Purp 2001 Jan;20(3):267-291. [doi: 10.1016/S0889-4906(00)00002-8]
17. Wette R, Hawken SJ. Measuring gains in an EMP course and the perspectives of language and medical educators as assessors. Engl Specif Purp 2016 Apr;42:38-49. [doi: 10.1016/j.esp.2015.11.002]

XSL•FO
RenderX

18. Hoekje BJ. Medical discourse and ESP courses for international medical graduates (IMGs). Engl Specif Purp 2007 Jan;26(3):327-343. [doi: 10.1016/j.esp.2006.09.002]
19. Willey I, Tanimoto K, McCrohan G, Nishiya K. An English needs analysis of medical doctors in western Japan. JALT J 2020 Nov 1;42(2):143. [doi: 10.37546/JALTJJ42.2-3]
20. Cook DA, Garside S, Levinson AJ, Dupras DM, Montori VM. What do we mean by web-based learning? A systematic review of the variability of interventions. Med Educ 2010;44(8):765-774. [doi: 10.1111/j.1365-2923.2010.03723.x]
21. Rose S. Medical student education in the time of COVID-19. JAMA 2020 Jun 2;323(21):2131-2132. [doi: 10.1001/jama.2020.5227]
22. Thepwongsa I, Kirby CN, Schattner P, Piterman L. Online continuing medical education (CME) for GPs: does it work? A systematic review. Aust Fam Physician 2014;43(10):717-721. [Medline: 25286431]
23. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA 2008 Sep 10;300(10):1181-1196. [doi: 10.1001/jama.300.10.1181] [Medline: 18780847]
24. Pei L, Wu H. Does online learning work better than offline learning in undergraduate medical education? A systematic review and meta-analysis. Med Educ Online 2019 Dec;24(1):1666538. [doi: 10.1080/10872981.2019.1666538] [Medline: 31526248]
25. Mustafa N, Raside D. Online learning readiness: a case study in the field of English for medical purposes. Particip Educ Res 2016:212-220 [FREE Full text]
26. Navidinia H, Zare Bidaki M, Hekmati N. Incorporating e-learning in teaching English language to medical students: exploring its potential contributions. Med J Islam Repub Iran 2016;30(1):462. [Medline: 28491837]
27. Sriwichai C. Students' readiness and problems in learning English through blended learning environment. Asian J Educ Train 2020;6(1):23-34. [doi: 10.20448/journal.522.2020.61.23.34]
28. Ogundiya O, Rahman TJ, Valnarov-Boulter I, Young TM. Looking back on digital medical education over the last 25 years and looking to the future: narrative review. J Med Internet Res 2024 Dec 19;26(1):e60312. [doi: 10.2196/60312] [Medline: 39700490]
29. Thepwongsa I, Muthukumar R, Sripa P, et al. The perspectives of learners at a public medical school on the evaluation of an online learning management system for degree and non-degree courses. Med Educ Online 2024 Dec 31;29(1):2299535. [doi: 10.1080/10872981.2023.2299535]
30. Thepwongsa I, Sripa P, Muthukumar R, Jenwitheesuk K, Virasiri S, Nonjui P. The effects of a newly established online learning management system: the perspectives of Thai medical students in a public medical school. Heliyon 2021 Oct;7(10):e08182. [doi: 10.1016/j.heliyon.2021.e08182]
31. Thammaboosadee S, Yanta S. An analytical data monetization value chain for educational process improvement under Thai University Central Admission System. Presented at: 2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON); Dec 11-13, 2019; Bangkok, Thailand. [doi: 10.1109/TIMES-iCON47539.2019.9024499]
32. Sripa P, Thepwongsa I, Muthukumar R. Factors associated with the entry to general practice training: a multicentre study. Med Teach 2020 Dec 1;42(12):1394-1400. [doi: 10.1080/0142159X.2020.1811846]
33. Mei A, Gao D, Jiang J, Qiao T, Wang F, Li D. The medical education systems in China and Thailand: a comparative study. Health Sci Rep 2022;5(6):e826. [doi: 10.1002/hsr2.826]
34. Jarunthawatchai W, Baker W. English language education and educational policy in Thailand. In: Moody AJ, editor. The Oxford Handbook of Southeast Asian Englishes: Oxford University Press; 2024:557-574. [doi: 10.1093/oxfordhb/9780192855282.013.30]
35. Baker W, Jarunthawatchai W. English language policy in Thailand. Eur J Lang Policy 2017 Apr;9(1):27-44. [doi: 10.3828/ejlp.2017.3]
36. Teng B, Sinwongsuwat K. Teaching and learning English in Thailand and the integration of conversation analysis (CA) into the classroom. Engl Lang Teach 2015;8(3). [doi: 10.5539/elt.v8n3p13]
37. Noom-ura S. English-teaching problems in Thailand and Thai teachers' professional development needs. Engl Lang Teach 2013;6(11):139. [doi: 10.5539/elt.v6n11p139]
38. Student admissions. Khon Kaen University. URL: https://apps.admissions.kku.ac.th/web/Port/PortProject/7 [accessed 2023-05-14]
39. Chulee W, Khanom K, Chesa N, Sha'ar M, Buddharat C. "Why still we cannot speak English?" Examining internal demotivating factors among Thai tertiary learners. MEXTESOL J 2023 Nov 1;47(4):1-17. [doi: 10.61871/mj.v47n4-5]
40. Na Nongkhai A. An investigation into English language motivation of Thai university students: understanding students' motivation over time, and their visions of future L2 selves, through narrative inquiry [PhD thesis]. : University of Sheffield; 2017.
41. Chan SMH, Mamat NH, Nadarajah VD. Mind your language: the importance of English language skills in an international medical programme (IMP). BMC Med Educ 2022 May 26;22(1):405. [doi: 10.1186/s12909-022-03481-w] [Medline: 35619080]

XSL•FO
RenderX

42. Suwanrot K, Sausukpaiboon K, Ketdao N, et al. Learning needs assessment for English language of medical students and residents in Srinagarind Hospital, Faculty of Medicine, Khon Kaen University [Article in Thai]. Srinagarind Med J 2017;32(5):454-460 [FREE Full text]
43. Wahyuni S. English language needs for medical students: a link and match of academic and professional career. ENGLISH FRANCA Acad J Engl Lang Educ 2021;5(1):169-184. [doi: 10.29240/ef.v5i1.2146]
44. Kaliyadan F, Thalamkandathil N, Parupalli SR, Amin TT, Balaha MH, Al Bu Ali WH. English language proficiency and academic performance: a study of a medical preparatory year program in Saudi Arabia. Avicenna J Med 2015;5(4):140-144. [doi: 10.4103/2231-0770.165126] [Medline: 26629471]
45. Sadeghi B, Kashanian NM, Maleki A, Haghdoost A. English language proficiency as a predictor of academic achievement among medical students in Iran. Theory Pract Lang Stud 2013;3(12):2315-2321. [doi: 10.4304/tpls.3.12.2315-2321]
46. Bączek M, Zagańczyk-Bączek M, Szpringer M, Jaroszyński A, Wożakowska-Kapłon B. Students' perception of online learning during the COVID-19 pandemic: a survey study of Polish medical students. Research Square. Preprint posted online on Jul 14, 2020. [doi: 10.21203/rs.3.rs-41178/v1]
47. Hays R, Veitch C. Continuing medical education and divisions. In: Hays R, Veitch C, editors. Continuing Medical Education for General Practitioners: James Cook University; 1999:5-17.
48. Cook DA, Dupras DM. A practical guide to developing effective web-based learning. J Gen Intern Med 2004 Jun;19(6):698-707. [doi: 10.1111/j.1525-1497.2004.30029.x] [Medline: 15209610]
49. Thepwongsa I. Education of rural and remote general practitioners (GPs) in Australia on type 2 diabetes: impact of online continuing medical education on GPs' knowledge, attitudes and practices and barriers to online learning [Thesis]. : Monash University; 2017 Feb 23. [doi: 10.4225/03/58ae4580ba641]
50. Abbasi S, Ayoob T, Malik A, Memon SI. Perceptions of students regarding e-learning during Covid-19 at a private medical college. Pak J Med Sci 2020 May;36(COVID19-S4):S57-S61. [doi: 10.12669/pjms.36.COVID19-S4.2766] [Medline: 32582315]
51. Lasanthika W, Tennakoon W. Assessing the adoption of learning management systems in higher education. Global J Bus Soc Sci Review 2019 Sep 27;7(3):204-209. [doi: 10.35609/gjbssr.2019.7.3(5)]
52. Dillman DA, Smyth JD, Christian LM. Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method, 4th edition: Wiley; 2014. [doi: 10.1002/9781394260645]
53. Fan W, Yan Z. Factors affecting response rates of the web survey: a systematic review. Comput Human Behav 2010 Mar;26(2):132-139. [doi: 10.1016/j.chb.2009.10.015]
54. Porter SR, Whitcomb ME. The impact of contact type on web survey response rates. Public Opin Q 2003;67(4):579-588. [doi: 10.1086/378964]
55. Dunning D. The Dunning–Kruger effect: on being ignorant of one's own ignorance. In: Olson JM, Zanna MP, editors. Advances in Experimental Social Psychology: Academic Press; 2011, Vol. 44:247-296. [doi: 10.1016/B978-0-12-385522-0.00005-6]
56. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. J Pers Soc Psychol 1999;77(6):1121-1134. [doi: 10.1037/0022-3514.77.6.1121]
57. Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. Acad Med 2005 Oct;80(10 Suppl):S46-S54. [doi: 10.1097/00001888-200510001-00015] [Medline: 16199457]
58. Norman G. Research in clinical reasoning: past history and current trends. Med Educ 2005 Apr;39(4):418-427. [doi: 10.1111/j.1365-2929.2005.02127.x] [Medline: 15813765]
59. Arani JA. Mobile educational SMSs as supplementary means to teach sentence paraphrasing in EMP course. Int J Interact Mob Technol 2016;10(1):45-51. [doi: 10.3991/ijim.v10i1.5188]
60. Mayer RE. Evidence-based principles for how to design effective instructional videos. J Appl Res Mem Cogn 2021 Jun 1;10(2):229-240. [doi: 10.1016/j.jarmac.2021.03.007]

Abbreviations

EMP: English for medical purposes
KKUMET: Khon Kaen University Medical English Test
LMS: learning management system
TCAS: Thailand University Central Admission System

XSL•FO
RenderX

Edited by B Lesselroth; submitted 26.08.24; peer-reviewed by N Riapina, Q Xie; revised version received 20.01.25; accepted 25.02.25; published 27.03.25.

Please cite as:
Muthukumar R, Thepwongsa I, Sripa P, Jindawong B, Jenwitheesuk K, Virasiri S
Preclinical Medical Students' Perspectives and Experiences With Structured Web-Based English for Medical Purposes Courses: Cross-Sectional Study
JMIR Med Educ 2025;11:e65779
URL: https://mededu.jmir.org/2025/1/e65779
doi:10.2196/65779

© Radhakrishnan Muthukumar, Isaraporn Thepwongsa, Poompong Sripa, Bangonsri Jindawong, Kamonwan Jenwitheesuk, Surapol Virasiri. Originally published in JMIR Medical Education (https://mededu.jmir.org), 27.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Factors Influencing Educators' Perspectives on Accepting Extended Reality in Health Care Education: Qualitative Study

Zuheir Khlaif[1], PhD; Nisreen Salama[2], PhD; Bilal Hamamra[1], PhD; Allam Mousa[3], PhD

[1]Faculty of Humanities and Educational Sciences, An-Najah National University, Nablus, Occupied Palestinian Territory

[2]Faculty of Nursing, Arab American University, Jenin, Occupied Palestinian Territory

[3]Faculty of Engineering, An Najah National University, Nablus, Occupied Palestinian Territory

Corresponding Author:
Zuheir Khlaif, PhD
Faculty of Humanities and Educational Sciences
An-Najah National University
Old Campus Street
Nablus, PS358
Occupied Palestinian Territory
Phone: 970 0592754908
Email: zkhlaif@najah.edu

Abstract

Background: Palestinian higher education institutions face limitations in providing interactive practical training for medical education. Extended reality (XR), which encompasses virtual reality and augmented reality, is increasingly recognized for addressing these challenges by offering immersive learning experiences.

Objective: This study investigates the factors influencing the acceptance and adoption of XR in health care education within Palestinian universities, exploring its potential to transform traditional teaching methods.

Methods: A qualitative approach was used in this study to collect data through semistructured interviews and artifacts from the participants. The participants of the study were 25 faculty members from 2 large Palestinian universities who teach in the field of medical sciences.

Results: Three primary categories—external, internal, and design-related factors—emerged as pivotal in influencing XR adoption. Professional development, technical support, and infrastructure were key external enablers. Internally, prior experience with digital tools and positive attitudes had a significant impact on the adoption of XR. Design factors, including ease of use and interactivity, played a crucial role but also posed challenges for less tech-savvy educators. Despite barriers such as cost and technical issues, XR demonstrated notable benefits, including enhanced learning outcomes, improved knowledge retention, and the ability to simulate complex medical scenarios.

Conclusions: XR technologies offer transformative potential for health care education in Palestine. By addressing challenges and leveraging XR's strengths, educational institutions can foster innovation and improve student engagement and skill acquisition. The study contributes to the theoretical understanding of technology acceptance in education by identifying the interplay of external, internal, and design factors. Practically, it emphasizes strategic investments in infrastructure, professional training, and institutional policies to optimize XR integration.

(JMIR Med Educ 2025;11:e65042) doi:10.2196/65042

KEYWORDS

extended reality (XR); health care education; educational technology; Sustainable Development Goals (SDGs); Palestine

Introduction

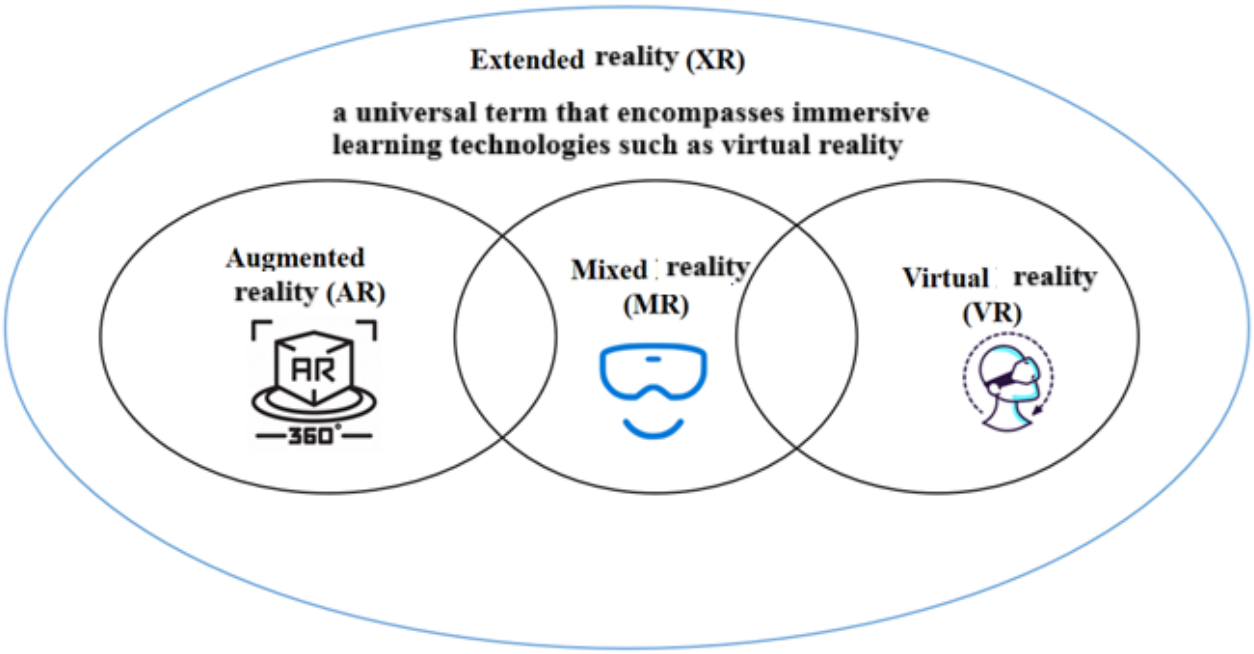
Background

The integration of emerging technologies such as artificial intelligence and XR has significantly enhanced learners' engagement and interaction with educational environments [1]. XR, an umbrella term encompassing virtual reality (VR), augmented reality (AR), and mixed reality (MR), facilitates human-machine interaction through computer-generated content [2]. VR provides a fully immersive digital environment, AR overlays virtual objects onto real-world settings, and MR blends physical and digital elements into a seamless interactive

XSL•FO
RenderX

experience [2]. By incorporating visual, auditory, and interactive elements, XR enhances real-world learning by immersing students in digital environments that facilitate deeper engagement with educational content [3]. As shown in Figure 1, XR encompasses AR, MR, and VR, creating a spectrum of immersive experiences. Figure 1, developed by the authors, shows how XR encompasses AR, MR, and VR.

Figure 1. Illustration showing how extended reality (XR) encompasses augmented reality (AR), mixed reality (MR), and virtual reality (VR), as developed by the authors.



XR technologies are transforming medical education, particularly in resource-constrained settings, by providing cost-effective, risk-free learning environments. Head-mounted devices enable students to practice medical skills repeatedly without compromising patient safety. A review of 27 studies found XR-based training to be highly effective in surgery and anatomy, although more large-scale research is needed to fully assess its impact [4]. While some studies, such as Behmadi et al [5], found no statistically significant improvements in knowledge retention when comparing XR-based learning with traditional methods, they reported enhanced student engagement and satisfaction. This suggests that XR should be integrated alongside conventional teaching methods to cater to diverse learning styles and develop critical decision-making skills. Additionally, XR reduces dependence on costly medical equipment and provides students with virtual laboratories to conduct experiments, mitigating financial and logistical barriers [3,6]. By enabling hands-on practice in safe, controlled environments, XR improves learning accessibility and instructional quality [6].

In Palestinian higher education, XR adoption remains in its early stages, particularly in health care disciplines such as medicine and nursing. Its integration is closely tied to broader digital transformation policies, which aim to modernize educational practices and align with global trends in technology-enhanced learning. However, the transition faces barriers such as limited resources, inadequate infrastructure, and the need for faculty capacity-building. Despite these challenges, Palestinian universities are increasingly prioritizing XR within their strategic frameworks, reflecting a commitment to leveraging emerging technologies to enhance learning outcomes.

This study contributes to the growing research on XR adoption by examining Palestinian higher education institutions, a unique and underrepresented context. Unlike studies from technologically advanced regions, where XR adoption is often supported by substantial funding and infrastructure, this research highlights the policy-related, infrastructural, and faculty-specific challenges of implementing XR in resource-limited settings. While previous studies in Latin America and Southeast Asia have explored XR's role in medical education, they have often overlooked institutional policies, faculty readiness, and regional constraints. By addressing these factors, this study provides a comparative perspective on XR adoption in Palestinian universities and offers global insights into key determinants for sustainable technology integration in medical education.

Literature Review

The Unified Theory of Acceptance and Use of Technology Model in XR Adoption for Health Care Education

Overview

VR and AR are highly adaptable technologies that use various systems, setups, and content types, ranging from immersive and dynamic to nonimmersive and static environments. These technologies are characterized by 3 key elements: immersion, presence, and interaction [7,8]. Immersion depends on the

XSL•FO
RenderX

technological medium, such as head-mounted displays, concave or 3D projections, or interactive videos where users engage as protagonists. Presence and interaction, by contrast, relate to an individual's perception of connectedness within the virtual environment and their ability to act upon it and receive feedback [7,8]. These elements are crucial in defining the effectiveness and adoption of XR technologies, particularly in educational contexts such as health care training.

The Unified Theory of Acceptance and Use of Technology (UTAUT), developed by Venkatesh et al [9], serves as the theoretical framework for examining the adoption of XR in health care education. Over time, various theories and models have been proposed to understand the factors influencing the acceptance of new technologies. One of the foundational models in this field is the Technology Acceptance Model (TAM), developed by Davis [10] and based on the Theory of Reasoned Action. The TAM has been widely used to investigate the adoption of emerging technologies, including XR in surgical training, artificial intelligence adoption, and mobile learning [11-13]. However, the TAM has been critiqued for its limited predictive accuracy, as it fails to account for technology acceptance in nearly 40% of cases [14].

To address these limitations, Venkatesh et al [9] developed the UTAUT, which integrates multiple previous models, including the Theory of Reasoned Action, to provide a more comprehensive framework for understanding technology adoption. The UTAUT identifies 4 key constructs that shape an individual's intention to use, as well as their actual usage behavior of a new technology: Performance Expectancy, Effort Expectancy, Social Influence, and Facilitating Conditions.

Performance Expectancy

In the context of XR adoption in health care education, performance expectancy refers to the extent to which faculty members perceive XR as beneficial for both educators and students. XR facilitates immersive learning, enhances knowledge retention, and allows for realistic simulations of medical procedures without risks to patients. Faculty members recognize its potential to improve teaching effectiveness, help achieve learning objectives, and develop students' practical skills in a safe environment.

Effort Expectancy

Effort expectancy concerns the perceived ease of use of XR technology in teaching and learning environments. Faculty members' willingness to integrate XR depends on how intuitive and user-friendly the technology is. Instructors with prior experience using digital tools often find XR more accessible, while others may require training and technical support to overcome usability challenges.

Social Influence

The role of social influence in XR adoption is significant, as faculty members' decisions are shaped by the expectations of colleagues, mentors, and students. In this study, social influence extends beyond professional networks to include university policies, peer recommendations, and online technology communities that encourage XR integration.

Facilitating Conditions

The successful adoption of XR technology in health care education depends heavily on facilitating conditions, including institutional support, infrastructure, and technical assistance. Universities must provide adequate resources, professional development programs, and robust digital infrastructure to support faculty members' continued use of XR. Without sufficient technical support and access to XR-compatible hardware, faculty adoption may be hindered.

Virtual Reality in Health Care

VR technology has revolutionized health care education by offering immersive, interactive learning experiences. It creates realistic simulations that allow students to practice procedures and decision-making without risks [15]. For instance, the Stanford Virtual Heart Project uses VR to help students understand cardiac anatomy through 3D models. Furthermore, VR allows for detailed exploration of the human body, aiding in the comprehension of anatomical structures. The HoloAnatomy program at Case Western Reserve University uses Microsoft HoloLens for holographic dissections, enhancing anatomy learning and critical thinking [15].

VR provides simulation-based training for medical procedures, enhancing skills and confidence before actual patient care. Apps such as Surgical Theater aid in understanding and practicing surgical procedures [16]. VR also enables collaborative learning through merged MR and mixed XR platforms, enhancing teamwork and communication skills by combining the virtual and physical worlds [15]. VR makes learning more engaging and accessible by bringing medical simulations to remote or underresourced areas, ensuring that high-quality medical education is available to a broader audience [17].

XR Adoption in Education

Researchers have reported that XR is the future of learning and teaching in higher education institutions worldwide [6]. XR promotes teamwork through collaboration in VR environments [6]. Moreover, using XR in education transforms the learning process from traditional methods to active learning, where learners engage directly in learning activities [18]. In addition, XR enhances personalized learning by creating a customized learning environment tailored to the learner's needs and abilities [19]. Educators can design digital content to simulate immersive and interactive environments, making XR particularly suitable for teaching high-risk procedures, using expensive equipment, and conducting practical training activities that require additional resources, such as those in medicine or medical sciences [20].

Other studies have confirmed that using XR in learning fosters creativity, innovation, and design among learners [21,22]. Aguayo and Eames [23] reported that virtual agents in XR facilitated language learning and teaching by making complex knowledge more accessible. Moreover, these technologies offer immersive and interactive simulations that replicate complex real-world scenarios [24]. Consequently, using XR in education enhances the real-world experience by incorporating sounds, videos, and graphics into the learning environment, enabling learners to interact more effectively [3].

XSL•FO
RenderX

Benefits and Challenges

VR and AR offer numerous benefits over conventional therapy, including cost reduction, fewer hospital visits for immobile patients, user-friendly experiences, and improved patient safety. They also facilitate data collection in research settings and reduce surgical errors in training [25,26]. However, many studies indicate that determining the benefits of VR and AR in health care is challenging due to their small sample sizes, heterogeneous nature, and lack of proper controls. Pain relief mechanisms remain debated, and implementation is technically complex and expensive, with lower acceptance among older adults [7,27].

One of the main advantages of using XR is its ability to provide practical training without the need for expensive physical equipment [28]. For example, medical students can practice surgical procedures in a virtual environment, reducing the need for costly medical equipment and minimizing the risks associated with real-life practice on patients [29]. The implementation of virtual laboratories through XR not only makes education more accessible but also enhances the learning experience by offering hands-on practice in a safe and controlled environment [6].

Factors That Influence the Continued Intention to Use XR in Health Care Sciences

A previous study explored the potential of VR and AR technologies, particularly XR, in enhancing patient care, medical education, and presurgical planning [30]. It highlights the potential of integrating XR into medical education to provide immersive, interactive experiences.

Burian et al [31] highlighted that XR technologies provide more effective training compared with traditional methods, particularly for novices. Chuah [32] conducted a study involving 45 relevant studies to assess and analyze user acceptance of XR technology from multiple theories, disciplines, and perspectives. Wearable XR technology is influenced by various factors, including cost, technical and performance issues, hardware size, sensory inputs, content quality, cognitive impacts, user satisfaction, attitudes, intentions to visit tourism sites, expectation confirmation, personality traits, knowledge transfer, support needs, presence, boundary considerations, consumer characteristics, spatial awareness, control, participation, effectiveness, familiarity, innovativeness, value perceptions, decision comfort, spatial understanding, cognitive load, virtual embodiment perception, sense of presence, health and privacy concerns, and psychological and physical risks. Curran et al [33] stated that XR technologies offer portability, standardization, replicability, accessibility, and the ability to function without heavy manikin parts. They can be widely distributed without the need for a live instructor, thereby increasing learner engagement and enhancing spatial representation.

Kluge et al [34] noted that despite limited experience with XR technology, staff and students at the University of Newcastle view it as a standard tool for teaching. They aim to develop a sustainable implementation framework within 5 years. VR, AR, and MR technologies are disrupting medical education by offering immersive experiences and alleviating traditional learning constraints. XR technologies, particularly in emergency medicine, enable remote clinical skill development, even amidst the challenges posed by COVID-19 [33]. Li and Keskitalo [35] emphasized that XR technology is commonly used in health care education for safe treatments, communication, and decision-making. It supports 5 cognitive-processing dimensions: remembering, understanding, applying, analyzing, and evaluating. AR, VR, and MR can positively impact medical education. To effectively implement XR, it is important to consider existing resources and Bloom's taxonomy, and select the most suitable technology. Optimizing and expanding XR utilization are crucial for promoting deeper learning in health care. There is a significant gap in research regarding the factors influencing the continued use of XR in medical science, emphasizing the need for further studies in this area.

Methods

Study Design

We used a qualitative approach to explore faculty members' lived experiences with using XR in medical education at 2 Palestinian universities. This approach allows researchers to gain insights into the phenomena from practitioners who are using XR in teaching [36]. We gave participants the opportunity to share their experiences with using XR in their courses. The 2 universities involved are The Arab American University-Palestine and An-Najah National University. Both universities have established digital transformation centers focused on immersive technologies to enhance teaching and learning. The Arab American University-Palestine has developed a VR laboratory within its nursing and medicine departments, using cutting-edge technology to improve educational experiences. Similarly, An-Najah National University has created VR laboratories for its Departments of Dental Medicine and Medicine, along with a general XR center that serves both the university and the surrounding community. These laboratories provide training in XR technologies for faculty and students, supporting the integration of immersive tools into the curriculum.

Participants

This study involved 25 faculty members (18 males and 7 females) from diverse disciplines within medical sciences, representing 2 universities in the West Bank of Palestine. All participants had a minimum of 1 year of experience using XR in their teaching of undergraduate medical sciences courses, ensuring their expertise and relevance to the study's focus. Participants contributed by engaging in semistructured interviews and providing examples of their own work and that of their students. Table 1 presents the demographic characteristics of the participants, highlighting their diversity and alignment with the study's objectives.

XSL•FO
RenderX

Table 1. Demographic information about the participants in the study.

Variable and category	Frequency, n
Gender	
Male	18
Female	7
Education level	
Doctorate	19
Master's degree	6
Teaching experience	
5 years or less	5
6-10 years	12
11 years or more	8
Medical sciences field	
Nursing	7
Human medicine	8
Pharmacy	4
Dentistry	6

Recruitment of Participants and Justification of Sample Size

This study used purposive sampling, a qualitative method that selects participants based on predefined criteria relevant to the research objectives [37]. Faculty members, aligned with the study's focus on XR adoption in medical education, were recruited through official invitations sent by the deans of the 2 Palestinian universities.

Eligible participants were required to meet several inclusion criteria: at least one year of experience using XR in teaching, an academic position within the Faculty of Medical Sciences, and a minimum of 3 years of higher education teaching experience. Their expertise was further validated through prior research or publications on XR, active involvement in curriculum development, and proficiency with XR tools, demonstrated through certifications, training, or workshops. Additionally, student evaluations and feedback on XR-based teaching were considered.

To enhance transferability, the study included faculty from the fields of nursing, human medicine, pharmacy, and dentistry, ensuring representation across disciplines with varying levels of reliance on XR-based learning. Participants ranged from early-career faculty with 5 or fewer years of teaching experience to senior faculty with over 10 years, providing a range of professional perspectives. The sample also encompassed both experienced and novice XR users, drawn from universities at different stages of XR adoption—one with an established XR laboratory and structured training programs, and the other in an early adoption phase—offering a comprehensive view of institutional challenges and implementation strategies.

The sample size of 25 faculty members was determined based on data saturation, a key principle in qualitative research. Saturation is reached when further data collection no longer generates new themes or insights [38]. In this study, saturation was achieved by the 22nd interview, as recurring patterns related to institutional barriers, faculty perceptions, and XR adoption challenges emerged. The remaining interviews were conducted to reinforce the robustness of these themes. Prior research [9,36] on technology adoption in higher education suggests that qualitative studies with 15-30 participants provide sufficient depth to capture rich, context-specific insights. Given the exploratory nature of this study, the selected sample size was deemed appropriate for gathering in-depth faculty perspectives on XR integration in medical education.

We acknowledge that the small sample size may limit the generalizability of the findings. However, this study was designed to explore educators' lived experiences and provide deep insights into XR adoption in medical education, rather than aiming for broad generalizability. The qualitative methodology, supported by data triangulation through interviews and artifacts, enhances the trustworthiness and validity of the findings despite the limited sample size. To mitigate this limitation, participants were purposively selected from diverse medical disciplines and institutions to ensure a variety of perspectives. Rigorous thematic analysis techniques were used, validated by multiple coders with high interrater reliability. Additionally, the findings were cross-referenced with existing literature to contextualize and strengthen the conclusions.

While future studies with larger samples are necessary to further validate these findings, this study provides valuable foundational insights into XR adoption in medical education. These findings can serve as a reference for similar educational contexts, aiding institutions and educators in navigating the complexities of XR implementation.

XSL•FO
RenderX

Study Context

The context of this study involved faculty members from the Faculty of Healthcare Sciences at 2 major universities in the West Bank of Palestine. Both universities have a clear vision and policy to integrate emerging technologies, such as XR, into medical education and research. To facilitate the adoption of XR in teaching, 4 training workshops were organized at each university to equip faculty members with the knowledge and skills necessary to understand and apply XR in their practices. The length of each training session varied depending on the topic, generally ranging from 2 to 7 days. These sessions provided teachers with the opportunity to create and develop learning objects using XR. Various topics were covered, including an introduction to VR, AR, and XR; designing lessons using objects on the platform; and creating avatars for lessons, among others.

Research Instrument

A semistructured interview was the primary research instrument for data collection. We developed an interview protocol to guide the process. The protocol (Multimedia Appendix 1) consisted of 2 sections. The first section introduced the study to the participants and confirmed the confidentiality of their responses. Participants were informed that the interview would be recorded, provided they agreed, and that they could withdraw at any time. They were also asked to sign a consent form before the interview was recorded. The second section contained interview questions developed from a literature review, aligned with the research questions. Another data source consisted of artifacts provided by the participants during the interviews or submitted via email, illustrating how XR was used in medical education.

Data Collection

The researchers sent invitations to the nominated participants to schedule semistructured interviews at their convenience. Participants were given the flexibility to choose the time and location of the interviews. Each interview, lasting 30-45 minutes, was conducted individually and recorded. The interview protocol began by asking participants to discuss their general experiences with technology, followed by specific inquiries about their experiences with XR, allowing them to share their stories and journeys. Follow-up questions were asked to delve deeper into their experiences, especially regarding student collaboration on projects. For example, when one participant mentioned that students learn more through exploration, a follow-up question was, “Can you provide more details about what exploring entails and the role of XR in facilitating that exploration?” The interviews were designed to capture as much detailed information as possible from the participants’ experiences. In addition to the interviews, participants provided various artifacts, including samples of student work and activities implemented using XR. These artifacts served as valuable secondary data sources.

Trustworthiness

To ensure trustworthiness and methodological rigor, this study followed established qualitative research principles, emphasizing credibility, confirmability, dependability, and transferability. Credibility was enhanced through the use of multiple data collection methods, including semistructured interviews, participant-submitted artifacts, and institutional policy documents. Member checking was conducted by sharing transcribed interviews and preliminary themes with participants, enabling them to verify the accuracy of interpretations and provide clarifications where necessary. Additionally, peer debriefing was used, where external researchers reviewed the coding process and emerging themes to refine definitions, minimize bias, and ensure analytical coherence.

Conformability was maintained by carefully documenting the research process and prioritizing participants’ statements to minimize researcher bias. The interview protocol was developed based on the research questions, a pilot study, and expert reviews, ensuring its relevance and alignment with the study’s objectives. Dependability was reinforced through rigorous coding procedures, with data being independently coded 3 times, achieving 92% interrater reliability. This process ensured consistency in theme identification and analysis. Transferability was supported by purposive sampling, which selected participants from diverse academic disciplines and career stages to capture a broad range of perspectives on XR adoption. Additionally, data triangulation—incorporating interviews, observational notes, and artifacts from faculty members—provided a comprehensive understanding of the factors influencing XR integration in medical education. By incorporating these robust qualitative validation strategies, the study strengthens its credibility, validity, and applicability, ensuring that the findings accurately reflect participants’ experiences while addressing potential biases in sample selection and data interpretation.

Ethical Considerations

This research was approved by the institutional review board committee at An-Najah National University under reference Med. April.2024/18. The study adhered to strict ethical guidelines to ensure participant confidentiality, informed consent, and data protection. Before participation, all participants were provided with detailed information about the study’s objectives, procedures, and their rights as participants. Participants were explicitly informed that their participation was entirely voluntary, and they could withdraw from the study at any time without facing any consequences or needing to provide an explanation. To document consent, participants signed an informed consent form that outlined the scope of the study, the types of data to be collected, and how the data would be used solely for research purposes. The consent form also detailed the measures taken to protect their identities and ensure the confidentiality of their responses.

In terms of data protection, stringent protocols were followed to safeguard participant information. Personal data, including contact details, were securely stored on a password-protected and encrypted computer. Both physical and digital access to these data was restricted to authorized researchers only. Furthermore, any identifying information was anonymized during the data analysis process to further protect participants’ privacy. Interview data were stored in encrypted files, and backup copies were securely maintained to prevent data loss.

XSL•FO
RenderX

Additionally, the researchers communicated the steps taken to comply with ethical research standards, including adherence to the principles outlined by Ngozwana [39]. These procedures ensured that participants felt confident their contributions were protected and respected throughout the research process. By incorporating these comprehensive protocols, the study's transparency and ethical rigor were enhanced.

Data Analysis

To address the research questions, inductive thematic analysis was used, following the 6-step methodology proposed by Braun and Clarke [38]. The research process involved conducting 10.5 hours of recorded interviews. Before data analysis, the researchers transcribed the audio recordings and shared the text files with participants for validation. Participants were given the opportunity to amend, rewrite, or supplement the content as needed. Once the files were returned, no further modifications were expected. The thematic analysis was carried out in 6 phases, as outlined by Braun and Clarke [38], and these phases guided the data analysis process in this study.

The researchers began their analysis by organizing the individual transcript files, labeling them as F1, F2, and so on up to F25. The initial phase involved familiarizing themselves with the data. The researchers, who had conducted the interviews, carefully read through each transcript while simultaneously listening to the corresponding audio files. This process of immersion allowed the researchers to take detailed notes in the margins of the transcripts, ensuring a thorough understanding of the data. In the second phase, coding, the researchers developed labels for significant features of the data that were relevant to the research questions. As they meticulously read through the transcripts line-by-line, they created a coding book to document these labels. This phase focused on identifying key concepts or ideas that were central to the research, such as how faculty members experience XR in teaching medical concepts and in transferring medical knowledge and skills to their students. Next, the researchers moved to the theme development phase, where they identified patterns related to the research questions. They grouped similar and contrasting codes into coherent themes, resulting in a structured coding book (Table 2 and Multimedia Appendix 2). In the fourth phase, reviewing themes, the researchers examined these themes across all data, using inductive analysis to refine and combine them as needed. This iterative process of theme development was essential for generating meaningful insights from the data.

The fifth phase involved defining and naming themes based on the research questions, literature review, and theoretical framework. This phase integrated both inductive and deductive approaches, which are common in qualitative data analysis. Finally, in the reporting phase, the researchers documented and presented the analysis results, supporting the findings with direct quotations from the participants' interviews. This approach helped provide concrete evidence and context for the identified themes.

The data analysis process began with a thorough review of each transcript to familiarize the researchers with the material. During this phase, ideas and concepts were identified as units of analysis, and a coding process was developed. A coding book was created to assist with data coding. Ideas and concepts sharing similar characteristics were organized into emerging subthemes, which were then grouped under primary themes, informed by prior research findings. The research questions guided the analysis process. Table 2 illustrates the inductive thematic analysis approach used to develop the coding book for each research question.

XSL•FO
RenderX

Table 2. An overview of the themes and data sources (interviews and artifacts).

Major themes and sub-themes	Findings from interviews	Findings from artifacts
External Factors		
Professional Development	Faculty attended training sessions on XR[a], viewed as beneficial for skill-building and knowledge sharing.	Training materials and lesson plans demonstrated faculty engagement with XR-based teaching.
Technical Support	Faculty emphasized the need for technical support, linking it to continued XR use.	Technical support records showed assistance provided for troubleshooting XR-related issues.
Infrastructure	Challenges related to limited XR devices, weak Wi-Fi, and lack of educational resources.	Limited XR-enabled classrooms and laboratories were documented in institutional reports.
Social Influence	Positive influence from colleagues, social media, and university policies encouraged XR adoption.	Shared lesson plans and peer-reviewed materials indicated knowledge exchange among faculty.
XR Features		
Ease of Use	Some faculty found XR easy to use, while others struggled with the complexity of designing activities.	User guides and simplified XR tools were developed to address faculty concerns about complexity.
Interactivity	Interactivity was valued for facilitating learning objectives and engagement.	Lesson artifacts included interactive 3D models and gamified simulations.
Imagination and Immersion	Imagination linked to visualization capabilities, enhancing the learning experience.	Students' assignments showed creative applications of XR for visualization.
Internal Factors		
Previous Experience With ICTs[b]	Prior experience with ICTs contributed to smoother XR adoption.	Faculty-created resources mirrored existing ICT teaching practices, aiding XR integration.
Digital Competencies	Digital literacy skills were essential for the effective use of XR in teaching.	Assessment rubrics reflected the need for digital competencies in grading XR-based tasks.
Attitudes Toward XR	Most faculty had positive attitudes, seeing XR as engaging; a minority found it too complex.	Students' reflections indicated excitement about XR use; some reported difficulty in self-directed learning.
Design Factors		
Design Challenges	Time constraints and lack of technical expertise made designing XR activities difficult.	Course syllabi showed attempts to integrate XR but highlighted gaps in structured activity design.

[a]XR: extended reality.

[b]ICT: information and communication technology.

Results

Key Factors

Participants identified key factors based on their experiences with XR in health care education. Faculty workload and responsibility were recognized as significant factors influencing the integration of XR into teaching practices. Additionally, experience with medical technology was found to be linked to the use of XR. Data analysis revealed various influencing factors, which were grouped into 3 categories: internal factors, design factors, and external factors, which include XR features (Table 2). Each category encompasses several specific factors, as detailed in the following sections. Table 2 provides an overview of the themes and data sources.

External Factors

Institutional Drivers of XR Adoption in Teaching

External factors related to higher education institution policies and readiness, such as professional development, technical support, infrastructure, social influence, and XR features, were identified as factors that could positively influence the acceptance and continued use of XR in teaching practices.

Professional Development

All participants confirmed their attendance at training sessions on using XR in their teaching. These sessions covered a range of topics, from recognizing the value of XR as an advanced technology to creating lessons and activities using existing platform assets. For example, one faculty member mentioned, "The training was helpful in various aspects, such as understanding the value of XR and learning how to use it in my class activities" [D1]. Additionally, some faculty members viewed the training sessions as opportunities to share their knowledge and receive feedback, enhancing their practices in medical education. One participant noted, "I shared the activities I designed from scratch to get feedback from my colleagues. It was a good chance to share experiences and learn from others" [P5].

Integrating VR into medical education, particularly in fields such as nursing, offers students the chance to learn in authentic, immersive environments and practice practical tasks through simulations. However, some participants raised concerns that

XSL•FO
RenderX

VR might not substantially enhance student learning, as they believed students could already visualize real-life situations without the need for VR. This underscores the importance of professional development programs in equipping educators with the strategies needed to design VR experiences that extend beyond imagination, offering unique, hands-on, and interactive learning opportunities that traditional methods cannot replicate.

Technical Support

Most participants emphasized the importance of technical support in ensuring the continued use of XR, as it helps minimize technical difficulties faced by faculty members in medical sciences. Technical support encompassed a range of services, from creating platform accounts to troubleshooting issues with platform assets and student access. One faculty member stated, "Technical support is essential to continue using XR as it's a new technology, and I had no previous experience with it" [D25]. Some faculty members associated the availability of technical support with saving time and being able to focus more on the quality of activities and assessments. However, a few reported a lack of technical support due to insufficient staffing at the XR center. Providing technical support also positively impacted students' timely completion of assignments and tasks using XR.

Infrastructure

Infrastructure plays a crucial role in the use and continued adoption of XR in medical education. Participants defined infrastructure in terms of the availability of suitable VR devices, strong Wi-Fi, and educational resources. One faculty member mentioned, "I confront challenges to find assets related to Nursing relevant to my teaching topic" [D24]. A few participants cited the lack of infrastructure as a significant challenge. For instance, one faculty member said, "I have 45 students in Human Medicine, and it's difficult to take them all to the computer lab to use the VR devices because there is only one device" [D5].

One of the novel findings of this study is its emphasis on the intersection of institutional policies, faculty readiness, and infrastructural limitations in shaping XR adoption in medical education. While faculty members acknowledged the pedagogical benefits of XR, they also highlighted significant challenges related to institutional support and funding constraints. Unlike institutions in high-income regions, where government and private sector investments facilitate the widespread adoption of XR, Palestinian universities rely primarily on limited internal budgets and external grants. Consequently, the lack of funding for VR-compatible hardware, insufficient training opportunities for faculty, and inadequate technical support staff emerged as critical barriers to adoption. This study underscores the importance of targeted policy interventions, including faculty incentives, resource-sharing initiatives, and digital transformation strategies, to address these systemic barriers and promote sustainable XR integration.

Social Influence

Positive Influence of Colleagues

When asked about the impact of colleagues on their use of XR, most participants reported a positive influence from both within and outside the university. This influence included sharing expertise, providing technical and instructional support, and exchanging lessons and learning objects on the XR platform. One participant stated, "It was challenging to design lessons using the XR platform, so I asked a colleague for help" [D20].

The Power of Social Media

Some participants reported being members of social media groups focused on advanced technology in engineering, such as Twitter groups. These communities helped them exchange ideas about designing VR activities and share lessons using 3D and 360-degree techniques. One participant shared a lesson about the human body on Facebook and received feedback to improve the lesson using advanced VR features. Another participant said, "I share my lessons and activities in the group and exchange ideas on using XR in teaching various topics" [D6].

XR Features

Impact of XR Features on Adoption in Medical Education

The XR features reported by the majority of participants included ease of use, imagination, interaction, and immersion, all of which could influence the use of XR in medical education.

Ease of Use

Most participants highlighted the importance of XR's ease of use in lesson activities and content presentation. Some linked the simplicity of designing activities and lessons on the platform to their intention to continue using it. One faculty member stated, "I liked using XR because it was easy to design activities to show hidden parts of the human body" [D4]. However, a few participants found XR complicated and challenging for designing lesson activities, which led them to stop using it, although they continued assigning XR-related tasks to students.

Interaction Feature

Many participants emphasized the role of interactive features in facilitating and achieving lesson objectives. One participant said, "interactivity is important for me and my students because it enables activities that are otherwise impossible" [D11]. They also highlighted the importance of designing interaction types between students and learning activities, as well as interactions among students.

Imagination and Immersion

All participants confirmed the significance of imagination in medical sciences education, which could lead to greater immersion in class activities. One faculty member reported, "My students used XR to virtually perform a surgery" [D3]. Many participants linked imagination to visualization features that attract faculty members to use XR in assignments and activities.

Internal Factors

Role of Digital Competencies and Experience in XR Adoption

Internal factors included previous experience with information and communication technologies and digital competencies.

XSL•FO
RenderX

Previous Experience With Information and Communication Technologies

The majority of respondents indicated that previous experience with information and communication technologies and mobile technology was crucial for accepting and using XR in medical education. Experience with smartphones also facilitated their use of mobile VR for course instruction and activities.

Digital Competencies

In this study, digital competencies refer to the knowledge, skills, and attitudes related to XR. Most interviewees reported that their digital competencies were crucial for continuing to use XR. One faculty member stated, “My experience and knowledge are important for using XR in medical education.” [D9].

Design Factors

Design factors refer specifically to the pedagogical and instructional aspects of XR integration, particularly in creating activities, materials, and assessments in health care. This process often requires the application of instructional design principles. Most participants indicated that their primary difficulty was designing course-related activities. Many felt they lacked the technological expertise required, and some reported insufficient time due to commitments at private hospitals and clinics.

Attitudes Toward Using XR in Medical Education

Positive Attitudes

Many participants expressed positive attitudes toward using XR in medical education, attributing these attitudes to features such as interaction, visualization, and immersion. Some mentioned that the simplicity of XR saved both time and effort. For instance, one faculty member noted, “My students were excited to use XR.” [D4].

Negative Attitudes

A minority of participants, fewer than one-third, reported obstacles that negatively influenced their attitudes toward using XR. These challenges were related to the complexity of XR usage.

Socioeconomic Factors and XR Adoption

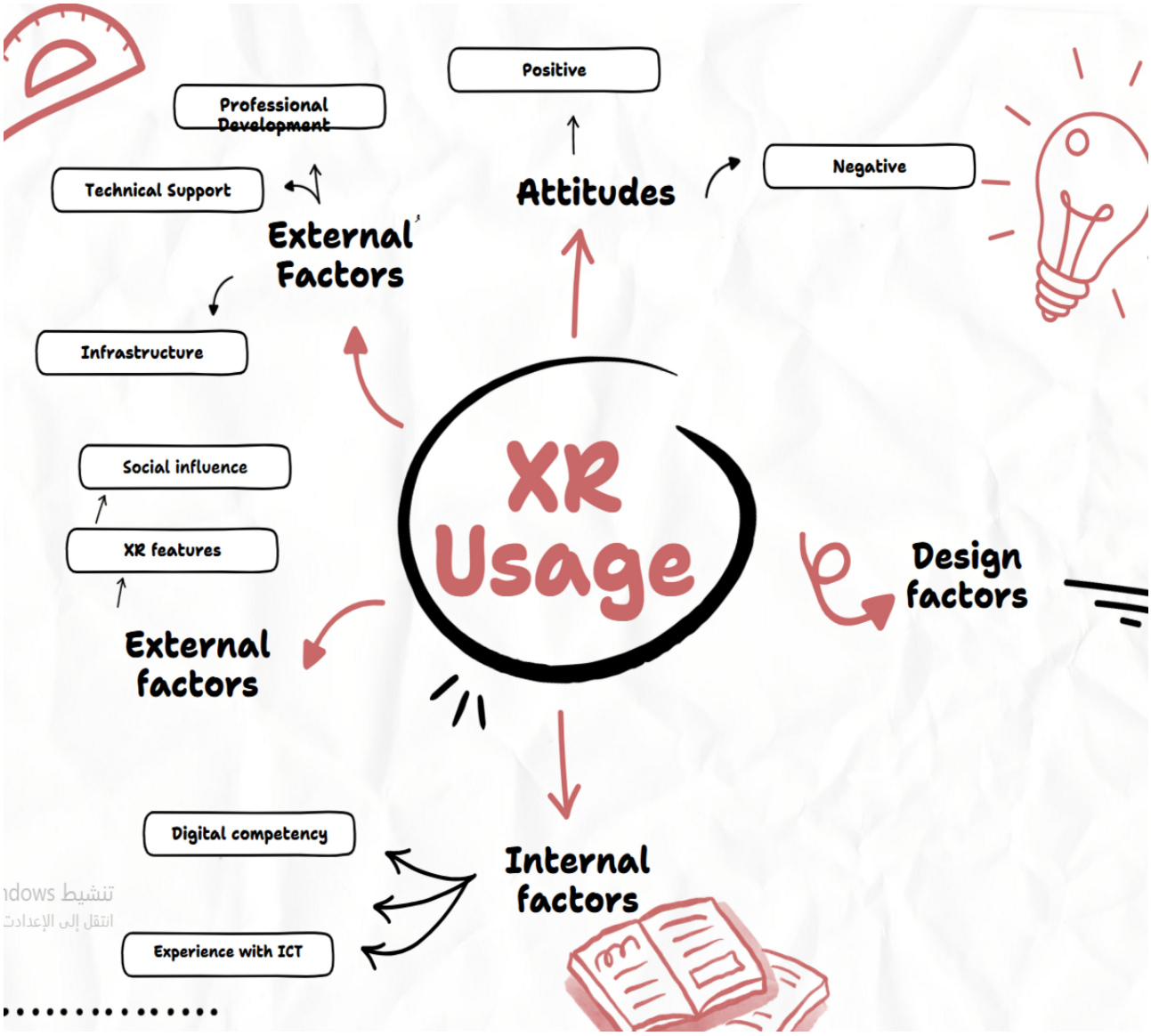
Socioeconomic factors significantly influence XR adoption in Palestinian higher education, particularly in infrastructure investment, faculty training, and institutional support. The high costs of XR hardware, software, and maintenance, coupled with limited government funding and restricted access to grants, present a major challenge. Many universities struggle to scale XR beyond pilot initiatives, limiting widespread faculty adoption. A key barrier is faculty training and professional development. While some institutions offer workshops, many educators lack consistent training and technical support, resulting in uneven adoption across disciplines. Comparisons with developing regions, such as Latin America and Southeast Asia, reveal similar constraints, while well-funded institutions in South Korea and Germany address these challenges through government investment, faculty incentives, and public-private partnerships.

To bridge this digital divide, Palestinian universities must increase funding, establish training programs, and explore resource-sharing models across institutions. Addressing these socioeconomic barriers will ensure sustainable XR integration, enabling faculty to effectively utilize immersive technologies in medical education. Future research should explore scalable funding models and institutional collaborations to support long-term XR adoption. In conclusion, we visualize the factors influencing faculty members’ use of XR in medical education.

Figure 2 visualizes the factors influencing faculty members’ use of XR in medical education.

XSL•FO
RenderX

Figure 2. Visualization of the factors influencing faculty members to use extended reality (XR) in medical education.



Discussion

Principal Findings

The findings of this study highlight the complex factors influencing XR adoption in medical education, categorized into external, internal, and design-related elements. Professional development emerged as a key enabler, with faculty members who participated in XR training reporting increased confidence and capability in integrating the technology into their teaching. These sessions provided both technical knowledge and collaborative learning environments, aligning with prior research emphasizing the role of continuous professional development in technology adoption [40]. However, technical support and infrastructure remain critical challenges. While access to reliable support enhances faculty engagement with XR, inconsistent availability of assistance and limited institutional investment in technical staff hinder seamless integration. Similarly, infrastructure gaps—such as limited access to XR devices, inadequate internet connectivity, and insufficient educational resources—remain substantial barriers, especially in resource-constrained settings [41]. These findings align with studies in other developing regions, where high costs and inadequate infrastructure are primary obstacles to XR adoption [42].

Although this study applies the UTAUT model to analyze XR adoption, the findings suggest that policy and institutional support are crucial facilitating conditions not explicitly accounted for in the framework. Additionally, socioeconomic constraints, including funding limitations and digital infrastructure challenges, significantly influence adoption behaviors [43]. To extend the theoretical framework, we propose a modified model that integrates 2 additional dimensions: Pedagogical Readiness, encompassing faculty training, instructional design capabilities, and institutional encouragement for XR use, and Technical and Logistical Support, emphasizing the role of digital infrastructure, maintenance, and technical assistance. These modifications offer a more contextualized perspective on XR adoption in developing regions, reinforcing the need for localized implementation strategies [44].

The role of social influence in XR adoption extends beyond institutional policies and peer encouragement. This study found

XSL•FO
RenderX

that faculty skepticism, generational differences in adoption, and student perceptions significantly influence XR use. While peer influence and institutional endorsement encourage adoption, some senior faculty members expressed skepticism about XR, fearing that it might disrupt traditional pedagogical methods rather than complement them. These concerns align with prior research on faculty resistance to emerging technologies [32]. By contrast, younger faculty members demonstrated greater openness, reflecting trends observed in broader educational technology adoption studies [30]. Additionally, students' positive engagement with XR significantly influenced faculty willingness to integrate the technology, reinforcing prior findings that student enthusiasm can drive faculty adoption [33]. However, some educators expressed concerns that XR might encourage passive rather than active learning, highlighting the need for interactive and problem-solving–oriented XR applications to maximize its educational impact [33].

To overcome financial barriers to XR adoption in Palestinian universities, alternative and sustainable funding models are essential. While current efforts often rely on short-term external grants, more resilient approaches—such as public-private partnerships, collaborations with technology firms, and the use of open-source XR platforms—could help support long-term implementation and scalability [45]. Although the return on investment in XR may not be immediately measurable in financial terms, it can be demonstrated through improvements in student performance, engagement, and retention. These outcomes contribute to institutional sustainability by reducing dropout rates and enhancing overall learning effectiveness [41].

Lastly, design-related challenges, particularly the complexity of XR tools and time constraints for faculty, emerged as barriers to effective integration. While many faculty members appreciated the interactive and immersive capabilities of XR, others found content creation and instructional design challenging, highlighting the need for user-friendly design tools and targeted training [46]. Digital competencies were also found to be a critical factor, with faculty members possessing stronger digital skills demonstrating greater ease in XR adoption. This underscores the importance of developing digital literacy as a core competency in medical education [47].

Overall, this study emphasizes the need for a holistic approach to XR adoption, integrating technical, economic, and pedagogical strategies. In comparison to universities in Latin America and Southeast Asia, where national digital education strategies and structured funding initiatives have facilitated XR adoption, Palestinian institutions require policy-driven interventions and regional partnerships to develop scalable, sustainable funding models [48]. Addressing these economic and infrastructural constraints will be essential to ensure that XR can be effectively integrated into medical education in underresourced contexts.

Theoretical and Practical Implications

The study on integrating XR in Palestinian health care education highlights key theoretical and practical implications. Theoretically, it advances technology acceptance models by identifying factors influencing XR adoption, including institutional policies, social influences, digital competencies, and attitudes. It also emphasizes the need for robust infrastructure and professional development to support technology integration.

Practically, effectively implementing XR in medical education requires a well-structured, phased approach to fully realize its transformative potential. The first priority for institutions should be establishing robust foundational infrastructure, including high-quality hardware and software solutions that are scalable and adaptable to evolving needs. This requires substantial initial investments, not only in technology but also in developing technical support systems to address challenges such as high costs, operational complexities, and the demands of maintaining cutting-edge solutions. Along with infrastructure development, it is essential to provide comprehensive training for educators and students, focusing on the digital literacy skills needed to use XR effectively. Clear guidelines should also be developed to ensure consistent, meaningful integration of XR into the curriculum.

Higher education institutions should also consider designing and adopting performance indicators specifically tailored to measure the success and impact of XR implementation. These indicators could include metrics such as student engagement levels, improvements in skill acquisition, and the cost-effectiveness of XR solutions. By establishing these benchmarks, institutions can monitor progress and identify areas for improvement, ensuring a data-driven approach to XR adoption.

Early adoption strategies should emphasize piloting risk-free, immersive simulations that allow educators and students to explore XR's capabilities in a controlled environment. These pilot programs serve to demonstrate the tangible value of XR, helping to build confidence among stakeholders and secure their buy-in for broader implementation. However, institutions should be cautious of potential pitfalls. For instance, underestimating the need for ongoing technical support can lead to system failures and diminished user satisfaction. Similarly, neglecting to align XR initiatives with specific, well-defined learning outcomes can result in unfocused or ineffective use of the technology. Finally, failing to allocate adequate resources for regular maintenance and updates may jeopardize the long-term sustainability of XR programs.

Addressing Technical and Economic Barriers to XR Adoption

Sustainable integration of XR in medical education requires a strategic focus on viable funding models, cost-effectiveness, and long-term impact. In resource-constrained settings, such as Palestinian universities, advancing XR implementation depends less on reiterating existing challenges and more on identifying innovative, context-sensitive solutions. Strategic partnerships—with private technology firms, medical institutions, and international funding bodies—can facilitate access to sponsored XR hardware, software, and training. These collaborations support the co-development of immersive learning programs and enable cost-sharing arrangements that reduce the financial burden on institutions.

XSL•FO
RenderX

Adopting open-source XR platforms also presents a promising avenue for sustainable integration. These tools offer flexibility in content creation and deployment without the high costs associated with proprietary systems, making them particularly suitable for universities with limited budgets. Beyond initial implementation, institutions must assess XR's return on investment through educational outcomes rather than direct financial metrics. Improvements in student engagement, knowledge retention, and academic performance are strong indicators of XR's value and can contribute to institutional sustainability by reducing dropout rates and enhancing graduate readiness.

To ensure scalability and impact, universities should adopt data-informed strategies, including cost-benefit analyses, pilot programs, and scalable deployment models. Aligning financial planning with pedagogical goals ensures that XR technologies are integrated not only as innovative teaching tools but also as sustainable investments in the future of medical education. A methodical, forward-looking approach enables institutions to transform economic limitations into opportunities for creative problem-solving and long-term growth.

Limitations

The study acknowledges several limitations. First, the research was conducted during the initial stages of XR adoption in Palestinian higher education, which may limit the generalizability of the findings. The small sample size and the focus on specific institutions further constrain the applicability of the results to other contexts. Additionally, the high upfront costs and technical challenges associated with XR technologies may pose barriers that were not fully explored due to the limited scope of the study. Finally, the study relies on self-reported data from participants, which could introduce bias or inaccuracies in the findings.

Future Research

Future research should focus on scaling the study to larger populations across multiple universities to provide a more comprehensive understanding of XR's applicability and effectiveness in diverse educational contexts. Expanding research across different institutions, disciplines, and settings will offer broader insights into how XR can be integrated into various pedagogical frameworks.

Additionally, longitudinal studies are essential to track XR adoption over time. These studies would assess the long-term impact of XR on educational outcomes, skill retention, learner engagement, instructor effectiveness, and curriculum integration. Examining how faculty and students interact with XR technologies over extended periods will help identify patterns of adoption, sustained challenges, and evolving best practices. This approach will also contribute to understanding the long-term sustainability of XR implementation in higher education.

Further research should also investigate the technical and pedagogical challenges associated with XR adoption. Identifying these challenges could lead to detailed, actionable guidelines that institutions can use to optimize XR deployment strategies. Beyond health care education, exploring XR's potential in fields such as engineering, humanities, and business would provide insights into its broader applicability. Moreover, examining how XR interacts with emerging technologies such as artificial intelligence, machine learning, and data analytics may reveal innovative ways to enhance teaching and learning experiences.

Another critical area for future research is the development of performance indicators to measure the success of XR adoption. These indicators should assess learning outcomes, user satisfaction, cost-effectiveness, and scalability, providing institutions with data-driven benchmarks to evaluate and refine their XR initiatives.

Finally, addressing the digital divide in XR adoption is crucial, particularly in developing regions. Investigating how educational institutions can ensure equitable access to XR technologies for students from varied socioeconomic backgrounds will help create inclusive and accessible learning environments. This research will be instrumental in bridging technological disparities and promoting digital equity in higher education.

Conclusion

XR technologies have the potential to revolutionize health care education by providing immersive learning experiences that enhance practical skills and knowledge retention. This study highlights several key factors for the successful adoption of XR in medical education, including professional development, adequate infrastructure, robust technical support, positive social influence, and user-friendly design. Strategic investments in these areas are vital to overcoming initial barriers and aligning XR adoption with the Sustainable Development Goals of quality education and good health. By addressing these complex factors, educational institutions can create an environment conducive to the successful integration of XR technology, ultimately improving teaching practices and student learning outcomes in medical and nursing programs, particularly in Palestine.

The findings of this study emphasize that successful XR adoption in medical education requires more than just technological availability—it demands strong institutional policies, sustained funding mechanisms, and structured faculty development programs. Higher education institutions must move beyond pilot initiatives and develop long-term strategies for integrating XR into curricula, supported by clear guidelines, resource-sharing models, and institutional incentives. Additionally, regional collaborations among universities in developing contexts could facilitate knowledge exchange and infrastructure sharing, reducing the financial burden on individual institutions. Future research should further explore **scalable policy interventions** that enable sustainable XR adoption, particularly in resource-constrained environments where technology-enhanced learning can play a crucial role in addressing educational inequalities.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview protocol.
[DOCX File , 15 KB - mededu_v11i1e65042_app1.docx]

Multimedia Appendix 2
A sample of the inductive thematic analysis used in this study.
[DOCX File , 17 KB - mededu_v11i1e65042_app2.docx]

References

1. Doolani S, Wessels C, Kanal V, Sevastopoulos C, Jaiswal A, Nambiappan H, et al. A review of extended reality (XR) technologies for manufacturing training. Technologies 2020 Dec 10;8(4):77. [doi: 10.3390/technologies8040077]
2. Morimoto T, Kobayashi T, Hirata H, Otani K, Sugimoto M, Tsukamoto M, et al. XR (extended reality: virtual reality, augmented reality, mixed reality) technology in spine medicine: status quo and quo vadis. J Clin Med 2022 Jan 17;11(2):470 [FREE Full text] [doi: 10.3390/jcm11020470] [Medline: 35054164]
3. Catbas FN, Luleci F, Zakaria M, Bagci U, LaViola JJ, Cruz-Neira C, et al. Extended Reality (XR) for Condition Assessment of Civil Engineering Structures: A Literature Review. Sensors (Basel) 2022 Dec 06;22(23):9560 [FREE Full text] [doi: 10.3390/s22239560] [Medline: 36502261]
4. Barteit S, Lanfermann L, Bärnighausen T, Neuhann F, Beiersmann C. Augmented, Mixed, and Virtual Reality-Based Head-Mounted Devices for Medical Education: Systematic Review. JMIR Serious Games 2021 Jul 08;9(3):e29080 [FREE Full text] [doi: 10.2196/29080] [Medline: 34255668]
5. Behmadi S. Virtual reality-based medical education versus lecture-based method in teaching start triage lessons in emergency medical students: Virtual reality in medical education. Journal of Advances in Medical Education & Professionalism 2022 Jul 31;10(1):1-6. [doi: 10.1164/ajrccm-conference.2023.207.1_meetingabstracts.a1676]
6. Alam A, Mohanty A. Foundation for the future of higher education or 'misplaced optimism'? Being human in the age of artificial intelligence. Innovations in Intelligent Computing and Communication 2023 Jan 1:17-29.
7. Xiong J, Hsiang E, He Z, Zhan T, Wu S. Augmented reality and virtual reality displays: emerging technologies and future perspectives. Light Sci Appl 2021 Oct 25;10(1):216 [FREE Full text] [doi: 10.1038/s41377-021-00658-8] [Medline: 34697292]
8. Bargeri S. , Scalea, S. , Agosta, F., Banfi, G., Corbetta, D., Filippi, M., ... & Gianola, S. Effectiveness and safety of virtual reality rehabilitation after stroke: an overview of systematic reviews.EClinicalMedicine. 2023 :64.
9. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
10. Davis F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
11. Khlaif ZN, Sanmugam M, Hattab MK, Bensalem E, Ayyoub A, Sharma RC, et al. Mobile technology features and technostress in mandatory online teaching during the COVID-19 crisis. Heliyon 2023 Aug;9(8):e19069 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e19069] [Medline: 37636397]
12. Park Y, Tak YW, Kim I, Lee HJ, Lee JB, Lee JW, et al. User Experience and Extended Technology Acceptance Model in Commercial Health Care App Usage Among Patients With Cancer: Mixed Methods Study. J Med Internet Res 2024 Dec 18;26:e55176 [FREE Full text] [doi: 10.2196/55176] [Medline: 39693615]
13. Wang X, Young GW, Iqbal MZ, Guckin CM. The potential of extended reality in Rural Education’s future – perspectives from rural educators. Educ Inf Technol 2023 Sep 04;29(7):8987-9011. [doi: 10.1007/s10639-023-12169-7]
14. Bach M. P.., Understanding Determinants of Management Simulation Games Adoption in Higher Educational Institutions Using an Integrated Technology Acceptance Model/Technology?Organisation?Environment Modelducator Perspective. Information 2025;16(1):E. [doi: https://doi.org/10.3390/info16010045]
15. Skalidis I. , Muller, O. , & Fournier, S. CardioVerse: The cardiovascular medicine in the era of Metaverse.Trends in cardiovascular medicine,2023 ;33(8):471. [doi: https://doi.org/10.1016/j.tcm.2022.05.004]
16. Tang Y. M.., A systematic review of immersive technology applications for medical practice and education-trends, application areas, recipients, teaching contents, evaluation methods, and performance. Educational Research Review 2022;35:100429. [doi: https://doi.org/10.1016/j.edurev.2021.100429]
17. Suh I. , T. McKinney, and K. -C. Siu. Current perspective of metaverse application in medical education, research and patient care. in Virtual Worlds. 2023. MDPI . [doi: https://doi.org/10.3390/virtualworlds2020007]
18. Garlinska M. The influence of emerging technologies on distance education. Electronics 2023;12(7):1550. [doi: https://doi.org/10.3390/electronics12071550]

XSL•FO
RenderX

19. Marougkas A. How personalized and effective is immersive virtual reality in education? A systematic literature review for the last decade. Multimedia Tools and Applications 2024;83(6):18185-18233. [doi: https://doi.org/10.1007/s11042-023-15986-7]
20. Lee A. T., R.K. Ramasamy, and A. Subbarao, Understanding Psychosocial Barriers to Healthcare Technology Adoption: A Review of TAM Technology Acceptance Model and Unified Theory of Acceptance and Use of Technology and UTAUT Frameworks. Healthcare 2025;13(3):A. [doi: https://doi.org/10.3390/healthcare13030250]
21. Kuleto V. Extended reality in higher education, a responsible innovation approach for generation y and generation z. Sustainability 2021;13(21):11814. [doi: https://doi.org/10.3390/su132111814]
22. Serna-Mendiburu G. M. C.R. Guerra-Tamez. Shaping the future of creative education: the transformative power of VR in artdesign learning. in Frontiers in Education. 2024. Frontiers Media SA . [doi: https://doi.org/10.3389/feduc.2024.1388483]
23. Aguayo C, Eames C. Using mixed reality (XR) immersive learning to enhance environmental education. The Journal of Environmental Education 2023 Mar 16;54(1):58-71. [doi: 10.1080/00958964.2022.2152410]
24. González-Erena P. V., S. Fernández-Guinea, and P. Kourtesis, Cognitive Assessment and Training in Extended Reality: Multimodal Systems, Clinical Utility, and Current Challenges. Encyclopedia 2025;5(1):8. [doi: https://doi.org/10.3390/encyclopedia5010008]
25. Goudman L. , Jansen, J. , Billot, M., Vets, N., De Smedt, A., Roulaud, M., ... & Moens, M. Virtual reality applications in chronic pain management: systematic review and meta-analysis.JMIR Serious Games,2022. , e3 ;10(2):4402. [doi: https://doi.org/10.2196/34402]
26. Catania V. , Rundo, F. , Panerai, S., & Ferri, R. Virtual reality for the rehabilitation of acquired cognitive disorders: a narrative review.Bioengineering,2023 ;11(1):35. [doi: https://doi.org/10.3390/bioengineering11010035]
27. Samnakay S. , Bell, E. , Evans, D., Sommerfield, D., Sommerfield, A., Hauser, N., & von Ungern?Sternberg, B. S. Assessing the Use and Acceptability of Virtual Reality to Assist Coping in Children Undergoing Clinical Procedures.Journal for Specialists in Pediatric Nursing,2025. , e7 ;30(1):0002. [doi: https://doi.org/10.1111/jspn.70002]
28. de G. Adopting extended reality? A systematic review of manufacturing training and teaching applications. Journal of manufacturing systems 2023;71:645-663. [doi: https://doi.org/10.1016/j.jmsy.2023.10.016]
29. Mistry D, Brock CA, Lindsey T. The Present and Future of Virtual Reality in Medical Education: A Narrative Review. Cureus 2023 Dec;15(12):e51124 [FREE Full text] [doi: 10.7759/cureus.51124] [Medline: 38274907]
30. Arthur T, Melendez-Torres G, Harris D, Robinson S, Wilson M, Vine S. Extended Reality Interventions for Health and Procedural Anxiety: Panoramic Meta-Analysis Based on Overviews of Reviews. J Med Internet Res 2025 Jan 08;27:e58086 [FREE Full text] [doi: 10.2196/58086] [Medline: 39778203]
31. Burian B, Ebnali M, Robertson J, Musson D, Pozner C, Doyle T, et al. Using extended reality (XR) for medical training and real-time clinical support during deep space missions. Appl Ergon 2023 Jan;106:103902. [doi: 10.1016/j.apergo.2022.103902] [Medline: 36162274]
32. Chuah SHW. Wearable XR-technology: literature review, conceptual framework and future research directions. IJTMKT 2019;13(3/4):205. [doi: 10.1504/ijtmkt.2019.104586]
33. Curran VR, Xu X, Aydin MY, Meruvia-Pastor O. Use of extended reality in medical education: an integrative review. Med Sci Educ 2023 Feb 19;33(1):275-286 [FREE Full text] [doi: 10.1007/s40670-022-01698-4] [Medline: 36569366]
34. Kluge MG, Maltby S, Keynes A, Nalivaiko E, Evans DJR, Walker FR. Current state and general perceptions of the use of extended reality (XR) technology at the University of Newcastle: interviews and surveys from staff and students. Sage Open 2022 Apr 26;12(2):21582440221093348. [doi: 10.1177/21582440221093348]
35. Li N, Keskitalo T. Literature review - XR in medical and healthcare education. WHO. 2022. URL: https://ec.europa.eu/programmes/erasmus-plus/project-result-content/f91f0f50-2e07-4d8e-a459-eaf7e0d3ebe4/PrepaCare-IO2_Literature_Review_Publication.pdf [accessed 2025-04-28]
36. Creswell J. Qualitative Research Designs Selection and Implementation. The Counseling Psychologist 2007;35:236-264.
37. Campbell S. Purposive sampling: complex or simple? Research case examples. J Res Nurs 2020;25(8):652-661.
38. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
39. Ngozwana N. Ethical dilemmas in qualitative research methodology: researcher's reflections. Int J Educ Methodol 2018 Feb 15;4(1):19-28. [doi: 10.12973/ijem.4.1.19]
40. Jin Y. Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. Computers and Education: Artificial Intelligence 2025;8:A. [doi: https://doi.org/10.1016/j.caeai.2024.100348]
41. Luo S. , D. Zou, and L. Kohnke, A systematic review of research on xReality (XR) in the English classroom: Trends, research areas, benefits, and challenges. Computers & Education: X Reality 2024;4:100049. [doi: https://doi.org/10.1016/j.cexr.2023.100049]
42. Dinh A. Perceptions about augmented reality in remote medical care: Interview study of emergency telemedicine providers. JMIR Formative Research 2023;7:e45211. [doi: https://doi.org/10.2196/45211]
43. Dingel J. Predictors of Health Care Practitioners? Intention to Use AI-Enabled Clinical Decision Support Systems: Meta-Analysis Based on the Unified Theory of Acceptance and Use of Technology. Journal of medical internet research 2024;26:e57224. [doi: https://doi.org/10.2196/57224]

XSL•FO
RenderX

44. Fugate J. , M. Tonsanger, and S. Macrine, Immersive Extended Reality (I-XR) in Medical and Nursing Education: A Systematic Review and Pedagogical Directives 2025 Jan 13:1-50. [doi: 10.20944/preprints202501.0928.v1]
45. Kelly S. , S.-A. Kaye, and O. Oviedo-Trespalacios, What factors contribute to the acceptance of artificial intelligence? A systematic review. Telematics and Informatics 2023;77:101925.
46. Meccawy M. Teachers? prospective attitudes towards the adoption of extended reality technologies in the classroom: Interests and concerns. Smart Learning Environments 2023;10(1):36. [doi: https://doi.org/10.1186/s40561-023-00256-8]
47. Mainz A. Measuring the Digital Competence of Health Professionals: Scoping Review. JMIR Medical Education 2024;10(1):e55737. [doi: https://doi.org/10.2196/55737]
48. Al-Adwan A. S.., Extending the technology acceptance model (TAM) to Predict University Students? intentions to use metaverse-based learning platforms. Education and Information Technologies 2023;28(11):15381-15413. [doi: https://doi.org/10.1007/s10639-023-11816-3]

Abbreviations

AR: augmented reality
MR: mixed reality
TAM: Technology Acceptance Model
UTAUT: Unified Theory of Acceptance and Use of Technology
VR: virtual reality
XR: extended reality

Edited by A Bahattab; submitted 03.08.24; peer-reviewed by M Smith, D Patel, T Moser; comments to author 01.01.25; revised version received 15.01.25; accepted 23.03.25; published 01.05.25.

Please cite as:
Khlaif Z, Salama N, Hamamra B, Mousa A
Factors Influencing Educators' Perspectives on Accepting Extended Reality in Health Care Education: Qualitative Study
JMIR Med Educ 2025;11:e65042
URL: https://mededu.jmir.org/2025/1/e65042
doi:10.2196/65042
PMID:40310667

©Zuheir Khlaif, Nisreen Salama, Bilal Hamamra, Allam Mousa. Originally published in JMIR Medical Education (https://mededu.jmir.org), 01.05.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Assessing ChatGPT's Capability as a New Age Standardized Patient: Qualitative Study

Joseph Cross[1*], PhD; Tarron Kayalackakom[2*], MD; Raymond E Robinson[3], MPH, MBA, MD, EdS; Andrea Vaughans[4], MD; Roopa Sebastian[4], MSc, PhD; Ricardo Hood[2], MD; Courtney Lewis[2], MD; Sumanth Devaraju[2], MD; Prasanna Honnavar[2], PhD; Sheetal Naik[2], MD; Jillwin Joseph[2], PhD; Nikhilesh Anand[5], MD; Abdalla Mohammed[6], BSc; Asjah Johnson[6], BSc; Eliran Cohen[6], BSc; Teniola Adeniji[6], BSc; Aisling Nnenna Nnaji[6], BSc; Julia Elizabeth George[6]

[1]Medical University of the Americas, PO Box 701, Charlestown, Saint Kitts and Nevis

[2]Department of Education Enhancement, College of Medicine, American University of Antigua, St Johns, Antigua and Barbuda

[3]Department of Health Informatics, School of Professional Studies, Northwestern University, Evanston, IL, United States

[4]Department of Biochemistry, Cell Biology and Genetics, College of Medicine, American University of Antigua, Basseterre, Antigua and Barbuda

[5]Department of Medical Education, School of Medicine, University of Texas Rio Grande Valley, Edinburgh, TX, United States

[6]School of Medicine, Xavier University, Orangestad, Aruba

[*]these authors contributed equally

Corresponding Author:
Joseph Cross, PhD
Medical University of the Americas, PO Box 701, Charlestown, Saint Kitts and Nevis

Abstract

Background: Standardized patients (SPs) have been crucial in medical education, offering realistic patient interactions to students. Despite their benefits, SP training is resource-intensive and access can be limited. Advances in artificial intelligence (AI), particularly with large language models such as ChatGPT, present new opportunities for virtual SPs, potentially addressing these limitations.

Objectives: This study aims to assess medical students' perceptions and experiences of using ChatGPT as an SP and to evaluate ChatGPT's effectiveness in performing as a virtual SP in a medical school setting.

Methods: This qualitative study, approved by the American University of Antigua Institutional Review Board, involved 9 students (5 females and 4 males, aged 22‑48 years) from the American University of Antigua College of Medicine. Students were observed during a live role-play, interacting with ChatGPT as an SP using a predetermined prompt. A structured 15-question survey was administered before and after the interaction. Thematic analysis was conducted on the transcribed and coded responses, with inductive category formation.

Results: Thematic analysis identified key themes preinteraction including technology limitations (eg, prompt engineering difficulties), learning efficacy (eg, potential for personalized learning and reduced interview stress), verisimilitude (eg, absence of visual cues), and trust (eg, concerns about AI accuracy). Postinteraction, students noted improvements in prompt engineering, some alignment issues (eg, limited responses on sensitive topics), maintained learning efficacy (eg, convenience and repetition), and continued verisimilitude challenges (eg, lack of empathy and nonverbal cues). No significant trust issues were reported postinteraction. Despite some limitations, students found ChatGPT as a valuable supplement to traditional SPs, enhancing practice flexibility and diagnostic skills.

Conclusions: ChatGPT can effectively augment traditional SPs in medical education, offering accessible, flexible practice opportunities. However, it cannot fully replace human SPs due to limitations in verisimilitude and prompt engineering challenges. Integrating prompt engineering into medical curricula and continuous advancements in AI are recommended to enhance the use of virtual SPs.

(JMIR Med Educ 2025;11:e63353) doi:10.2196/63353

KEYWORDS

medical education; standardized patient; AI; ChatGPT; virtual patient; assessment; standardized patients; LLM; effectiveness; medical school; qualitative; flexibility; diagnostic

XSL•FO
RenderX

Introduction

Standardized patients (SPs) have been a cornerstone of medical education since the 1960s, offering students an immersive, real-world experience in a controlled environment. Studies have demonstrated that SP programs are superior for teaching consultation skills compared with traditional methods, with medical students trained using SPs showing increased confidence and competency compared with those trained through other modalities [1,2].

While SPs provide valuable opportunities for students to practice diagnostic and interpersonal skills under standardized conditions, several inherent challenges exist. The resource-intensive nature of SP programs has been a persistent issue, with significant costs associated with recruitment, training, and maintenance of an SP bank [1,3]. Additionally, questions have emerged about SPs' ability to adequately represent the nuances of real patient presentations.

These challenges are particularly pronounced in specific contexts. For instance, Caribbean medical schools face unique obstacles due to limited local health care infrastructure and varying access to clinical training resources. Many offshore institutions in countries such as Aruba and Antigua and Barbuda must rely on partnerships with local health care providers, often resulting in inconsistent access across student cohorts [4,5]. The COVID-19 pandemic exposed additional vulnerabilities in traditional SP programs. The discontinuation of the USMLE Step 2 Clinical Skills examination in 2022, for instance, highlighted the risks of relying solely on in-person SP encounters for assessment [5].

In the 21st century, virtual SPs have emerged. These are computer programs that simulate specific illnesses and respond to learner inputs [6]. They have become invaluable tools in both teaching and assessment. However, their development also requires significant resources, making it challenging for institutions without robust educational technology support departments [7].

As the field of artificial intelligence (AI) has advanced, the potential for its application in medical education has expanded. Large language models (LLMs), such as ChatGPT (OpenAI), have revolutionized natural language processing. These sophisticated neural networks, trained on vast amounts of web-based data, are adept at predicting subsequent words in a sequence [8]. ChatGPT, a chatbot based on the GPT-3.5 model, has an enormous 175 billion parameters and displays a remarkable capacity for understanding and reasoning, bordering on human-like proficiency [9]. Since its introduction in November 2022, sectors spanning from history to entertainment have rapidly adopted the LLM [10].

This advancement in AI has led to the development of virtual SP chatbots. A number of major educational material suppliers and specialized companies are offering chatbot SPs, based on LLMs capable of natural language interactions, for students to practice clinical skills. One example is Oscer, which can present more than 200 virtual patient conditions and boasts above 90% accuracy in symptomology [11]. Similarly, the University of Texas Medical Branch makes use of an AI agent termed Virti, which they use to conduct virtual Observed Structured Clinical Examinations with medical students [12]. Other publicly accessible sites offering virtual patients include Soma Lab [13] and Body Interact [14]. However, for this new generation of virtual patients there is again considerable time and resources required for the company or the institution to develop the program and train the LLM on specific datasets and student access can be limited by cost and locality [7].

The debut of ChatGPT sparked inquiries into its potential as an SP. Liu et. al [15] crafted 10 medical histories with ChatGPT, which were then vetted by experienced physicians. Their results highlighted ChatGPT's promise in clinical education, although some responses came across as robotic [15]. Suarez et.al [16] gathered dental student's feedback after interacting with an AI chatbot. The majority found the experience valuable, especially those who made a correct diagnosis. This underscores the potential of integrating AI into health sciences training [16].

Weidener and Fischer [17] emphasized the growing consensus on incorporating AI into medical education. Their study indicated the importance of both practical and technological skills for leveraging AI in medicine [17]. Similarly, Jowsey et. al [18] have recommended adoption of AI into medical education as a way of preparing future physicians for the reality of modern practice.

We were aware that SPs at our school, American University of Antigua (AUA), were in limited supply and had received feedback indicating that while SPs are effective, students would like greater access to them. In fact, some students had no access during their course, depending on their cohort.

One of our study's aims was to assess medical students' perceptions and experiences regarding the use of AI in medicine—specifically by examining their views before and after interacting with ChatGPT as an SP. A second aim was to evaluate whether ChatGPT can perform adequately as a virtual SP in a medical school setting. Guided by these aims, our investigation focused on the following research questions: (1) How do students perceive the effectiveness of ChatGPT compared with traditional SPs in medical training scenarios? (2) To what extent can ChatGPT function effectively as a virtual SP in medical education?

By addressing these questions, our study seeks to inform the potential integration of AI-driven virtual SPs into medical curricula, particularly in settings where access to traditional SPs is limited.

Methods

Ethical Considerations

This study was given expedited approval by the AUA Research committee (no. AUAIRBa23011). Eleven medical student volunteers enrolled in the MD course at AUA were recruited via a campus-wide email. Two participants were lost to follow-up, leaving a total of 9 participants. Students were 5 females and 4 males, aged 22-48 years, comprising students from both first and second years of the basic sciences course

XSL•FO
RenderX

section of the MD program. Participants were explicitly informed that their involvement in the research was completely voluntary. They were also assured that their responses would remain confidential and anonymous, and all participants signed informed consent agreements. All data were anonymized and no compensation was provided to participants.

Study Design

Students were given access to ChatGPT version 4.0 accounts, the most recent available at the time of the study. Students were observed during a live role-play, in which a student inputted a predetermined prompt, provided by the study authors, into ChatGPT. The prompt directed the LLM to present as a patient with a neurological condition (Figure 1).

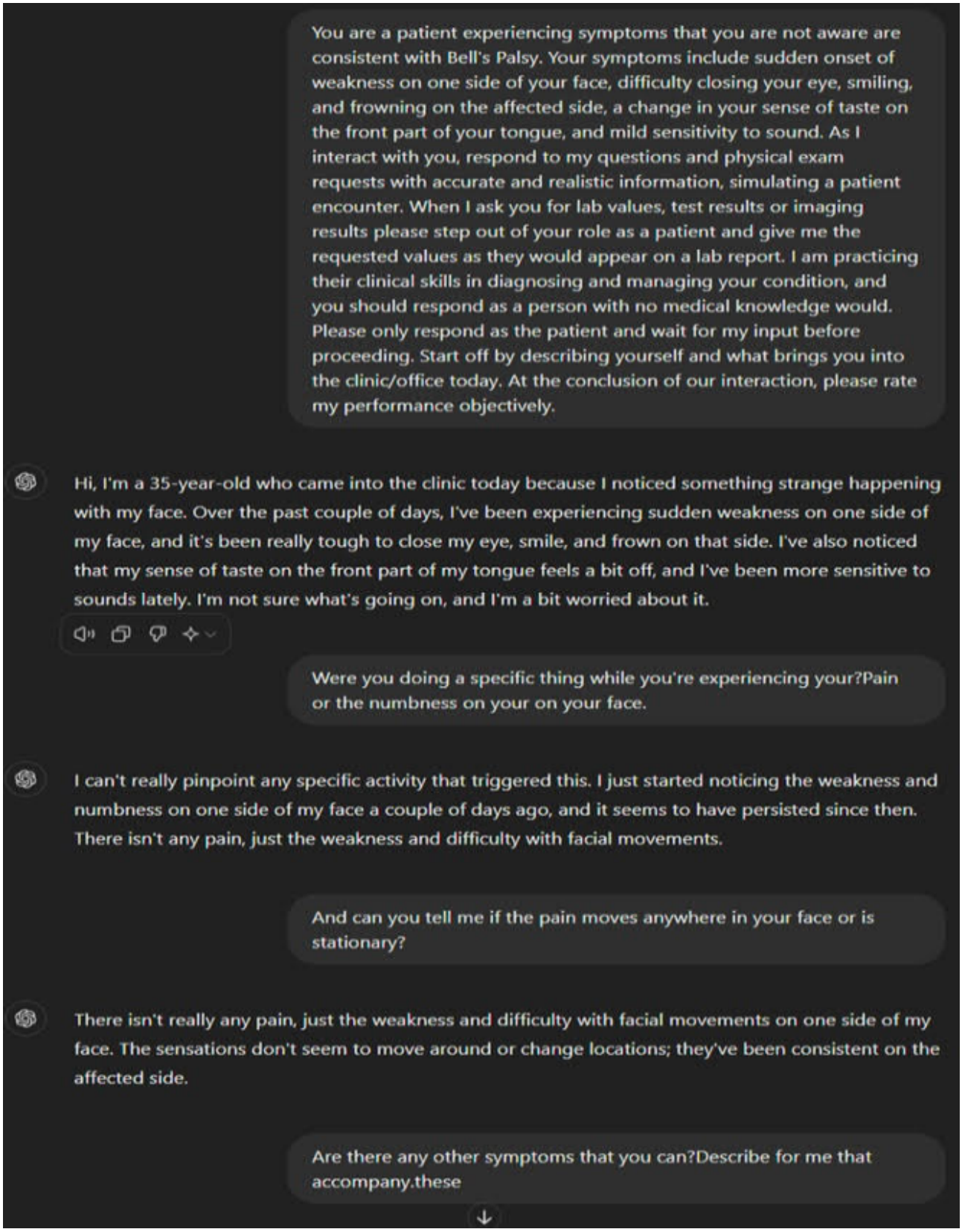
The student, in the role of physician, then interviewed the ChatGPT and attempted to make a differential diagnosis (Figure 2).

Figure 1. Prompt used in ChatGPT role-play.

You are a patient experiencing symptoms that you are not aware are consistent with Bell's Palsy. Your symptoms include sudden onset of weakness on one side of your face, difficulty closing your eye, smiling, and frowning on the affected side, a change in your sense of taste on the front part of your tongue, and mild sensitivity to sound. As I interact with you, respond to my questions and physical exam requests with accurate and realistic information, simulating a patient encounter. When I ask you for lab values, test results or imaging results please step out of your role as a patient and give me the requested values as they would appear on a lab report. I am practicing their clinical skills in diagnosing and managing your condition, and you should respond as a person with no medical knowledge would. Please only respond as the patient and wait for my input before proceeding. Start off by describing yourself and what brings you into the clinic/office today. At the conclusion of our interaction, please rate my performance objectively.

XSL•FO
RenderX

Figure 2. Screenshot example of ChatGPT standardized patient interaction.



Prompt Development

The development of the prompt for ChatGPT's simulated patient interaction underwent an iterative process prior to its use by students. This process involved a 6-member faculty team comprising both clinical and nonclinical faculty, ensuring a diverse range of perspectives and expertise. The faculty were tasked with using the prompt in simulated interactions with ChatGPT, assessing the following factors:

XSL•FO
RenderX

1. Consistency: ensuring the chatbot consistently adhered to the patient role and provided responses aligned with the illness script.
2. Accuracy: evaluating whether the responses were medically plausible and aligned with the provided case information.
3. Likelihood of misleading the SP: assessing whether the chatbot responses could inadvertently lead users to incorrect assumptions or conclusions.
4. Quality of output: reviewing the depth and appropriateness of responses to ensure a realistic and effective simulation experience.
5. Adherence to prompt instructions: verifying that ChatGPT's responses followed the specific behavioral and informational instructions embedded in the prompt.

Faculty provided detailed feedback based on their observations, leading to refinements in the prompt. Suggestions included adjustments to phrasing, additional clarifications to the illness script, and enhancements to behavioral instructions to minimize the potential for ChatGPT to deviate from the assigned patient role. This iterative process was instrumental in optimizing the prompt's effectiveness before deployment in the study.

Rationale for Clinical Case Selection

Bell palsy was chosen as the clinical condition for the simulation due to its relevance to the material being taught at the time. This alignment ensured that the scenario was both clinically pertinent and integrated with the participants' ongoing coursework in both basic sciences and clinical disciplines. The familiarity of the students with the foundational aspects of Bell palsy was intended to facilitate meaningful engagement with the simulated patient, allowing them to focus on the interaction and diagnostic questioning rather than struggling with unfamiliar content.

Purpose of the Evaluation

It is important to note that the primary goal of this study was not to evaluate the students' diagnostic accuracy. Instead, the focus was on assessing their perceptions of ChatGPT's performance as a simulated patient. This distinction was critical to the study design, as it allowed for an emphasis on the usability, realism, and educational value of AI-driven SP interactions without conflating these aspects with the participants' clinical competencies.

The role-play was conducted verbally, as a voice control extension added to the ChatGPT accounts allowed natural language conversation between the student and the LLM [19]. A structured questionnaire consisting of 15 open-ended questions was administered before and after interaction with ChatGPT in the role of an SP. Students were asked about specific elements of their interaction and interviews were conducted in person by faculty team members (Multimedia Appendix 1).

Participating students were introduced to the ethical considerations of using LLMs such as ChatGPT. This included training on the importance of deidentifying patient data, recognizing the limitations of AI, and understanding the potential biases inherent in AI responses, such as those related to gender or ethnicity. This ethical orientation aimed to ensure that students approached the interactions responsibly and with an awareness of the technology's constraints.

Thematic Analysis

The results of the students' group work were recorded, transcribed, and coded by 3 different authors (JC, TK, and RER). Following discussions in regular meetings, findings were summarized, and a category system consisting of main and subcategories, according to Mayring's [20] qualitative content analysis, was agreed upon. Selected text passages were used as quotations to illustrate each category. Inductive category formation, a qualitative research method used to analyze data by identifying patterns, themes, or categories that emerge directly from the data itself, without predefined hypotheses or coding frameworks, was used to analyze open-ended survey responses and interview transcripts.

To explore differences in prompt engineering techniques across academic levels, we asked students to describe how they approached questioning and refining their prompts during the postsession interviews. First-year students, who had less clinical exposure, were expected to rely more on general inquiry methods, while second-year students might leverage slightly more clinical insight. Recording these observations allowed us to compare prompt engineering strategies between these groups and understand how curriculum familiarity influenced interactions with the AI-driven simulated patient interactions.

Results

A total of 9 students participated (5 females and 4 males, aged 22 - 48 years) (Table 1). All students had had some prior experience with traditional SPs, with more senior students having had a greater number of encounters. This contextualizes their perceptions of ChatGPT as a supplement and provides a baseline for understanding the comparative effectiveness of the AI-based approach.

XSL•FO
RenderX

Table . Demographic data.

Characteristics	Participants (n=9), n
Age (years)[a]	
22 - 30	4
31 - 40	4
41+	1
Sex	
Male	4
Female	5
Semester	
1	0
2	7
3	1
4	1
5	0
Prior experience with SPs[b]	
Yes	9
No	0
Prior experience with AI[c]/ChatGPT	
Yes	6
No	3

[a]Mean age: 31.22 (SD 6.8) years

[b]SPs: standardized patients.

[c]AI: artificial intelligence.

The thematic analysis of student feedback prior to interaction with ChatGPT as an SP identified several key themes and subthemes (Table 2). Under the theme of technology limitations, students noted challenges with prompt engineering, such as difficulty in asking effective questions, because the AI could not role-play a physical examination. In terms of learning efficacy, students mentioned the potential for personalized learning materials, grammatical assistance, and the ability for repeated practice without the constraints of limited SP availability. Additionally, some students highlighted the potential for increased convenience, as they could practice as often and whenever they wanted. A potential reduction in SP interview stress was also seen as a benefit of increasing virtual practice. However, under the theme of verisimilitude (ie, the degree to which a simulation mirrors real-life scenarios, including the subtle behaviors and interactions that contribute to a convincing experience), students expressed concerns about the absence of visual cues and rapport, which are important in real patient interactions. Finally, trust issues were raised regarding the accuracy of the LLMs output.

XSL•FO
RenderX

Table . Thematic analysis of student feedback preinteraction with ChatGPT standardized patient.

Themes and subthemes	Representative quotations
Theme 1. Technology limitations	
Prompt engineering	"The challenges might be just asking the right questions, because it's an AI, you can't ask them to do physical examinations."
Theme 2. Learning efficacy	
Personalized learning materials	"Triple checking work and not only getting the right answer, but getting explanations for the right answer and then why the wrong answer is wrong."
Grammatical assistance	"It would be helpful because English is not my first language."
Repetition	"There's usually 10 medical students to one patient, and sometimes you're fighting over each other to get the interview, so this allows us to get more repetitions."
Depth of medical knowledge	"The sky's the limit with regards to what we can practice."
Interview stress or anxiety	"It will kind of be a bit more stress free because you know you're talking to a computer rather than an actual patient."
Convenience	"Be able to practice it as much as I want, as often as I want and any time I want."
Theme 3. Verisimilitude	
Absence of visual cues	"You have to figure out ways to ask the question without the visual cues."
Absence of rapport or empathy	"Building the communication and the relationship with your patient is important."
Theme 4. Trust	
Inaccurate output	"One incident was in the small group activity, where it gave us the wrong answer."

Following interaction with ChatGPT, the thematic analysis of student feedback revealed some changes in perceptions (Table 3). While technology limitations were still noted, students mentioned that they had learnt to improve the output from ChatGPT by tailoring prompts. They also reported alignment issues, such as ChatGPT not providing information on sensitive topics such as patient sexual history. Learning efficacy remained a significant theme, with students appreciating the convenience and repetition benefits. They found the ability to practice history taking without stress and receive feedback useful for skill development. However, verisimilitude issues persisted as a theme, with students noting the absence of visual and tonal cues, and the lack of rapport and empathy, all of which impacted the effectiveness of the patient interview and the ability to make a diagnosis. Some students experienced information overload, feeling that ChatGPT provided more information than a real patient would.

XSL•FO
RenderX

Table . Thematic analysis of student feedback postinteraction with ChatGPT standardized patient.

Themes and subthemes	Representative quotations
Theme 1. Technology limitations	
Prompt engineering	"You could put in the prompt that you want to tailor the responses you want to get back."
Alignment	"When I asked like about sexual history, they were not able to give information."
Theme 2. Learning efficacy	
Convenience	"Having ChatGPT to practice history whenever we want, I think that's the improvement."
Repetition	"You are able to have a lot more repetitions than you are in lab."
Interview stress or anxiety	"Since it's a computer, it's not real. I had less anxiety."
Feedback	"I can ask 'hey, how did you think I did?'"
Skills development	"It highlighted the importance of on-the-spot thinking and memory recall in a medical scenario."
Overall enhanced learning	"It's going to make you sharper. You know, you're probably going to be ahead of your peers, you're going to be able to answer a patient in a better, more detailed manner. Give them a better treatment or care plan."
Theme 3. Verisimilitude	
Absence of visual cues	"For the standardized patient you physically see them. You can see if they're in pain, they don't have to explain where they are in pain."
Absence of tonal cues	"ChatGPT had the same tone, even if it was saying something sad."
Absence of rapport or empathy	"It takes away the personal connection between the doctor and the patient."
Information overload	"It felt like it was offering more information than a regular patient would."

To provide broader context, we compared ChatGPT with some other virtual SP platforms or platforms that could provide this function (Table 4). The comparison highlights the unique strengths and weaknesses of ChatGPT identified in this study in comparison with other platforms, including Claude AI (another chatbot often ranking near the top of benchmarking tables), Body Interact, Oscer AI, and Soma Lab [13,14,21-23]. Both ChatGPT and Claude AI offer flexibility and unlimited practice but are limited by uncurated outputs and reliance on prompt engineering. Oscer AI and Som Lab provide curated clinical cases with tailored feedback, yet their visual representation and interactivity vary, with Soma Lab integrating natural conversational voice modes. Body Interact enhances verisimilitude through patient avatars and curated cases but lacks voice interaction. Cost structures range from free access for basic use to subscription-based models for advanced features.

Table . Comparison of various platforms able to function as standardized patients.

Platform	Technology limitations	Learning efficacy	Verisimilitude	Model cost
ChatGPT	Requires effective prompt engineering; uncurated outputs	Offers flexibility and unlimited practice	Limited visual and tonal cues; natural conversational voice mode	Free and subscription-based options
Claude AI	Requires effective prompt engineering; uncurated outputs	Offers flexibility and unlimited practice	Limited visual and tonal cues; limited voice interaction	Free and subscription-based options
Body Interact	Requires effective prompt engineering; curated clinical cases	Facilitates skill development through realistic scenarios	Patient avatars; lacks voice interaction	Subscription or licensing fees
Oscer AI	Requires effective prompt engineering; wide range of curated clinical cases	Focus on history-taking skills with limited versatility	Limited visual cues; voice interaction possible	Free. Subscription for full access
Soma Lab	Requires effective prompt engineering; wide range of curated clinical cases	Counseling-focused; supports repeated practice with tailored feedback	Static patient avatars; natural conversational voice mode	Variable costs based on usage and features

XSL•FO
RenderX

Discussion

Principal Findings

This study investigated the use of ChatGPT as an SP by qualitative analysis of students' responses to a questionnaire, preinteraction and postinteraction, with ChatGPT performing the role of SP. In terms of diagnostic skill development, our conclusions were drawn from a combination of faculty observations and student self-report. Faculty members who observed the sessions noted that students demonstrated more structured reasoning and improved question formulation after repetitive practice with ChatGPT. In postsession interviews, students themselves expressed feeling more confident and organized in their clinical reasoning steps. This alignment between external observation and self-assessment suggests that the interaction with ChatGPT, although lacking nonverbal cues and certain realistic elements, still provides a valuable platform for honing diagnostic interviewing skills. Thematic analysis provided insights into student perceptions. Major themes identified were technology limitations, learning efficacy, and verisimilitude. Our results suggest that the current version of ChatGPT (ChatGPT version 4.0 at the time of this study) can function effectively as an augmentation to traditional SPs but cannot fully substitute for SPs. These results are broadly in line with those of other studies using LLMs in the role of SP [24-29].

The technological limitations of LLMs in the context of SP exercises were both anticipated and confirmed in our study. The subtheme of prompt engineering was particularly important. Students were made aware of the importance of correctly worded prompts before the exercise, and we found that the faculty-provided prompt, developed through a trial and error process, proved effective in this regard.

The significance of prompt engineering when using LLMs as virtual SPs, or in developing related materials, is also supported by other studies [28,30-33]. It has been suggested that prompt engineering could be incorporated into medical curricula through, for example, hands-on workshops, simulation-based learning, and courses on AI in health care [28,30-32].

The postinteraction interviews also revealed an additional subtheme of alignment. Alignment refers to the problem of ensuring that AI acts in accordance with human intentions and human values [34]. Students noted that the LLM did not provide a response when asked about a patient's sexual history, a standard question in any medical consultation. Ensuring that ChatGPT does not output material which could be considered offensive under societal norms is a component of alignment [35]. However, our results demonstrate an "alignment tax," in that the model becomes less useful due to constraints imposed by the alignment. The development of LLMs designed specifically for medical education may overcome this issue [36].

Learning efficacy was also a major theme identified in this study. Important subthemes in this category were repetition and convenience. Students noted the benefits of having access to ChatGPT for practice at any time or place and having virtually unlimited ability to repeat the exercises. As mentioned earlier, access to SPs is limited in many medical schools [15]. The ability to augment this shortfall with a virtual SP may be a positive option for many medical students and medical schools.

Interestingly, some students expressed that they experienced considerable anxiety as much as a day before they were scheduled to interact with an SP, although they were aware that the SP was not a real patient. The ability to practice with an LLM such as ChatGPT was seen as beneficial, because students could develop questioning techniques to a point where even during the session with a real SP they could still perform well.

Some differences between preinteraction and postinteraction in terms of subthemes were evident under the major theme of learning efficacy. Before the exercise students were focused more on anticipated or previous experiences in using LLMs for personalized learning materials, for example, developing mnemonics, practice questions, or flash cards. This reflects the experience of other medical students [37]. Responses following the exercise were focused on diagnostic patient interaction skills. This is to be expected as students now had actual experience of ChatGPT in this role and knew that this was to be the focus of our study.

Verisimilitude was a major theme in both preinteraction and postinteraction responses. All students mentioned this as a limiting factor. Absence of facial cues, changes in tone, or body language and an inability to develop rapport were all seen as drawbacks of the virtual SP. Some students also mentioned that this impacted their role as physician. For example, a student physician leaning into the patient to show interest, or other types of body language, was redundant in the exercise. Other studies have also highlighted that the output from ChatGPT cannot replicate the true stimuli a physician relies on in a patient visit [28,31,38,39]. We note that virtual patients are developing rapidly, so issues with verisimilitude may be overcome in future, although it may take some time before ChatGPT, specifically, is able to incorporate a visual or physical layer.

Trust as a theme was evident in preinterview responses but had disappeared in postinterview responses. We note that our faculty team, consisting of clinicians and PhD-qualified members, did not notice any "hallucinations" in output, despite multiple repetitions of the exercise. Yanagita et al [40] recently found that high-quality illness scripts, used for improving medical student's clinical reasoning, could be generated by ChatGPT with relatively few errors. Magalhaes et al [25] also found that a majority of students trusted ChatGPT's output. Nevertheless, even a single error in ChatGPT output, given multiple health care providers may receive the same output, could affect many patients. It is therefore imperative that the veracity of AI output be thoroughly tested before it is fully integrated into health care and medical education settings [28].

Other subthemes for learning efficacy evident postinteraction were feedback and information overload. Our prompt included a direction for ChatGPT to provide feedback on how students could improve their performance. We note that it was necessary to revise the prompt several times during the study, as initially it provided only positive feedback, which did not help in identifying areas for improvement. Responses under the information overload subtheme suggested that students found that the LLM tended to provide more information in regard to

XSL•FO
RenderX

a given question than perhaps a real patient or SP would. This presumably related to the depth of medical knowledge of the LLM but should be considered in further iterations of this exercise. It may be possible to refine the prompt to reduce this effect.

Table 4 compares various platforms able to be used as SPs in medical education, highlighting strengths and limitations across technology, learning efficacy, verisimilitude, and cost. ChatGPT and Claude AI offer affordable, flexible options for unlimited practice but face challenges with uncurated outputs and limited realism in visual and tonal cues. In contrast, platforms such as Body Interact and Soma Lab provide curated cases and interactive features, although often at a higher cost. These findings reinforce that while ChatGPT is a valuable and accessible tool for augmenting SP training, it cannot fully replicate the nuances of human SPs. Addressing limitations such as effective prompt engineering and enhancing realism through improved visual and auditory features could significantly improve its use.

It is possible that the use of ChatGPT as a virtual SP may influence trainees' sensitivity toward patients through the absence of the genuine human interaction students may have with SPs and real patients [41]. The rapid evolution of AI technologies is addressing these gaps to an extent. For instance, the advanced voice mode (AVM) in newer versions of ChatGPT incorporates natural speech patterns and emotional intonations, which may help simulate more realistic interactions. While AI cannot yet replicate the full nuances of real patient encounters, it serves as a consistent and flexible supplementary tool for medical training. Future advancements in AI capabilities may further enhance their ability to foster empathy and connection, thereby reducing potential concerns around decreased sensitivity in trainees.

A number of recent studies have confirmed the use of ChatGPT, or similar LLMs, as virtual SPs [28,29,42]. Similarly to our study, these studies have highlighted ChatGPT's potential to reduce resource constraints and improve accessibility in medical training while offering immersive, flexible practice opportunities. At the same time, limitations created by a lack of verisimilitude were also noted.

Both the necessity and challenges of integrating AI, including LLMs, into medical curricula have also been widely acknowledged [43-47]. Addressing inequities in AI models derived from biased training data is crucial, as these can perpetuate disparities in patient care. Strategies to ensure fairness and equitable outcomes, such as transparency in algorithmic design, have been emphasized in recent studies [45,48]. Additionally, resource allocation, faculty training, and the development of tailored content for medical applications add layers of complexity to curricula integration [46,48]. To move forward, curricula must incorporate foundational AI competencies, including ethical considerations, algorithmic fairness, and practical skills such as prompt engineering. Embedding these competencies into existing core courses, rather than as electives, will ensure comprehensive and equitable learning opportunities [43,44,46,48].

To effectively integrate AI into medical curricula, assessments should be designed to balance the use of AI tools while maintaining the integrity of evaluation processes [44]. Educators should implement secure examination protocols, such as locked-down computers and stricter proctoring, to prevent misuse of AI during assessments. However, assessments can also creatively incorporate AI by engaging students in critiquing AI-generated responses or using these tools to identify knowledge gaps and provide tailored feedback. Generative AI can enhance formative assessments by offering immediate and individualized feedback, guiding students' learning trajectories. We note that our results demonstrate the efficacy of this approach, with the virtual SP providing valuable insights to each student individually on how to improve their patient interactions.

Study Limitations

The small sample size, comprising only 9 participants from a single institution, and potential ascertainment bias, with tech-savvy volunteers possibly skewing results, limited the study's generalizability. This lack of diversity in the sample highlights the need for future studies to include larger and more diverse participant pools to enhance the robustness and generalizability of the findings. Our team is currently working on a multicenter, randomized controlled trial with a mixed methods approach. The study uses a convergent parallel mixed methods design and will span 8 months across multiple medical schools. It will use the new AVM of ChatGPT to simulate an SP. The AVM offers several advantages over the original voice mode, including reduced latency and an ability to inject emotion into its voice [29]. The study aims to draw conclusions based on robust statistical data comparing the average percentage improvement of the experimental group with the control groups on Observed Structured Clinical Examination scores, as well as qualitative data exploring students' learning and perceptions of the AI through focus groups.

Conclusions

This study found ChatGPT to be an effective supplement, although not a full replacement, to traditional SPs. Students and faculty appreciated its potential, noting benefits such as flexible practice times, reduced stress, and improved diagnostic skills. Some shortcomings were noted, including the need for effective prompt engineering and the lack of nonverbal cues affecting realism. Despite these challenges, its reliability and convenience make it a valuable training tool.

Students' diagnostic skills were not formally assessed in this study. However, based on their self-reported perceptions and observations of their interactions with ChatGPT, it appears that the AI can be a valuable tool for practicing clinical reasoning and problem-solving skills. Future research could explore the impact of ChatGPT on students' diagnostic accuracy and clinical performance.

Overall, ChatGPT offers a significant adjunct to traditional SPs, providing accessible, flexible practice opportunities for medical students. The study underscores the importance of integrating prompt engineering into medical curricula and refining AI interactions for balanced information delivery. Continuous

XSL•FO
RenderX

advancements in virtual patient technology and AI capabilities, including improved verbal and auditory flow, are expected to further enhance ChatGPT's use in medical education. Future studies are planned with a larger sample size and using the recently released ChatGPT version 4.o1 with AVM.

Acknowledgments

The authors thank Richard Millis, PhD, for support and guidance during this study. JC was affiliated with the American University of Antigua at the time of the study and is currently affiliated with the Medical University of the Americas. NA was affiliated with the American University of Antigua at the time of the study and is currently affiliated with the School of Medicine, University of Texas. RER and SN were affiliated with the American University of Antigua at the time of the study and are currently unaffiliated.

Authors' Contributions

JC and TK contributed equally to this work. JC conceived the original idea and study design. JC, TK, RER, SD, AV, RH, PH, SN, RS, CL, NA, and JJ refined the study design and conducted the study activity. TK, RS, and RER conducted study participant interviews. AM, AJ, EC, TA, ANN, and JEG contributed to data organization, input, and analysis. JC, TK, and RER contributed to coding, theme construction, and writing original draft of paper. All authors contributed to the review and editing of the paper and approved the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview questions.
[DOCX File, 14 KB - mededu_v11i1e63353_app1.docx]

References

1. Cleland JA, Abe K, Rethans JJ. The use of simulated patients in medical education: AMEE Guide No 42. Med Teach 2009 Jun;31(6):477-486. [doi: 10.1080/01421590903002821] [Medline: 19811162]
2. Flanagan OL, Cummings KM. Standardized patients in medical education: a review of the literature. Cureus 2023 Jul;15(7):e42027. [doi: 10.7759/cureus.42027] [Medline: 37593270]
3. Mühling T, Schreiner V, Appel M, Leutritz T, König S. Comparing virtual reality-based and traditional physical Objective Structured Clinical Examination (OSCE) stations for clinical competency assessments: randomized controlled trial. J Med Internet Res 2025 Jan 10;27:e55066. [doi: 10.2196/55066] [Medline: 39793025]
4. Shankar PR, Dwivedi NR. Using standardized patients for teaching-learning and assessment in a Caribbean medical school. EIMJ 2015 Jun 10;7(2). [doi: 10.5959/eimj.v7i2.358]
5. Mladenovic J, van Zanten M, Pinsky WW. Evolution of educational commission for foreign medical graduates certification in the absence of the USMLE step 2 clinical skills examination. Acad Med 2023 Apr 1;98(4):444-447. [doi: 10.1097/ACM.0000000000005051] [Medline: 36538680]
6. Phanudulkitti C, Puengrung S, Meepong R, Vanderboll K, Farris KB, Vordenberg SE. A systematic review on the use of virtual patient and computer-based simulation for experiential pharmacy education. Explor Res Clin Soc Pharm 2023 Sep;11:100316. [doi: 10.1016/j.rcsop.2023.100316] [Medline: 37635840]
7. Huang G, Reynolds R, Candler C. Virtual patient simulation at US and Canadian medical schools. Acad Med 2007 May;82(5):446-451. [doi: 10.1097/ACM.0b013e31803e8a0a] [Medline: 17457063]
8. Bubeck S, Chandrasekaran V, Eldan R, et al. Sparks of artificial general intelligence: early experiments with GPT-4. arXiv. Preprint posted online on Mar 22, 2023. [doi: 10.48550/arXiv.2303.12712]
9. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023 Feb;2(2):e0000198. [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
10. Temsah O, Khan SA, Chaiah Y, et al. Overview of early ChatGPT's presence in medical literature: insights from a hybrid literature review by ChatGPT and human experts. Cureus 2023 Apr;15(4):e37281. [doi: 10.7759/cureus.37281] [Medline: 37038381]
11. About Oscer. Oscer. 2024. URL: https://www.oscer.ai/about [accessed 2024-06-11]
12. Educational affairs - Office of Clinical Education. The University of Texas Medical Branch. 2024. URL: https://www.utmb.edu/som/som-educational-affairs/office-of-clinical-education-(oce)/clerkship---pediatrics/pedi-clerkship [accessed 2025-05-14]
13. Soma Lab. 2024. URL: https://counselor.somalab.ai/signup [accessed 2024-12-16]

XSL•FO
RenderX

14. Body Interact. URL: https://bodyinteract.com/ [accessed 2024-12-15]
15. Liu X, Wu C, Lai R, et al. ChatGPT: when the artificial intelligence meets standardized patients in clinical training. J Transl Med 2023 Jul 6;21(1):447. [doi: 10.1186/s12967-023-04314-0] [Medline: 37415217]
16. Suárez A, Adanero A, Díaz-Flores García V, Freire Y, Algar J. Using a virtual patient via an artificial intelligence chatbot to develop dental students' diagnostic skills. Int J Environ Res Public Health 2022 Jul 18;19(14):8735. [doi: 10.3390/ijerph19148735] [Medline: 35886584]
17. Weidener L, Fischer M. Artificial intelligence teaching as part of medical education: qualitative analysis of expert interviews. JMIR Med Educ 2023 Apr 24;9:e46428. [doi: 10.2196/46428] [Medline: 36946094]
18. Jowsey T, Stokes-Parish J, Singleton R, Todorovic M. Medical education empowered by generative artificial intelligence large language models. Trends Mol Med 2023 Dec;29(12):971-973. [doi: 10.1016/j.molmed.2023.08.012] [Medline: 37718142]
19. Voice control for ChatGPT x MIA AI. Chrome Web Store. 2024. URL: https://chromewebstore.google.com/detail/voice-control-for-chatgpt/eollffkcakegifhacjnlnegohfdlidhn?hl=en&pli=1 [accessed 2025-05-14]
20. Mayring P. Qualitative content analysis theoretical foundation, basic procedures and software solution. In: Approaches to Qualitative Research in Mathematics Education Advances in Mathematics Education 2015. [doi: 10.1007/978-94-017-9181-6_13]
21. Claude. URL: https://claude.ai/new [accessed 2024-12-15]
22. MMLU—wikipedia. Wikipedia Contributors. 2024. URL: https://en.wikipedia.org/wiki/MMLU [accessed 2024-12-16]
23. Learn clinical reasoning—free sign up. Oscer. URL: https://www.oscer.ai/dashboard/home [accessed 2024-12-15]
24. Kavadella A, Dias da Silva MA, Kaklamanos EG, Stamatopoulos V, Giannakopoulos K. Evaluation of ChatGPT's real-life implementation in undergraduate dental education: mixed methods study. JMIR Med Educ 2024 Jan 31;10(1):e51344. [doi: 10.2196/51344] [Medline: 38111256]
25. Magalhães Araujo S, Cruz-Correia R. Incorporating ChatGPT in medical informatics education: mixed methods study on student perceptions and experiential integration proposals. JMIR Preprints. Preprint posted online on Jul 27, 2023. [doi: 10.2196/preprints.51151]
26. Moldt JA, Festl-Wietek T, Madany Mamlouk A, Nieselt K, Fuhl W, Herrmann-Werner A. Chatbots for future docs: exploring medical students' attitudes and knowledge towards artificial intelligence and medical chatbots. Med Educ Online 2023 Dec;28(1):2182659. [doi: 10.1080/10872981.2023.2182659] [Medline: 36855245]
27. Sarangi PK, Panda BB, P S, Pattanayak D, Panda S, Mondal H. Exploring radiology postgraduate students' engagement with large language models for educational purposes: a study of knowledge, attitudes, and practices. Indian J Radiol Imaging 2025 Jan;35(1):35-42. [doi: 10.1055/s-0044-1788605] [Medline: 39697505]
28. Holderried F, Stegemann-Philipps C, Herschbach L, et al. A Generative Pretrained Transformer (GPT)-powered chatbot as a simulated patient to practice history taking: prospective, mixed methods study. JMIR Med Educ 2024 Jan 16;10(1):e53961. [doi: 10.2196/53961] [Medline: 38227363]
29. Barra FL, Costa A, Rodella G, Semeraro F, Carenzo L. Shaping the future of simulator interactions: the role of ChatGPT's advanced voice mode. Resuscitation 2024 Dec;205:110452. [doi: 10.1016/j.resuscitation.2024.110452] [Medline: 39617251]
30. Gray M, Baird A, Sawyer T, et al. Increasing realism and variety of virtual patient dialogues for prenatal counseling education through a novel application of chatgpt: exploratory observational study. JMIR Preprints. Preprint posted online on Jul 17, 2023. [doi: 10.2196/preprints.50705]
31. Webb JJ. Proof of concept: using ChatGPT to teach emergency physicians how to break bad news. Cureus 2023 May;15(5):e38755. [doi: 10.7759/cureus.38755] [Medline: 37303324]
32. Meskó B. Prompt engineering as an important emerging skill for medical professionals: tutorial. J Med Internet Res 2023 Oct 4;25:e50638. [doi: 10.2196/50638] [Medline: 37792434]
33. Heston T, Khun C. Prompt engineering in medical education. Int Med Educ 2023 Aug 31;2(3):198-205. [doi: 10.3390/ime2030019]
34. Aligning language models to follow instructions. Open AI. 2024. URL: https://openai.com/index/instruction-following [accessed 2025-05-04]
35. Our approach to AI safety. Open AI. 2024. URL: https://openai.com/index/our-approach-to-ai-safety [accessed 2025-05-04]
36. Tu T, Palepu A, Schaekermann M, et al. Towards conversational diagnostic AI. arXiv. Preprint posted online on Jan 10, 2024. [doi: 10.48550/arXiv.2401.05654]
37. Hashemi-Sabet F, Leung FH. Equity-driven MCAT prep: a ChatGPT advantage. Can Med Educ J 2024 May;15(2):105-106. [doi: 10.36834/cmej.78169] [Medline: 38827918]
38. Musallam E, Alhaj Ali A, Alkhafaji M. OpenAI's ChatGPT clinical simulation. Nurse Educ 2024;49(6):E361-E362. [doi: 10.1097/NNE.0000000000001657]
39. Safranek CW, Sidamon-Eristoff AE, Gilson A, Chartash D. The role of large language models in medical education: applications and implications. JMIR Med Educ 2023 Aug 14;9:e50945. [doi: 10.2196/50945] [Medline: 37578830]
40. Yanagita Y, Yokokawa D, Fukuzawa F, Uchida S, Uehara T, Ikusaka M. Expert assessment of ChatGPT's ability to generate illness scripts: an evaluative study. BMC Med Educ 2024 May 15;24(1):536. [doi: 10.1186/s12909-024-05534-8] [Medline: 38750546]

XSL•FO
RenderX

41. Wu Y, Zheng Y, Feng B, Yang Y, Kang K, Zhao A. Embracing ChatGPT for medical education: exploring its impact on doctors and medical students. JMIR Med Educ 2024 Apr 10;10:e52483. [doi: 10.2196/52483] [Medline: 38598263]
42. Jiang Y, Fu X, Wang J, et al. Enhancing medical education with chatbots: a randomized controlled trial on standardized patients for colorectal cancer. BMC Med Educ 2024 Dec 20;24(1):1511. [doi: 10.1186/s12909-024-06530-8] [Medline: 39707245]
43. Hersh W. Artificial Intelligence: Implications for Health Professions Education: AMIA Academic Forum; 2024.
44. Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. Acad Med 2024 Jan 1;99(1):22-27. [doi: 10.1097/ACM.0000000000005439] [Medline: 37651677]
45. Ray PP, Majumder P. The potential of ChatGPT to transform healthcare and address ethical challenges in artificial intelligence-driven medicine. J Clin Neurol 2023 Sep;19(5):509-511. [doi: 10.3988/jcn.2023.0158] [Medline: 37635433]
46. Xu T, Weng H, Liu F, et al. Current status of ChatGPT use in medical education: potentials, challenges, and strategies. J Med Internet Res 2024 Aug 28;26:e57896. [doi: 10.2196/57896] [Medline: 39196640]
47. McCoy LG, Nagaraj S, Morgado F, Harish V, Das S, Celi LA. What do medical students actually need to know about artificial intelligence? NPJ Digit Med 2020;3:86. [doi: 10.1038/s41746-020-0294-7] [Medline: 32577533]
48. Gupta N, Khatri K, Malik Y, et al. Exploring prospects, hurdles, and road ahead for generative artificial intelligence in orthopedic education and training. BMC Med Educ 2024 Dec 28;24(1):1544. [doi: 10.1186/s12909-024-06592-8] [Medline: 39732679]

Abbreviations

AI: artificial intelligence
AUA: American University of Antigua
AVM: advanced voice mode
LLM: large language model
MUA: Medical University of the Americas
SP: standardized patient

Edited by D Chartash; submitted 20.06.24; peer-reviewed by PK Sarangi, R Saifi, SM Araujo; revised version received 18.03.25; accepted 18.03.25; published 20.05.25.

Please cite as:
Cross J, Kayalackakom T, Robinson RE, Vaughans A, Sebastian R, Hood R, Lewis C, Devaraju S, Honnavar P, Naik S, Joseph J, Anand N, Mohammed A, Johnson A, Cohen E, Adeniji T, Nnenna Nnaji A, George JE
Assessing ChatGPT's Capability as a New Age Standardized Patient: Qualitative Study
JMIR Med Educ 2025;11:e63353
URL: https://mededu.jmir.org/2025/1/e63353
doi:10.2196/63353

© Joseph Cross, Tarron Kayalackakom, Raymond E Robinson, Andrea Vaughans, Roopa Sebastian, Ricardo Hood, Courtney Lewis, Sumanth Devaraju, Prasanna Honnavar, Sheetal Naik, Jillwin Joseph, Nikhilesh Anand, Abdalla Mohammed, Asjah Johnson, Eliran Cohen, Teniola Adeniji, Aisling Nnenna Nnaji, Julia Elizabeth George. Originally published in JMIR Medical Education (https://mededu.jmir.org), 20.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Virtual Reality Simulation for Undergraduate Nursing Students for Care of Patients With Infectious Diseases: Mixed Methods Study

Wen Chang[1,2], PhD; Chun-Chih Lin[3,4], PhD; Julia Crilly[5,6], PhD; Hui-Ling Lee[1], MSN; Li-Chin Chen[1,2,4], MSN; Chin-Yen Han[1,4], PhD

[1]Department of Nursing, Chang Gung University of Science and Technology, Taoyuan City, Taiwan
[2]Nursing Management Department of Administration Center, Chang Gung Medical Foundation, Taoyuan City, Taiwan
[3]Department of Nursing, Chang Gung University of Science and Technology, Chiayi County, Taiwan
[4]Department of Nursing, New Taipei Municipal TuCheng Hospital (Built and Operated by Chang Gung Medical Foundation), New Taipei City, Taiwan
[5]Department of Emergency Medicine, Gold Coast Health, Southport, Australia
[6]School of Nursing and Midwifery, Griffith University, Southport, Australia

Corresponding Author:
Chin-Yen Han, PhD
Department of Nursing
Chang Gung University of Science and Technology
261
Wenhua 1st Road, Guishan Dist.
Taoyuan City, 33303
Taiwan
Phone: 886 3 2118999 ext 3417
Fax: 886 3 2118866
Email: cyhan@mail.cgust.edu.tw

Abstract

Background: Virtual reality simulation (VRS) teaching offers nursing students a safe, immersive learning environment with immediate feedback, enhancing learning outcomes. Before the COVID-19 pandemic, nursing students had limited training and opportunities to care for patients in isolation units with infectious diseases. However, the pandemic highlighted the ongoing global priority of providing care for patients with infectious diseases.

Objective: This study aims to (1) examine the effectiveness of VRS in preparing nursing students to care for patients with infectious diseases by assessing its impact on their theoretical knowledge, learning motivation, and attitudes; and (2) evaluate their experiences with VRS.

Methods: This 2-phased mixed methods study recruited third-year undergraduate nursing students enrolled in the Integrated Emergency and Critical Care course at a university in Taiwan. Phase 1 used a quasi-experimental design to address objective 1 by comparing the learning outcomes of students in the VRS teaching program (experimental group) with those in the traditional teaching program (control group). Tools included an infection control written test, the Instructional Materials Motivation Survey, and a learning attitude questionnaire. The experimental group participated in a VRS lesson titled “Caring for a Patient with COVID-19 in the Negative Pressure Unit” as part of the infection control unit. In phase 2, semistructured interviews were conducted to address objective 2, exploring students’ learning experiences.

Results: A total of 107 students participated in phase 1, and 18 students participated in phase 2. Both the VRS and control groups showed significant improvements in theoretical knowledge scores (for the VRS group t_{46}=–7.47; P<.001, for the control group t_{59}=–4.04; P<.001). However, compared with the control group, the VRS group achieved significantly higher theoretical knowledge scores ($t_{98.13}$=2.70; P=.008) and greater learning attention (t_{105}=2.30; P=.02) at T1. Additionally, the VRS group demonstrated a statistically significant higher regression coefficient for learning confidence compared with the control group (β=.29; P=.03). The students’ learning experiences in the VRS group were categorized into 4 themes: Applying Professional Knowledge to Patient Care, Enhancing Infection Control Skills, Demonstrating Patient Care Confidence, and Engaging in Real Clinical Cases. The core theme identified was Strengthening Clinical Patient Care Competencies.

Conclusions: The findings suggest that VRS teaching significantly enhanced undergraduate nursing students’ infection control knowledge, learning attention, and confidence. Qualitative insights reinforced the quantitative results, highlighting the holistic

XSL•FO
RenderX

benefits of VRS teaching in nursing education, including improved learning outcomes. The positive impact on student motivation and attitudes indicates a potentially transformative approach to nursing education, particularly in the post–COVID-19 era, where digital and remote learning tools play an increasingly vital role.

(JMIR Med Educ 2025;11:e64780) doi:10.2196/64780

KEYWORDS

virtual reality; infection control; learning motivation; learning attitudes; nursing education

Introduction

Background

Virtual reality (VR) has emerged as an innovative teaching strategy in nursing education. VR technology leverages simulated scenarios to overcome time and space limitations, offering students opportunities to learn in safe, realistic settings and receive immediate feedback [1]. VR simulation (VRS) teaching strategies enhance learning motivation, student immersion, knowledge and skill acquisition, confidence [2,3], active participation, and learning effectiveness [4-6]. The goal of undergraduate nursing education is to prepare students for clinical practice, making it essential to strengthen their professional competencies and attitudes. Integrating information technology into nursing education enhances students' learning outcomes. Nursing education should align with the broader clinical practice environment, incorporating technology to support students in developing their competencies [7].

The COVID-19 pandemic has profoundly impacted nursing curricula and teaching worldwide. In emergency and critical care, university-level nursing curricula must reflect clinical environments. Emphasizing situated learning enhances students' abilities and confidence in providing emergency patient care [8,9]. Before the pandemic, nursing students rarely had opportunities to care for patients with infectious diseases in isolation units. However, the demand for care related to infectious diseases remains a global priority [9]. Strengthening courses on infectious diseases can help students develop positive attitudes toward clinical practice [10]. Updating infectious disease courses with more practical experiences can further support nursing students in developing positive attitudes when caring for patients with infectious diseases during clinical practice.

Learning theories related to VR teaching include constructivism, situated learning, and experiential learning. In VR learning, learners actively absorb information and construct new knowledge [11]. Situated learning theory emphasizes real-world interactions and activities in authentic contexts, transforming these experiences into applicable knowledge [12]. VR offers an interactive virtual environment, using visual effects to present abstract problems and providing opportunities for active manipulation and repeated practice [11]. Experiential learning theory posits that learning is the transformation of experience, with knowledge creation emerging from interactions, conflicts, and problem-solving between individuals and their environment. This theory highlights the potential of immersive technology to provide meaningful experiences [12]. Compared with other teaching methods, VR teaching is easy to use, generates positive and active learning experiences [13], and enhances learning outcomes, including improvements in knowledge, skills, and clinical decision-making [14,15]. Engagement in VR environments provides students with experiences closely aligned with clinical practice, boosting their motivation and attitudes and leading to better educational outcomes [14,15].

Motivation and attitude play a significant role in influencing learning outcomes. Enhanced motivation strengthens active learning and improves results [12,16]. Studies have shown a positive correlation between motivation and learning outcomes, making learning easier and fostering proactive engagement [17,18]. Keller's Attention, Relevance, Confidence, and Satisfaction (ARCS) model of motivation incorporates a learning motivation scale to assess motivational aspects within a course [12,16]. Designing courses with integrated motivational models can inspire learners, enhance motivation, and increase classroom engagement [19,20]. In nursing education, particularly in emergency and critical care courses, VR can address the limitations of clinical settings and traditional teaching methods caused by resource constraints [21-24]. VR stimulates learners' motivation, promotes active participation, and enhances learning outcomes [25,26]. By incorporating VR teaching, courses can more closely align with clinical practice, providing students with a solid foundation in professional knowledge and skills.

Even long after the pandemic, there will remain a global need for the care of patients with infectious diseases. However, opportunities for students to participate in the actual care of such patients in isolation units remain limited. To date, little attention has been given in the literature to identifying educational strategies that address this gap in developing nursing students' professional knowledge and skills. This mixed methods study was guided by 2 research questions: (1) What is the effectiveness of VRS teaching on nursing students' theoretical knowledge, learning motivation, and attitudes toward the care of patients with infectious diseases? and (2) What are the learning experiences of nursing students in a VRS program? Our a priori hypothesis was that VRS teaching would significantly improve nursing students' infection control theoretical knowledge, learning motivation, and attitudes regarding the care of patients with infectious diseases.

Objectives

This study had 2 objectives: (1) to evaluate the effectiveness of VRS teaching on nursing students' theoretical knowledge, learning motivation, and attitudes toward the care of patients with infectious diseases, and (2) to explore their learning experiences in a VRS program designed for this target population.

XSL•FO
RenderX

Methods

Study Design

This study used a 2-phased mixed methods approach to comprehensively evaluate a VRS teaching program on the care of patients with infectious diseases, which was part of the infection control unit within the Integrated Emergency and Critical Care course. Phase 1 utilized a quantitative study design to assess the learning effectiveness of the VRS teaching method, while phase 2 used qualitative phenomenography to explore students' experiences and perceptions of the program.

Phase 1: Outcomes of the VRS Program on Students' Knowledge, Learning Motivation, and Attitudes to the Care of Patients With Infectious Diseases

Overview

A quasi-experimental design was used to compare learning outcomes—knowledge, motivation, and attitude—between students in the VRS teaching program (experimental group) and those in the traditional teaching course (control group). Data were collected from August 2022 to July 2023.

Participants

This study used convenience sampling and was conducted at a clinical competence center at a university in Taiwan. Third-year undergraduate nursing students enrolled in the Integrated Emergency and Critical Care course were eligible to participate. One class of students was assigned to the experimental group, and another to the control group. The Integrated Emergency and Critical Care course is an elective offered in both the first and second semesters. Researchers used a random selection process to assign students in the infection control unit to the VRS program in the first semester and to traditional teaching in the second semester. The participating school provided a 2-week add/drop period, during which members of the research team gave in-class briefings about the study, and students were free to choose whether to participate in the experimental group. The selection criteria for the experimental group were (1) aged ≥20 years, (2) enrolled in the Integrated Emergency and Critical Care course, and (3) willing to participate in this study. Students with a history of epilepsy were excluded from the VRS. Sample size estimation was conducted using G*Power software version 3.1 [27]. Following Cohen's rule [28], 2 groups were included, with a medium effect size of f=0.25, a correlation of 0.5, a power of 0.8, and an α value of .05, resulting in a required sample size of ≥86, with ≥43 participants per group. A total of 47 students were recruited for the experimental group. None of the students in the experimental group refused to participate in the VRS program. All participants were taught by the same instructor, and the course content was consistent across both groups.

Instruments

Infection Control Written Test

Previous research has shown that VRS teaching can enhance the development of both knowledge and practical skills in undergraduate nursing students, with outcomes effectively assessed using a written test [29]. In this study, the infection control knowledge assessment involved a written test administered to students before (T0) and after (T1) the infection control lesson. The test consisted of 10 questions aligned with the learning objectives of the infection control unit. These included single- and multiple-choice items covering both theoretical knowledge and practical skills, such as donning and doffing personal protective equipment (PPE). The test addressed the same key infection control techniques for all students, aiming to evaluate their baseline abilities and the changes in knowledge following the lesson. To better capture postlearning changes and minimize the influence of memory recall on the posttest results, the order of the questions was adjusted, and some questions were modified. The test items were reviewed by the course instructor and clinical experts (senior emergency nurses) to ensure content validity.

Instructional Materials Motivation Survey

The Instructional Materials Motivation Survey (IMMS) is a self-reported questionnaire administered before (T0) and after (T1) the infection control lesson. Designed primarily to evaluate students' motivation in learning a course [12], the IMMS comprises 36 items distributed across 4 subscales based on the ARCS motivation model: Attention (12 items), Relevance (9 items), Confidence (9 items), and Satisfaction (6 items). Each item is rated on a 5-point Likert scale, with higher scores indicating greater learning motivation. The original IMMS scale has demonstrated high reliability, with Cronbach α values ranging from 0.81 to 0.96 [12]. In this study, the IMMS exhibited excellent reliability, with a Cronbach α of 0.94.

Learning Attitude Questionnaire

A learning attitude questionnaire was administered at T0 and T1. This 20-item self-reported questionnaire was developed by several members of the research team to assess students' attitudes toward caring for patients with infectious diseases and their participation in the infection control unit. Items were rated on a 5-point Likert scale, where 1 indicates "strongly disagree," 2 "disagree," 3 "neutral," 4 "agree," and 5 "strongly agree." Higher scores reflected a more positive learning attitude. The questionnaire demonstrated strong validity and reliability, with an average Content Validity Index of 0.9 and a Cronbach α of 0.955.

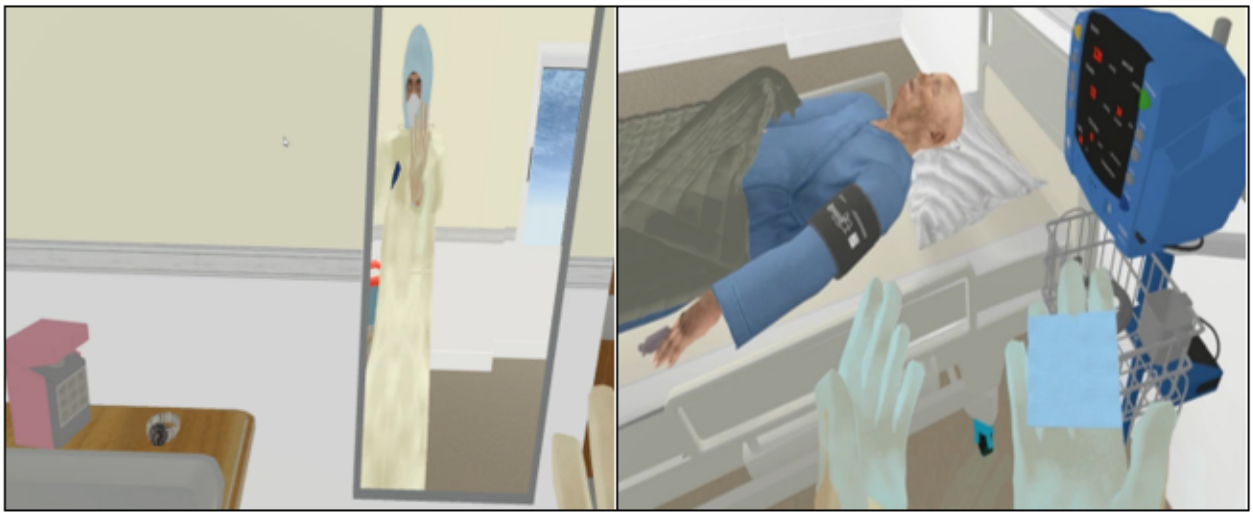
VRS Lesson Plan for Caring for a Patient With COVID-19 in the Negative Pressure Unit

In the infection control unit, VRS teaching was implemented for the experimental group. The VRS scenario, developed by several research team members with VR training certification, depicted a real clinical case of a febrile patient visiting the emergency department for triage, later confirmed to have COVID-19, and subsequently admitted to a negative pressure isolation unit (Figure 1). The teaching content emphasized a nurse's role in providing care within a negative pressure isolation room, including the proper techniques for donning and doffing PPE. The lesson's learning objectives were for students to differentiate care for patients with infectious diseases, correctly don and doff PPE, and provide appropriate patient care. The VR Oculus Quest equipment, including a headset and controllers, was supplied by the School of Nursing of the participating university. The VRS lesson plan was reviewed by the course instructor and clinical experts (senior emergency

XSL•FO
RenderX

nurses) to ensure content validity. The lesson utilized VR technology to deliver immersive visual effects and interactive scenarios, aiming to enhance students' awareness and provide opportunities for practical exercises, thereby improving learning outcomes [30]. The assessment included the standard procedure for applying PPE, such as N95 masks, goggles, hair caps, and gloves. Upon completing the assessment, students received immediate feedback, with the computer screen highlighting missed items. This direct feedback was intended to reinforce learning effectiveness [1].

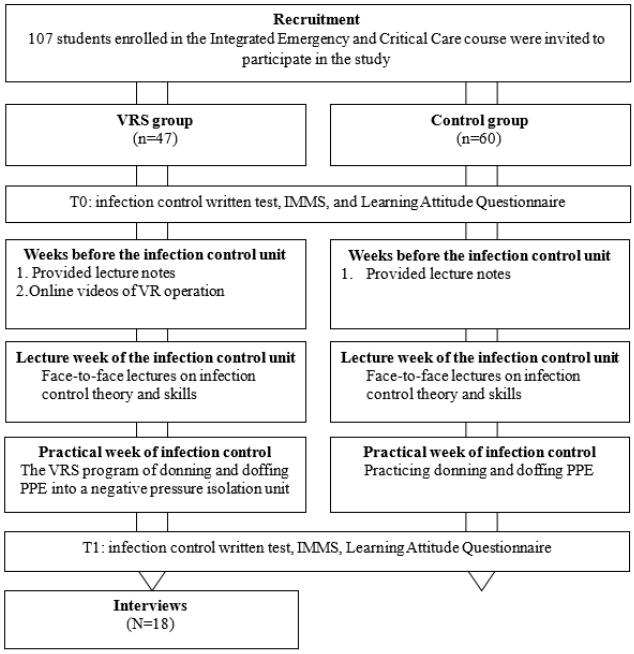
Figure 1. Screen captures of virtual reality simulation videos.



Procedure

To maintain neutrality, none of the research team members were involved in teaching either the experimental or control group. Instead, several team members focused on designing the VRS program and creating a VR system operation video to help students become familiar with operating the VR system. During the experimental group's class, research team members were available to address any technical issues that participants encountered. They also met with the unit instructors before the start of the unit to ensure consistency in teaching between the 2 groups and alignment in the course delivery process. The study protocol is illustrated in Figure 2.

Figure 2. Study protocol process. IMMS: Instructional Materials Motivation Survey; PPE: personal protective equipment; VR: virtual reality; VRS: virtual reality simulation.



The infection control unit spanned 2 weeks and included 1 lecture (100 minutes) and 1 practical session (100 minutes). Both the VRS group and the control group received lecture notes before the start of the course. Additionally, the VRS group was provided with prerecorded online VR videos demonstrating donning and doffing PPE in a negative pressure isolation unit, as well as instructions on operating the VR system. For the VRS group, the first week consisted of a 100-minute lesson on

XSL•FO
RenderX

infection control theory and skills, while the second week involved 100 minutes of VRS instruction. The VRS class was divided into 5 groups, each comprising 8-10 students, who worked collaboratively on the drills. The lesson began with an introduction to VR system operation (5-10 minutes), followed by group VRS scenario drills (30-40 minutes). Each student then executed their part of the VRS lesson, which lasted approximately 6-8 minutes. Upon completion, the system provided feedback, serving as the students' learning outcomes. Group members first discussed the session among themselves, followed by a 10-minute instructor-led debriefing session. During this session, students were encouraged to ask questions and share their reflections on the VRS program execution. Feedback and reflection were incorporated to help students consolidate their learning and transform it into meaningful learning outcomes. The groups then switched roles and conducted a second round of drills and discussions for another 30-40 minutes. A pretest (T0) and posttest (T1) on infection scenario cases were conducted to evaluate the students' learning outcomes, motivation, attitudes, and knowledge related to the infection control unit. For the control group, a traditional teaching strategy was used. During the first week, theoretical lectures were delivered, accompanied by a video to aid students in understanding the process. Lecture notes, identical to those provided to the VRS group, were distributed before class and included a video link demonstrating the standard PPE procedure. In the second week, students were divided into 6 groups of 9-11 members to practice donning and doffing PPE. Instructors provided individualized guidance to correct mistakes. Within the same groups, students evaluated and discussed the PPE practice. Although an instructor-led debriefing session was planned for the control group, it was postponed to the following week due to the large number of students and time constraints.

Data collected for this study were individually coded and entered into a computer for analysis using SPSS version 22.0 (IBM Corp.). Descriptive statistics, including frequency, percentage, mean, and SD, were calculated. Inferential statistics, such as independent *t* tests, paired *t* tests, and generalized estimating equations, were applied. Results with a *P* value of <.05 were considered statistically significant.

Phase 2: The Students' Learning Experiences of the VRS in Caring for Patients With Infectious Diseases

Overview

Phase 2 utilized qualitative phenomenography to explore students' experiences and perceptions of the VRS program. The key concepts in phenomenography are "phenomenon" and "experience." This methodology aims to identify the shared and generalized aspects of participants' thoughts or concepts regarding their experiences of a specific phenomenon, with a focus on describing their understanding of these experiences [31]. In this study, phenomenography was applied to understand how learners organize and structure the content they acquire during the learning process [31]. Interviews with students were analyzed to uncover their learning experiences and outcomes, with the goal of providing evidence to support the ongoing improvement of educational programs.

Participants

Students in the experimental group who met the following inclusion criteria were recruited: (1) aged 20 years or older, (2) enrolled in the Emergency and Critical Care course and participating in VRS teaching, and (3) consenting to participate in and have interviews recorded.

Procedure

Participants took part in in-depth semistructured interviews. These interviews facilitate meaningful conversations, providing a detailed understanding of complex issues [31]. In a phenomenographic study, interview questions need to be as open-ended as possible to accurately capture the participants' thoughts. The interview guide is provided in Multimedia Appendix 1. The in-depth interviews in this study contributed to understanding nursing students' experiences with VRS teaching. Each eligible participant received a consent form outlining the study's purpose, the voluntary nature of participation, and the confidentiality of their data. Participants completed the consent forms, and suitable interview times were scheduled. The interviews were conducted by a single researcher (CYH) who had prior experience in qualitative research gained during doctoral studies, had served as a principal investigator on research projects, and had published several qualitative research articles. CYH was not involved in teaching this subject and was unfamiliar with the participants in the experimental group. During the interviews, CYH utilized interview skills to encourage participants to articulate their VR learning experiences. The interviews were audio-recorded and lasted between 42 and 62 minutes. A sample size of 18 students was sufficient to generate rich data and achieve saturation.

Qualitative Analysis and Trustworthiness

Following each interview, the same researcher (CYH) transcribed the audio recordings verbatim to ensure detailed documentation and analyzed the interview data. Data analysis followed the 7 steps of phenomenographic analysis: familiarization, condensation, comparison, grouping, articulating, labeling, and contrasting [32]. The trustworthiness of the research findings was established using Lincoln and Guba's [33] criteria of credibility, transferability, dependability, and confirmability. In terms of credibility, phenomenographic research emphasizes the precise description of each stage of the study process, the application of the researcher's ideas to the phenomena, the careful formulation of interview questions and processes, and the thorough analysis and presentation of conclusions. Peer debriefing, which involves collaborative data analysis to explore diverse interpretations, enhances data interpretation and credibility, contributing to the development of credible research outcomes. Transferability is supported by providing in-depth data that represent a comprehensive view of the research, highlighting its relevance and context. Dependability is ensured by supporting categorizations with excerpted interview content, illustrating the similarities and differences among participants in relation to the phenomenon, and confirming the logical connection between the collected data and the phenomena captured by the descriptive categorization. Confirmability is established by documenting the interviewer's feelings and thoughts during the interview

XSL•FO
RenderX

process, thereby creating an audit trail. The data analysis is thoroughly described, with detailed records of decisions made and strategies adopted during concept formation. These reflections on theoretical and methodological aspects further contribute to the audit trail and the confirmability of the findings [34].

Ethical Consideration

Ethical approval was obtained from the Institutional Review Board of Chang Gung Medical Foundation (approval number 202002386B0). Potential participants were fully informed about the nature and purpose of the study, emphasizing that participation was entirely voluntary and that they had the right to withdraw from the study at any time. They were explicitly assured that their academic results would not be affected by their decision to participate or not. Participants were also guaranteed that their data would remain confidential and that they would not be identifiable in any reports. All participants provided written informed consent, and none of the students withdrew from the study.

Results

Phase 1: Outcomes of the VRS Program on Students' Knowledge, Learning Motivation, and Attitudes to the Care of Patients With Infectious Diseases

Overview

Participants in phase 1 consisted of 107 third-year undergraduate students: 47 in the experimental group, who received VRS teaching, and 60 in the control group, who participated in traditional practical sessions on donning and doffing isolation gowns. As shown in Table 1, the majority of participants (99/107, 92.5%) were female, with an average age of 21.14 (SD 0.69) years.

Table 1. Demographic characteristics of participants in phase 1 (N=107).

Participant demographics	Study groups		Total participants (N=107)
	Virtual reality simulation (n=47)	Control (n=60)	
Gender, n (%)			
Male	2 (4.3)	6 (10)	8 (7.5)
Female	45 (95.7)	54 (90)	99 (92.5)
Age (years), mean (SD)	N/A[a]	N/A	21.14 (0.69)
Male	22.00 (1.41)	20.50 (0.55)	20.88 (0.99)
Female	21.40 (0.54)	20.96 (0.70)	21.16 (0.67)

[a]N/A: not applicable.

Effectiveness of VRS in Infection Control Theoretical Knowledge

Pre- and posttest assessments of infection control knowledge were conducted at T0 and T1, with a maximum score of 10 points. The combined results of all participants showed an average pretest knowledge score of 7.58 (SD 1.13) and a posttest knowledge score of 8.58 (SD 1.16), indicating improved knowledge after course completion (t_{106}=–7.08; P<.001). Both the VRS and control groups demonstrated significant improvements in theoretical knowledge scores (for the VRS group t_{46}=–.747; P<.001 and for the control group t_{58}=–4.04; P<.001). However, the VRS group achieved significantly higher posttest scores compared with the control group ($t_{98.13}$=2.70; P=.008), suggesting that VRS teaching was more effective in enhancing students' knowledge (Table 2 and Figure 3).

XSL•FO
RenderX

Figure 3. Change in infection control knowledge in the two groups. T0: pretest; T1: posttest; **: *P*<.01.

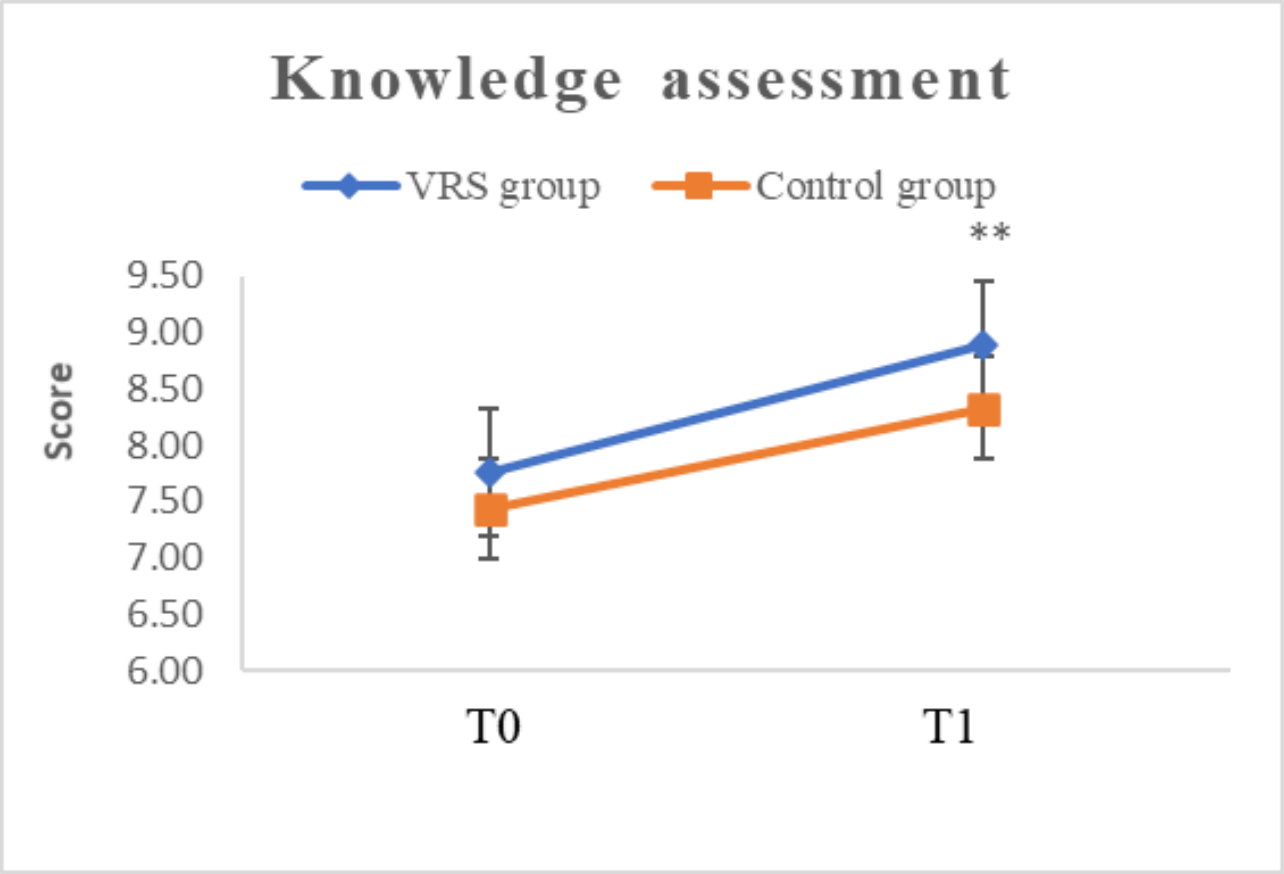


Table 2. Comparison of theoretical knowledge scores in the 2 groups (N=107).

Variable	T0[a], mean (SD)	T1[b], mean (SD)	*t* test (*df*)	*P* value
Knowledge assessment				
Virtual reality simulation	7.77 (1.05)	8.89 (0.79)	–7.47 (46)	<.001
Control	7.43 (1.18)	8.33 (1.34)	–4.04 (58)	<.001
t test (*df*)	1.52 (105)	2.70 (98.13)	N/A[c]	N/A
P value	.13	.008[d]	N/A	N/A
Total	7.58 (1.13)	8.58 (1.16)	–7.08 (106)	<.001

[c]N/A: not applicable.

Effectiveness of VRS on Learning Motivation

The learning motivation of all students increased slightly from T0 (mean 3.84, SD 0.47) to T1 (mean 3.94, SD 0.40) (t_{106}=–3.10; *P*=.002), with no significant differences between the groups at either T0 ($t_{76.50}$=0.09; *P*=.93) or T1 ($t_{80.95}$=1.43; *P*=.16). At T0, except for the Confidence dimension—which was lower in the VRS group compared with the control group ($t_{78.53}$=–2.12; *P*=.04)—the other dimensions of the ARCS model (ie, Attention, Relevance, and Satisfaction) did not differ significantly between the groups. At T1, only the Attention dimension differed significantly between the VRS and control groups, being higher in the VRS group (t_{105}=2.30; *P*=.02), as shown in Table 3.

XSL•FO
RenderX

Table 3. Comparison of learning outcomes between the 2 groups at different time points (N=107).

Variable and groups	T0[a]			T1[b]		
	Mean (SD)	*t* test (*df*)	*P* value	Mean (SD)	*t* test (*df*)	*P* value
Motivation		0.09 (76.50)	.93		1.43 (80.95)	.16
Virtual reality simulation	3.84 (0.57)			4.01 (0.46)		
Control	3.83 (0.38)			3.89 (0.34)		
Attention		0.91 (105)	.36		2.30 (105)	.02[c]
Virtual reality simulation	3.84 (0.67)			4.03 (0.52)		
Control	3.74 (0.48)			3.82 (0.42)		
Relevance		0.76 (82.99)	.45		0.88 (105)	.38
Virtual reality simulation	4.15 (0.54)			4.21 (0.48)		
Control	4.07 (0.41)			4.14 (0.37)		
Confidence		–2.12 (78.53)	.04[c]		0.86 (105)	.39
Virtual reality simulation	3.45 (0.58)			3.65 (0.49)		
Control	3.66 (0.40)			3.58 (0.43)		
Satisfaction		0.51 (78.30)	.61		0.48 (105)	.63
Virtual reality simulation	3.98 (0.64)			4.18 (0.51)		
Control	3.92 (0.44)			4.13 (0.50)		
Attitude		–0.60 (105)	.55		0.03 (105)	.98
Virtual reality simulation	4.09 (0.57)			4.34 (0.52)		
Control	4.15 (0.50)			4.34 (0.56)		

[a]T0: pretest.

[b]T1: posttest.

[c]$P<.05$.

To address potential bias stemming from the differences in the Confidence dimension between the VRS and control groups at T0, the Generalized Estimating Equations model was applied to analyze and compare changes in both groups throughout the study period and to evaluate the outcomes of the VRS intervention. For the VRS group, the regression coefficients for the Confidence dimension were significant (β=.29; P=.03), with positive parameter estimates compared with the control group. This finding indicates that the VRS intervention enhanced students' learning confidence.

Effectiveness of VRS on Learning Attitude

The learning attitude score increased slightly in the VRS group from T0 (mean 4.09, SD 0.57) to T1 (mean 4.34, SD 0.52) and in the control group from T0 (mean 4.15, SD 0.50) to T1 (mean 4.34, SD 0.56). However, no significant differences were observed between the groups at either T0 or T1, as shown in Table 3.

Phase 2: The Students' Learning Experiences of the VRS in Caring for Patients With Infectious Diseases

Overview

In phase 2 of this study, 18 students from the VRS group who had expressed willingness to be interviewed were recruited for qualitative interviews. All interview participants were female. Data analysis followed the phenomenographic steps of familiarization, condensation, comparison, grouping, articulating, labeling, and contrasting [32]. Each theme elicited from the participants' pool of meaning represented a concept of their learning experiences associated with engaging in the VRS program. The core theme captured the relationship between each theme and participants' overall understanding of their VRS learning experiences. The students' learning experiences were categorized into 4 themes: (1) Application of Professional Knowledge to Patient Care, (2) Enhanced Infection Control Skills, (3) Demonstrated Confidence in Patient Care, and (4) Participation in Real Clinical Cases. The core theme was identified as Strengthening Clinical Patient Care Competencies.

Theme 1: Application of Professional Knowledge to Patient Care

The students described how they applied their theoretical knowledge of infection control during the VRS teaching process, particularly regarding the various factors that need to be considered when entering and exiting the negative pressure isolation unit. Through the feedback and debriefing provided by the VRS teaching, they were able to reflect on the content of their infection control learning, thereby deepening their professional knowledge in this area.

XSL•FO
RenderX

Theme 2: Enhanced Infection Control Skills

The participants shared their experiences of using VRS to practice the care skills learned in the infection control unit. Through hands-on practice and observing their classmates, they noted improvements in their skills. They also reported that the course's practical and interactive scenarios enhanced their learning interest, which translated into positive learning outcomes.

Theme 3: Participation in Real Clinical Cases

Most of the students in the integrated care course for Emergency and Critical Care expressed a desire to intern or work in emergency departments or intensive care units. However, during the pandemic, many hospitals' critical care units halted the acceptance of nursing interns or prevented students from participating in the care of patients with infectious diseases during their ward internships. The participants reported that the realistic, scenario-based case studies in the VRS enabled them to practice clinical skills that would otherwise have been unavailable, thereby bridging the gap between theory and practice. The experience of providing care for simulated patients in a context closely resembling clinical settings enhanced their learning experience.

Theme 4: Demonstrated Confidence in Patient Care

During the interviews, participants shared that they had been concerned about their ability to provide effective clinical care for patients with infectious diseases due to the impact of the pandemic. However, after participating in the VRS course, they reported a boost in their confidence in providing this type of care.

Core Theme: Strengthening Clinical Patient Care Competencies

The core theme that emerged from the analysis of the qualitative data on students' learning experiences in the VRS teaching program was the strengthening of their clinical care competencies. The VRS learning program allowed students to apply the professional knowledge and skills they had learned in the course to carefully design patient scenarios. Through VRS practical exercises, students improved the skills required to care for patients with infectious diseases. By engaging with clinical cases and performing learning tasks in a realistic setting, they gained greater confidence in caring for patients with infectious diseases. In other words, this approach of connecting learning experiences enhanced their clinical care competence, better preparing them for the future care of patients with infectious diseases.

Discussion

Principal Findings

The results of this study showed significant improvements in infection control knowledge scores in both groups, with the VRS group achieving higher scores, highlighting the effectiveness of VRS teaching in enhancing theoretical knowledge. The VRS group also achieved a higher attention score at T1 compared with the control group. Additionally, the VRS intervention enhanced students' learning confidence. Students' reflections on their learning experiences and perceptions of the VRS teaching emphasized the following themes: Application of Professional Knowledge to Patient Care, Enhanced Infection Control Skills, Demonstrated Confidence in Patient Care, Participation in Real Clinical Cases, and Strengthening Clinical Patient Care Competencies.

This study made every effort to control variables to ensure consistency in learning content and teaching quality between the 2 groups, including the use of the same teaching materials and the same instructor for both groups. The VRS provided immediate feedback, allowing students to actively engage in the care process within a VR patient scenario, which contributed to enhanced learning outcomes in the VRS group. However, due to the larger number of students in the control group, their debriefing session was delayed. Future studies could investigate the impact of debriefing timing on the effectiveness of VRS-based teaching.

Comparison With Prior Work: Effectiveness of VRS in Infection Control Knowledge

The results of this study demonstrated significant improvement in infection control knowledge in both groups after the learning process. The VRS group achieved significantly higher scores on the infection control written test compared with the control group at T1, indicating that VRS teaching was more effective in enhancing students' theoretical knowledge. This finding aligns with previous research. Systematic reviews and meta-analyses on VR in nursing education have demonstrated its effectiveness in improving knowledge [5,35]. Another review reported a moderate effect size (g=0.47) for VR teaching in knowledge acquisition [29]. Additionally, this review noted that subgroup analysis showed VR training involving multiple self-practice sessions of less than 30 minutes was effective in imparting procedural knowledge to undergraduate nursing students [29]. This finding is consistent with the results of our study, where each student engaged in VRS learning for 6-8 minutes, with the option for continued practice for those wishing to further develop their skills. An integrative review also concluded that VRS teaching is effective in enhancing the acquisition of clinical skills and knowledge [36]. An extensive review of 29 randomized controlled trials involving 2722 students found that VR, augmented reality, and mixed reality were as effective as traditional methods in enhancing knowledge, highlighting their potential role in preclinical education [37]. Similarly, a German study on teaching tracheal suction skills observed no statistical differences among various teaching methods in terms of knowledge and skill improvement, suggesting that VR can serve as a supplementary resource to existing learning strategies, supporting students in preparing for clinical practice [23].

Impact of VRS on Learning Motivation and Attitude

A systematic review and meta-analysis of 26 studies found no significant impact of VR on nursing students' motivation and cognitive load compared with traditional teaching methods [38]. This aligns with the findings of our study, which also showed no significant difference in learning motivation. However, other studies have reported higher motivation and satisfaction with VR, though it may also increase cognitive load [39].

XSL•FO
RenderX

Additionally, several studies have shown that VR positively impacts learners by enhancing attention and motivation, building self-efficacy, and reinforcing learning confidence and performance [5,6,21,25]. A Taiwanese study comparing traditional and VR teaching on nasogastric tube feeding found nonsignificant higher scores in the VR group, which demonstrated greater motivation and satisfaction but also experienced a higher cognitive load [39]. These findings highlight the need to carefully consider cognitive load in future course designs [39].

A South Korean study reported higher neonatal resuscitation knowledge, motivation, problem-solving skills, and confidence in the VR group compared with the control groups, along with lower anxiety levels [26]. An integrative review on VR teaching for emergency patients also revealed increased confidence in handling emergencies [25]. Although some studies have reported no significant differences in anxiety and confidence [3,40], further research is needed to determine VR's impact on learning confidence and stress. A Chinese study on disaster nursing courses found significant improvements in preparedness, confidence, and performance in the experimental group, highlighting VR's potential as a cost-effective simulation method [41]. Technical issues with VR were noted as disadvantages, which may explain the lower precourse confidence in the VRS group compared with the control group. Ensuring that students are familiar with the VR system before the course begins may help improve their confidence [1,27,42,43].

Student's Learning Experiences of VRS

Analysis of the qualitative data obtained in this study revealed 4 themes in students' experiences and perceptions of VR learning: Application of Professional Knowledge to Patient Care, Enhanced Infection Control Skills, Demonstrated Confidence in Patient Care, and Participation in Real Clinical Cases. The core theme identified was the Strengthening Clinical Patient Care Competencies. Similarly, previous qualitative studies have demonstrated a positive impact of VR learning on knowledge [23,24,43], skills [21,23,24], confidence [23], and engagement [4]. Some of these studies focused on VR learning as a tool, while others examined the characteristics of the VR environment [21,23,43]. By contrast, our study applied a phenomenographic methodology to explore students' experiences and perceptions of VR learning, linking these to various outcomes. As a result, our findings provide unique insights into students' conceptions of VR learning.

Strengths and Limitations

The strength of this study lies in its combination of quantitative and qualitative methods, providing a comprehensive understanding of the effectiveness of VRS teaching. However, some potential limitations and weaknesses should be considered. Although the sample size was adequate, it may not fully represent the diversity of nursing students, as it was drawn from only 1 university. The course in which the VRS program was applied was an elective unit offered in both the first and second semesters, with a maximum enrollment of 60 students per class. The number of students enrolled in each class was beyond the researchers' control, resulting in an imbalance in the number of students between the 2 groups. It is recommended that future studies compare the effectiveness of VRS using groups of equal size and a larger number of participants. This study did not conduct a formal survey on potential side effects among students in the experimental group. However, during the VRS session, research team members and the instructor periodically checked in with the students. None of the students reported any discomfort that required them to pause or stop the activity. The control group engaged in practical exercises for donning and doffing PPE, which differs from the traditional nursing classroom teaching methods used in previous studies. Future research is needed to build on our findings and develop a more detailed understanding of the effectiveness of VRS programs.

Conclusions

This study highlights the effectiveness of VRS teaching in enhancing infection control knowledge, learning motivation, attitudes, and course satisfaction among undergraduate nursing students. By combining insights from qualitative data with quantitative information, we have provided a holistic understanding of the potential role of VRS in nursing education. Despite its limitations, this study opens avenues for future research and presents a compelling case for the broader implementation of VR in nursing education curricula. Future studies should consider longitudinal designs to evaluate the long-term impacts of VRS teaching on nursing education. Additionally, expanding the participant pool to include a more diverse range of students could yield more generalizable results. The findings have significant implications for nursing education, suggesting that VRS teaching can effectively enhance learning outcomes, particularly in areas that require high levels of practical knowledge and skills. The positive impact on student motivation and attitudes also points to a potentially transformative shift in how nursing education can be delivered, especially in a post–COVID-19 era, where digital and remote learning tools are becoming increasingly important.

Use of Generative Artificial Intelligence

During the preparation of this work, the authors used ChatGPT (OpenAI) to enhance the clarity of the content. After using ChatGPT, the authors reviewed and edited the content as needed and took full responsibility for the content of the published article.

Acknowledgments

This research was funded by the National Science and Technology Council (grants MOST 110-2511-H-255-013 and MOST 111-2410-H-255-010). The authors thank those students who participated in this research.

XSL•FO
RenderX

Data Availability

The data that support the findings of this study are available on reasonable request from the corresponding author.

Authors' Contributions

The conception and design of the study were carried out by WC, CCL, and CYH, while data acquisition and collection were performed by WC and CYH. The analysis and interpretation of data involved WC, CCL, LCC, and CYH. WC, CCL, and CYH drafted the manuscript, and critical revisions were made by WC, CCL, JC, HLL, LCC, and CYH. The final approval of the manuscript was provided by WC, CCL, JC, HLL, LCC, and CYH. Administrative, technical, or material support was provided by WC, CCL, and CYH, with resources provided by HLL and LCC. Supervision and validation were undertaken by CCL and CYH, and funding for the study was obtained by CYH.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[PDF File (Adobe PDF File), 169 KB - mededu_v11i1e64780_app1.pdf]

References

1. Shorey S, Ng ED. The use of virtual reality simulation among nursing students and registered nurses: a systematic review. Nurse Educ Today 2021 Mar;98:104662. [doi: 10.1016/j.nedt.2020.104662] [Medline: 33203545]
2. Butt A, Kardong-Edgren S, Ellertson A. Using game-based virtual reality with haptics for skill acquisition. Clinical Simulation in Nursing 2018 Mar;16:25-32 [FREE Full text] [doi: 10.1016/j.ecns.2017.09.010]
3. Günay İ, Zaybak A. Comparison of the effectiveness of a virtual simulator with a plastic arm model in teaching intravenous catheter insertion skills. Comput Inform Nurs 2018;36(2):98-105. [doi: 10.1097/cin.0000000000000405]
4. Helle N, Vikman MD, Dahl-Michelsen T, Lie SS. Health care and social work students' experiences with a virtual reality simulation learning activity: qualitative study. JMIR Med Educ 2023 Sep 20;9:e49372 [FREE Full text] [doi: 10.2196/49372] [Medline: 37728988]
5. Park S, Shin HJ, Kwak H, Lee HJ. Effects of immersive technology-based education for undergraduate nursing students: systematic review and meta-analysis using the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) approach. J Med Internet Res 2024 Jul 24;26:e57566 [FREE Full text] [doi: 10.2196/57566] [Medline: 38978483]
6. Yoon H. Effects of immersive straight catheterization virtual reality simulation on skills, confidence, and flow state in nursing students. Comput Inform Nurs 2024;42(12):872-887. [doi: 10.1097/cin.0000000000001141]
7. Abou Hashish EA, Alnajjar H. Digital proficiency: assessing knowledge, attitudes, and skills in digital transformation, health literacy, and artificial intelligence among university nursing students. BMC Med Educ 2024 May 07;24(1):508. [doi: 10.1186/s12909-024-05482-3] [Medline: 38715005]
8. Andreoli KG, Musser LA. Challenges confronting the future of emergency nursing. J Emerg Nurs 2020 Sep;46(5):573-578. [doi: 10.1016/j.jen.2020.04.006] [Medline: 32828478]
9. Jones T, Curtis K, Shaban RZ. Academic and professional characteristics of Australian graduate emergency nursing programs: a national study. Australas Emerg Care 2020 Sep;23(3):173-180. [doi: 10.1016/j.auec.2020.02.003] [Medline: 32115399]
10. Provenzano S, Santangelo O, Armetta F, Pesco G, Allegro A, Lampasona M, et al. COVID-19 infection: comparing the knowledge, attitude and practices in a sample of nursing students. Acta Biomed 2020 Nov 30;91(12-S):e2020001 [FREE Full text] [doi: 10.23750/abm.v91i12-S.10252] [Medline: 33263338]
11. Mallek F, Mazhar T, Faisal Abbas Shah S, Ghadi Y, Hamam H. A review on cultivating effective learning: synthesizing educational theories and virtual reality for enhanced educational experiences. PeerJ Comput Sci 2024;10:e2000 [FREE Full text] [doi: 10.7717/peerj-cs.2000] [Medline: 38855256]
12. Keller J. Motivational Design for Learning and Performance: The ARCS Model Approach. New York, NY: Springer Nature; 2010.
13. Thompson D, Thompson A, McConnell K. Nursing students' engagement and experiences with virtual reality in an undergraduate bioscience course. Int J Nurs Educ Scholarsh 2020 Sep 16;17(1):20190081 [FREE Full text] [doi: 10.1515/ijnes-2019-0081] [Medline: 32941161]
14. Fealy S, Jones D, Hutton A, Graham K, McNeill L, Sweet L, et al. The integration of immersive virtual reality in tertiary nursing and midwifery education: a scoping review. Nurse Educ Today 2019 Aug;79:14-19. [doi: 10.1016/j.nedt.2019.05.002] [Medline: 31078869]

XSL•FO
RenderX

15. Rourke S. How does virtual reality simulation compare to simulated practice in the acquisition of clinical psychomotor skills for pre-registration student nurses? A systematic review. Int J Nurs Stud 2020 Feb;102:103466. [doi: 10.1016/j.ijnurstu.2019.103466] [Medline: 31783192]
16. Keller JM. Motivation, learning, and technology: applying the ARCS-V motivation model. Particip Educ Res 2016;3(2):1-13. [doi: 10.17275/per.16.06.3.2]
17. Killam LA, Timmermans KE, Shapiro SJ. Motivation and engagement of nursing students in 2 gamified courses. Nurse Educ 2021 Aug 6;46(6):E173-E178. [doi: 10.1097/nne.0000000000001065]
18. Wu H, Li S, Zheng J, Guo J. Medical students' motivation and academic performance: the mediating roles of self-efficacy and learning engagement. Med Educ Online 2020 Dec;25(1):1742964 [FREE Full text] [doi: 10.1080/10872981.2020.1742964] [Medline: 32180537]
19. Mohamed Mohamed Bayoumy H, Alsayed S. Investigating relationship of perceived learning engagement, motivation, and academic performance among nursing students: a multisite study. AMEP 2021 Apr;Volume 12:351-369. [doi: 10.2147/amep.s272745]
20. Kowitlawakul Y, Tan JJM, Suebnukarn S, Nguyen HD, Poo DCC, Chai J, et al. Utilizing educational technology in enhancing undergraduate nursing students' engagement and motivation: A scoping review. J Prof Nurs 2022;42:262-275. [doi: 10.1016/j.profnurs.2022.07.015] [Medline: 36150870]
21. Chang YM, Lai CL. Exploring the experiences of nursing students in using immersive virtual reality to learn nursing skills. Nurse Educ Today 2021 Feb;97:104670. [doi: 10.1016/j.nedt.2020.104670] [Medline: 33264739]
22. Mäkinen H, Haavisto E, Havola S, Koivisto J. Graduating nursing students' user experiences of the immersive virtual reality simulation in learning - a qualitative descriptive study. Nurs Open 2023 May;10(5):3210-3219 [FREE Full text] [doi: 10.1002/nop2.1571] [Medline: 36598872]
23. Plotzky C, Loessl B, Kuhnert B, Friedrich N, Kugler C, König P, et al. My hands are running away - learning a complex nursing skill via virtual reality simulation: a randomised mixed methods study. BMC Nurs 2023 Jun 27;22(1):222. [doi: 10.1186/s12912-023-01384-9] [Medline: 37370124]
24. Saab M, McCarthy M, O'Mahony B. Virtual reality simulation in nursing and midwifery education: a usability study. Comput Inform Nurs 2023;41(10):815-824. [doi: 10.1097/cin.0000000000001010]
25. Wood J, Ebert L, Duff J. Implementation methods of virtual reality simulation and the impact on confidence and stress when learning patient resuscitation: an integrative review. Clin Simul Nursing 2022;66:5-17 [FREE Full text] [doi: 10.1016/j.ecns.2022.02.006]
26. Yang S, Oh Y. The effects of neonatal resuscitation gamification program using immersive virtual reality: a quasi-experimental study. Nurse Educ Today 2022 Oct;117:105464 [FREE Full text] [doi: 10.1016/j.nedt.2022.105464] [Medline: 35914345]
27. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191 [FREE Full text] [doi: 10.3758/bf03193146] [Medline: 17695343]
28. Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd Edition). Hillside, NJ: Lawrence Erlbaum Associates; 1998.
29. Woon A, Mok W, Chieng Y, Zhang H, Ramos P, Mustadi H, et al. Effectiveness of virtual reality training in improving knowledge among nursing students: a systematic review, meta-analysis and meta-regression. Nurse Educ Today 2021 Mar;98:104655 [FREE Full text] [doi: 10.1016/j.nedt.2020.104655] [Medline: 33303246]
30. George R, Titus SK. Benefits and barriers of using virtual reality in teaching undergraduate nursing students. Nurse Educ 2024;49(5):E244-E249. [doi: 10.1097/nne.0000000000001660]
31. Marton F. Phenomenography-A research approach to investigating different understanding of reality. J Thought 1986;21(3):28-49.
32. Dahlgren L, Fallsberg M. Phenomenography as a qualitative approach in social pharmacy research. J Soc Admin Pharm 1991;8:150-156. [doi: 10.4135/9781412963909.n316]
33. Lincoln YS, Guba EG. Naturalistic Inquiry. Thousand Oaks, CA: Sage; 1985.
34. Sjöström B, Dahlgren LO. Applying phenomenography in nursing research. J Adv Nurs 2002 Nov;40(3):339-345. [doi: 10.1046/j.1365-2648.2002.02375.x] [Medline: 12383185]
35. Chen F, Leng Y, Ge J, Wang D, Li C, Chen B, et al. Effectiveness of virtual reality in nursing education: meta-analysis. J Med Internet Res 2020 Sep 15;22(9):e18290. [doi: 10.2196/18290] [Medline: 32930664]
36. Jallad ST, Işık B. The effectiveness of virtual reality simulation as learning strategy in the acquisition of medical skills in nursing education: a systematic review. Ir J Med Sci 2022 Jun;191(3):1407-1426. [doi: 10.1007/s11845-021-02695-z] [Medline: 34227032]
37. Ryan GV, Callaghan S, Rafferty A, Higgins MF, Mangina E, McAuliffe F. Learning outcomes of immersive technologies in health care student education: systematic review of the literature. J Med Internet Res 2022;24(2):e30082. [doi: 10.2196/30082]
38. Huai P, Li Y, Wang X, Zhang L, Liu N, Yang H. The effectiveness of virtual reality technology in student nurse education: a systematic review and meta-analysis. Nurse Educ Today 2024 Jul;138:106189. [doi: 10.1016/j.nedt.2024.106189] [Medline: 38603830]

XSL•FO
RenderX

39. Lo Y, Yang C, Yeh T, Tu H, Chang Y. Effectiveness of immersive virtual reality training in nasogastric tube feeding education: a randomized controlled trial. Nurse Educ Today 2022 Dec;119:105601. [doi: 10.1016/j.nedt.2022.105601] [Medline: 36244254]
40. Lee H, Han J, Park J, Min S, Park J. Development and evaluation of extracorporeal membrane oxygenation nursing education program for nursing students using virtual reality. BMC Med Educ 2024 Jan 26;24(1):92. [doi: 10.1186/s12909-024-05057-2] [Medline: 38279179]
41. Shujuan L, Mawpin T, Meichan C, Weijun X, Jing W, Biru L. The use of virtual reality to improve disaster preparedness among nursing students: a randomized study. J Nurs Educ 2022 Feb;61(2):93-96. [doi: 10.3928/01484834-20211213-05] [Medline: 35112954]
42. Fontenot J, Hebert M, Lin H, Kulshreshth AK. Examining the perceptions among undergraduate nursing students using virtual reality in a community course: a mixed-methods explanatory study. J Community Health Nurs 2024;41(3):145-155. [doi: 10.1080/07370016.2023.2280617] [Medline: 37966021]
43. Lau ST, Siah CJR, Loh WL, Rusli KDB, Schmidt LT, Lim FP, et al. Enhancing professional competency in clinical procedures using head-mounted display virtual reality - a mixed method study. Med Educ Online 2023 Dec;28(1):2232134. [doi: 10.1080/10872981.2023.2232134] [Medline: 37406175]

Abbreviations

ARCS: Attention, Relevance, Confidence, and Satisfaction
IMMS: Instructional Materials Motivation Survey
PPE: personal protective equipment
VR: virtual reality
VRS: virtual reality simulation

Edited by B Lesselroth; submitted 28.07.24; peer-reviewed by LH Wang, SY Wang, T Loetscher; comments to author 10.09.24; revised version received 17.09.24; accepted 14.01.25; published 11.02.25.

Please cite as:
Chang W, Lin CC, Crilly J, Lee HL, Chen LC, Han CY
Virtual Reality Simulation for Undergraduate Nursing Students for Care of Patients With Infectious Diseases: Mixed Methods Study
JMIR Med Educ 2025;11:e64780
URL: https://mededu.jmir.org/2025/1/e64780
doi:10.2196/64780
PMID:

©Wen Chang, Chun-Chih Lin, Julia Crilly, Hui-Ling Lee, Li-Chin Chen, Chin-Yen Han. Originally published in JMIR Medical Education (https://mededu.jmir.org), 11.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mono-Professional Simulation-Based Obstetric Training in a Low-Resource Setting: Stepped-Wedge Cluster Randomized Trial

Anne A C van Tetering[1,2,3], PhD, MD; Ella L de Vries[1,3], MSc, MD; Peter Ntuyo[4], MD; E R van den Heuvel[5], Prof Dr; Annemarie F Fransen[1,3], PhD, MD; M Beatrijs van der Hout-van der Jagt[1,3,6], MSc, PhD; Imelda Namagembe[4], PhD, MD; Josaphat Byamugisha[7], Prof Dr, MD; S Guid Oei[1,3], Prof Dr, MD

[1]Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
[2]Department of Obstetrics and Gynaecology, Amphia Ziekenhuis, Breda, The Netherlands
[3]Department of Obstetrics and Gynaecology, Máxima Medical Center, De Run 4600, Veldhoven, The Netherlands
[4]Department of Maternal Fetal Medicine, Mulago Specialised Women and Neonatal Hospital, Kampala, Uganda
[5]Dean of Mathematics & Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands
[6]Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
[7]Department of Obstetrics and Gynaecology, School of Medicine, Makerere University College of Health Sciences, Kampala, Uganda

Corresponding Author:
Anne A C van Tetering, PhD, MD
Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

Abstract

Background: Emergency obstetric simulation-based training has increasingly been used to improve emergency obstetric care provision in sub-Saharan Africa. For determining the optimal methodology for effective training sessions in resource-constrained settings, it is crucial to conduct high-quality research.

Objective: We aim to investigate the impact of a train-the-trainer model for providing technology-enhanced, mono-professional, simulation-based training in obstetrics in a resource-constrained setting on maternal and perinatal outcomes.

Methods: A stepped-wedge cluster randomized trial was conducted from October 2014 until March 2016 at the medium- to high-risk ward at Mulago National Referral Hospital, Uganda, with an annual delivery rate of over 23,000. The intervention consisted of a train-the-trainer model in which training was cascaded down from master trainers to local facilitators (obstetric senior staff members) to learners (senior house officers). The training of senior house officers was provided to 7 fixed clusters by a computer-generated random sequential roll-out. The training comprised a 1-day (8 h), mono-professional, simulation-based training in obstetrics, and half-day repetition training sessions targeted at every 7 weeks. Both medical technical skills and teamwork skills were taught. The primary outcome comprised a combined maternal and perinatal mortality rate. Secondary outcomes comprised the maternal mortality rate, the perinatal mortality rate, the percentage of births by vacuum extraction and cesarean section, and the Weighted Adverse Outcome Score.

Results: Overall, there were 17,496 births. The combined mortality rate was 9.05% (95% CI 8.37% - 9.77%) in the intervention group, and 8.73% (95% CI 8.21% - 9.28%) in the control group (odds ratio [OR] 0.98, 95% CI 0.86 - 1.12; *P*=.81). No statistically significant change was found in the maternal mortality rate (OR 0.80, 95% CI 0.27 - 2.32; *P*=.68) or the perinatal mortality rate (OR 0.99, 95% CI 0.87 - 1.13; *P*=.87). This study did not identify any difference in the percentage of vacuum extractions, the percentage of cesarean sections, or Weighted Adverse Outcome Scores.

Conclusions: This train-the-trainer model for providing technology-enhanced, mono-professional, simulation-based training in obstetrics was not able to change maternal and perinatal mortality outcomes. This study, in combination with literature, suggests that future research should consider multiprofessional team training in obstetrics involving all staff within their units.

Trial Registration: ISRCTN Registry ISRCTN98617255; https://www.isrctn.com/ISRCTN98617255

International Registered Report Identifier (IRRID): RR2-10.2196/17277

(JMIR Med Educ 2025;11:e54911) doi:10.2196/54911

KEYWORDS

obstetric; simulation; training; low income; middle income; LMIC; simulation training; medical education; mono-professional; obstetric training; low-resource setting; low-income setting; stepped-wedge; RCT; obstetric care; pregnancy care; Africa; maternal; perinatal; outcome; hospital; train-the-trainer; facilitator; trainer; learner; mortality rate; randomized controlled trial

XSL•FO
RenderX

Introduction

Emergency Obstetric Care in Uganda

Uganda continues to face challenges in providing safe obstetric care. Despite an increase in the rate of institutional births from 59% to 74%, the maternal mortality ratio was still high at 375 per 100,000 live births in 2017, accompanied by a neonatal mortality rate of 21 deaths per 1000 live births [1]. Key barriers for providing safe childbirth include shortcomings in the management of emergency obstetric care, delays in referral practices, and insufficient coordination among health care staff, all of which obstruct the provision of adequate emergency obstetric care [2].

Simulation-Based Obstetric Training

To address these challenges, simulation-based training for emergency obstetric care has evolved as a promising approach in sub-Saharan Africa. Growing evidence suggests that this type of training improves health care providers' knowledge and skills, while also leading to positive changes in their behavior [3-5]. Additionally, evidence from other studies has shown encouraging effects on patient outcomes, including reported reductions in neonatal and perinatal mortality rates, as well as potential decreases in maternal mortality and postpartum hemorrhage [6-9]. Despite these promising findings, assessments of patient outcomes remain infrequent, and the results are often inconsistent [3,4].

Evaluating Simulation-Based Training

One limitation of current evaluations is the reliance on 1-group pretest-posttest designs, which often fail to control for external variables that may influence the results. Furthermore, significant variability exists in training length, content, and design, with programs ranging from mono-professional to multi-professional approaches. This variation makes it difficult to identify which components most effectively contribute to the success of the training. Additional challenges, such as resource constraints, difficulties in sustaining training programs, staff shortages, and high turnover rates, further hinder the implementation and long-term impact of simulation-based training in sub-Saharan Africa. To overcome these challenges, high-quality research is essential to determine the most effective methodologies for emergency obstetric simulation-based training in sub-Saharan Africa.

This study aimed to evaluate the effect of a train-the-trainer program designed to provide technology-enhanced, mono-professional, simulation-based obstetric training on patient outcomes in Uganda [10].

Methods

Setting

A stepped-wedge cluster randomized trial was conducted from October 2014 until March 2016 at the medium- to high-risk labor ward at Mulago National Referral Hospital in Uganda. This hospital also functions as the main teaching facility for Makerere University College of Medicine and Health Sciences. During this study's period, over 23,000 women gave birth annually at the medium to high-risk labor ward.

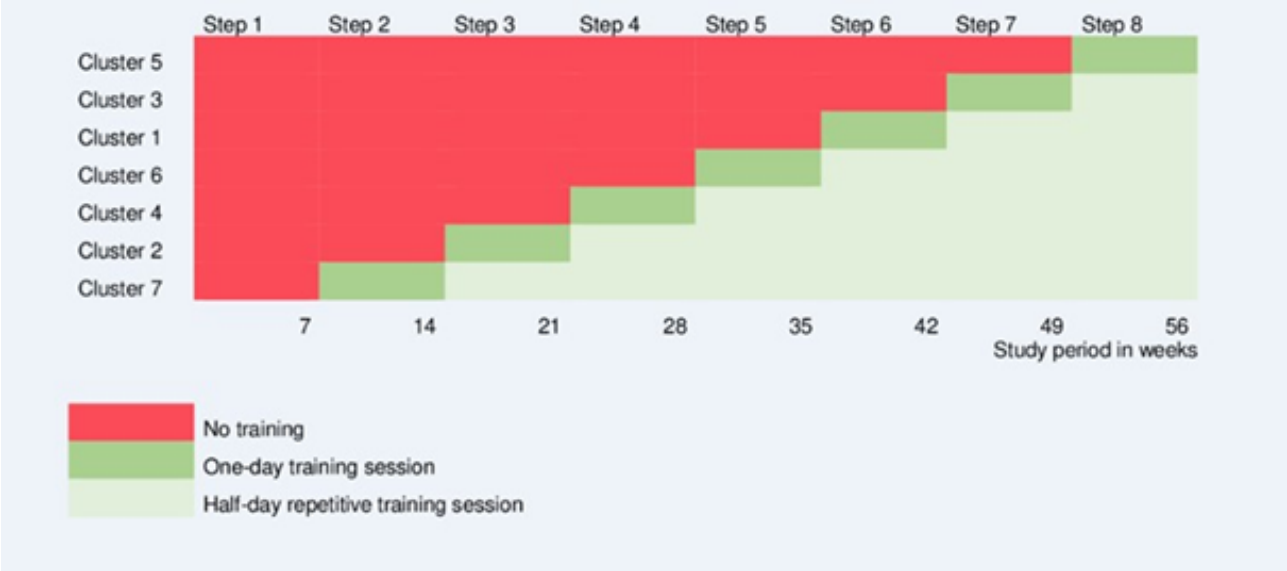
Design and Recruitment

The stepped-wedge cluster randomized trial design facilitated the phased implementation of the training program, with different clusters receiving the intervention at different periods to assess its impact on patient outcomes. This approach allowed for the measurement of the intervention's effect both within and between clusters. Additionally, it enabled the intervention to be provided as a standard service to all participants, while being implemented in stages [11]. As training of all obstetric ward staff was not feasible due to financial and logistical challenges, senior house officers (SHOs) were chosen as the target group for the training program due to their coordinating role in providing emergency obstetric care.

In October 2014, a total of 7 fixed clusters of SHOs were recruited to receive the training. All participants provided written informed consent before this study began. A computer-generated random sequential roll-out of the training program was conducted to determine the order in which the different clusters would receive the intervention (Figure 1). Examination and holiday periods were excluded from the schedule, as fixed clusters could not be maintained in the SHOs' work schedules during these times.

XSL•FO
RenderX

Figure 1. Stepped-wedge cluster randomized design with 7 clusters and 8 steps.



Train-the-Trainer Model

The training program was conducted using a train-the-trainer model, in which training was cascaded down from master trainers to local facilitators, and then to the learners, who were the SHOs. In this model, the master trainers, who were obstetricians from a high-resource setting, had been previously certified as simulation-based trainers by institutions such as EuSim or the Center for Medical Simulation. These master trainers provided a 4-day train-the-trainer program to 14 local facilitators. The facilitators, all gynecologists, were selected based on their clinical and teaching experience by the head of the department. The program included both lectures and practical teaching sessions using simulation-based obstetric scenarios. The train-the-trainer course concluded with an assessment session. During this session, the local facilitators trained intern doctors using a draft SHO training program. Afterward, the master trainers provided feedback to the facilitators. Based on this, 12 of the facilitators were certified as simulation trainers. The SHO training program was then adjusted based on feedback from both trainers and trainees. Subsequently, the local facilitators delivered the training to fixed clusters of 6 to 9 SHOs, comprising first-, second-, and third-year SHOs. A 1-day annual refresher training was offered to all local facilitators. The local facilitators were compensated for lost clinical income by being paid for their participation in the training sessions.

Course Content

Course content was developed by Medsim, a medical simulation center in the Netherlands, in cooperation with senior staff members of Mulago National Referral Hospital. The SHO training program included a 1-day (8 h) mono-professional, simulation-based sessions, followed by half-day refresher sessions every 7 weeks. These refresher sessions started after the switch from the control to the intervention group. Each training session was provided by 2 local facilitators. Scenarios were based on the main local causes of maternal and perinatal mortality and tailored to local clinical protocols and availability of medical equipment. This led to the creation of 2 different scenarios for postpartum hemorrhage, a scenario for eclampsia, a scenario involving fetal distress with a ventouse delivery, and a breech delivery scenario. Both medical-technical and teamwork skills were included in the training, with the difficulty level increasing throughout the day. Every SHO participated in at least 2 scenarios during the 1-day training, while having an observer role in the nonparticipating scenarios. During the repetition training sessions, a single clinical scenario was executed and repeated until skills were mastered.

Data Collection and Outcomes

The primary outcome of this study was the combined maternal and perinatal mortality rate, expressed as a percentage of maternal and perinatal deaths per total number of births. Perinatal deaths were defined as stillbirths and deaths occurring within the first week of life in the special care unit. Data about each delivery and maternal and perinatal outcomes were prospectively registered using the maternity register and transcribed without identification of the subjects. Data about maternal deaths in the high dependency unit, and neonatal deaths in the special care unit were obtained from registration books in these units. These data were matched to and merged with data from the maternity register of the medium to high-risk ward into 1 final electronic database.

Secondary outcomes comprised the maternal mortality rate (maternal deaths per 100,000 births), the perinatal mortality rate (perinatal deaths per 1000 births), percentage of births by vacuum extraction, percentage of births by cesarean section, and the Weighted Adverse Outcome Score (WAOS). The WAOS was defined as the total weighted score of each adverse outcome divided by the total number of births [12]. Four out of 10 index measures (maternal death [750 points], intrapartum or perinatal death [400 points], uterine rupture [100 points], Apgar score less than 7 after 5 minutes [25 points]) were available for registration and assessment. Finally, the maternal mortality ratio (maternal mortality per 100,000 live births), and the perinatal mortality ratio (perinatal mortality per 1000 live births) were calculated for the control and intervention group. As data were analyzed on the cluster level, the authors could not identify individual participants' results.

XSL•FO
RenderX

To provide a comprehensive understanding of the training's effectiveness, additional secondary outcomes were included and published separately, such as the evaluation of the instructional design, participants' reactions (corresponding to Kirkpatrick level 1), and the effects on knowledge, teamwork, and medical-technical skills (corresponding to Kirkpatrick level 2) [13].

Sample Size Calculation

The power calculation was conducted following the methods described by Hussey et al [14] and Woertman et al [10,15,16]. Initially, the sample size for a standard randomized clinical trial was calculated. To show a 20% reduction in combined maternal and perinatal mortality with an α of .05 and a power of 80%, a total of 6398 births would be required for a simple randomized clinical trial design. The design effect was then determined, assuming an intracluster correlation of 0.05, a cluster size of 3343 births per year, and 7 clusters in total. Accounting for the design effect, 2367 births per measurement period would be required. To meet this target, each measurement period would need to last at least 5 weeks. However, for logistical feasibility, the duration of each step was set at 7 weeks, resulting in a total study duration of 56 weeks. Statistical significance was defined as a 2-sided P value of <.05.

Statistical Analysis

Patient baseline characteristics were summarized with medians and IQRs for continuous variables and with counts (percentages) for categorical variables. A generalized linear mixed-effects model was used for the estimation of an intervention effect. Here, the outcome is the binary event on the individual (whether a birth was or was not complicated by maternal mortality, perinatal mortality, or both), and a logit link function was used to model the probability of the event. In the logit scale, the cluster indicator served as a random effect on the intercept, and the period of the stepped wedge was treated as a fixed effect, as well as the intervention effect. The effect of the intervention was reported as an odds ratio, and the performance per treatment group was reported as a percentage or as an incidence rate ratio with 95% CI.

Ethical Considerations

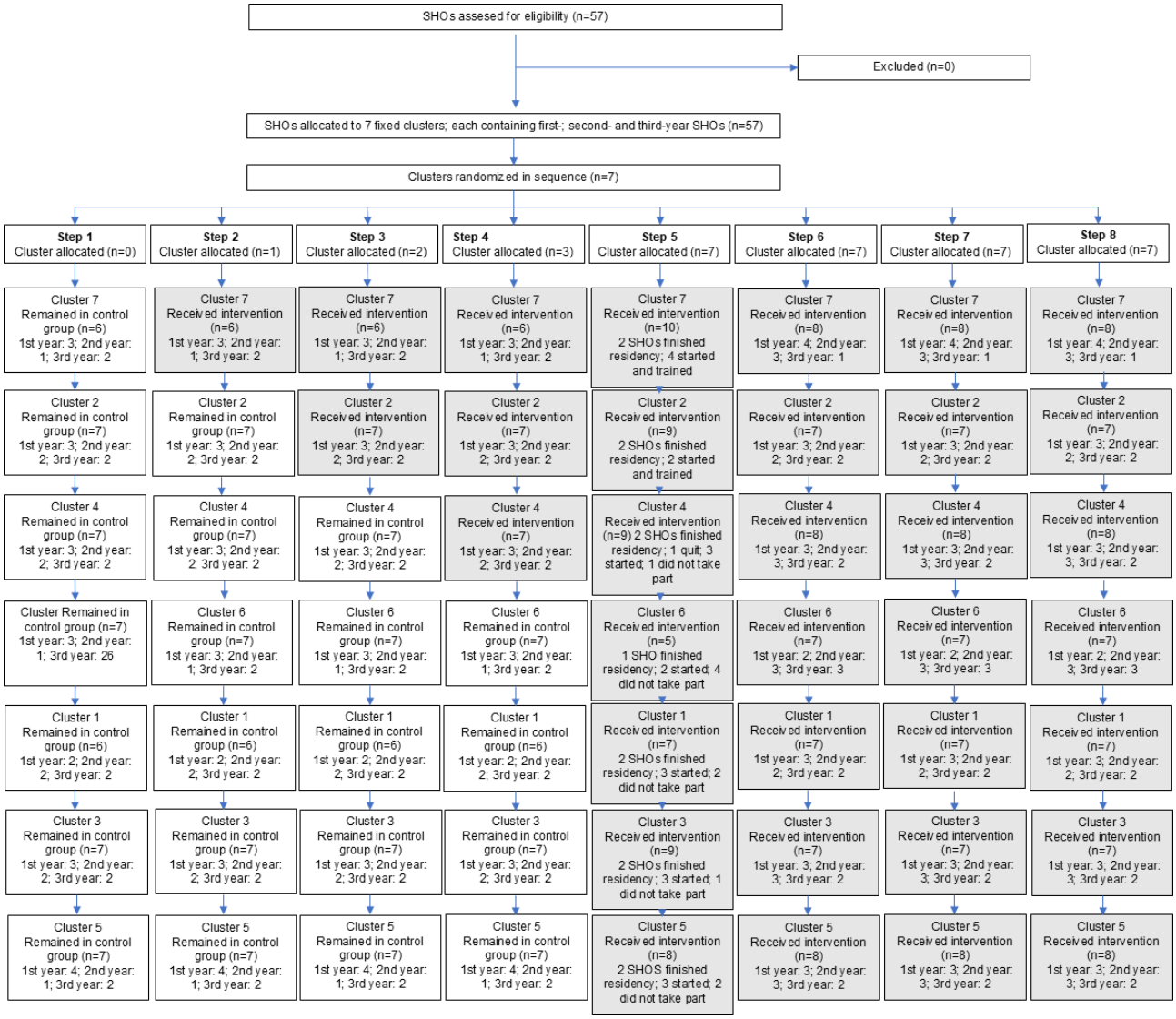
Ethical approval was obtained from the Mulago Research and Ethics Committee (Protocol MREC: 674), and the Uganda National Council for Science and Technology (UNCST, SS 3927). Written informed consent to participants in this study was obtained at the beginning of the first training day, and all participants had the ability to opt out during this study's period. No compensation was provided for the participants. All data were anonymized.

Results

From October 2014 until March 2016, a total of 57 SHOs were randomized into 7 clusters. The first 3 included clusters received the main training on schedule according to protocol. Afterward, when SHOs heard about the training experience from their peers, they requested to expedite the training schedule, rather than waiting for the allocated time slot. Therefore, these 4 remaining clusters were trained within the same week and they simultaneously switched from the control group to the intervention group during this week (Figure 2). After the deviation from the stepped-wedge design in the timing of the intervention, no amendment was needed by the ethical committee, as the intervention itself had not changed.

XSL•FO
RenderX

Figure 2. Flowchart of stepped-wedge cluster randomized design. SHO: senior house officer.



Baseline characteristics are shown in Table 1. Overall, there were 17,496 births. There were fewer female and more male neonates in the intervention group compared to the control group. The results on maternal and perinatal mortality rates, mode of delivery, and the WAOS are shown in Table 2. No differences were found between the intervention and the control group in the combined maternal and perinatal mortality rate, the maternal mortality rate, and the perinatal mortality rate. Results on the percentage of vacuum extractions, the percentage of cesarean sections, and the WAOS did not show any difference between the intervention and the control group. The maternal mortality ratio in the control and intervention group was, respectively, 160.4 (95% CI 94.9 - 266.6) and 116.9 (95% CI 51.2 - 252.4). The perinatal mortality ratio in the control and intervention group was 94.4 (95% CI 88.8 - 100.4) and 98.7 (95% CI 91.3 - 106.6).

XSL•FO
RenderX

Table . Baseline characteristics.

Characteristics			Total	Control group	Intervention group	*P* value
Maternal characteristics						
	Age (years), median (IQR)		24 (21-28)	24 (21-28)	24 (21-29)	.11[a]
	Parity, n					.71[b]
		Primiparous	6760	4219	2541	
		Multiparous	9594	5960	3634	
	Gestation, n (%)					.18[b]
		Preterm (<37 weeks)	1183 (8.2)	753 (8.5)	430 (7.8)	
		Full-term (>37 weeks)	13,215 (91.8)	8149 (91.5)	5066 (92.2)	
	Pregnancy, n (%)					.09[b]
		Singleton	16,166 (92.4)	10,099 (92.5)	6067 (92.2)	
		Twin	1299 (7.4)	795 (7.3)	504 (7.7)	
		Triplet	30 (0.2)	24 (0.2)	6 (0.1)	
Neonatal characteristics						
	Gender, n (%)					.002[b]
		Female	8381 (48.4)	5328 (49.3)	3053 (46.9)	
		Male	8926 (51.6)	5471 (50.7)	3455 (53.1)	
	Birth weight (kg), median (IQR)		3.1 (2.7-3.4)	3.1 (2.6-3.4)	3.1 (2.7-3.4)	.70[a]

[a]Wilcoxon rank sum test.

[b]Pearson chi-square test.

Table . Study results.

	Total	Before intervention	After intervention	Odds ratio	*P* value
Combined mortality rate, % (95% CI)	8.9 (8.4 - 9.3)	8.7 (8.2 - 9.3)	9.1 (8.4-9.8)	0.98 (0.86 to 1.12)	.81
Maternal mortality rate, event rate per 100,000 births (95% CI)	131.5 (85.3 - 200.6)	146.5 (86.7 - 243.6)	106.4 (46.6 - 229.7)	0.80 (0.27 to 2.32)	.68
Perinatal mortality rate, event rate per 1000 births (95% CI)	87.6 (83.5 - 91.9)	86.3 (81.1 - 91.7)	89.8 (83.1 - 97.1)	0.99 (0.87 to 1.13)	.87
Births by vacuum extraction, % (95% CI)	2.3 (2.1 - 2.6)	2.43 (2.2 - 2.8)	2.15 (1.8 - 2.5)	1.00 (0.76 to 1.33)	.99
Births by cesarean section, % (95% CI)	26.6 (25.9 - 27.2)	26 (25.2 - 26.8)	27.5 (26.4 - 28.6)	1.06 (0.94 to 1.2)	.33
Weighted Adverse Outcome Score (WAOS), median score (IQR)	39.6 (0 - 282.6)	39.1 (0 - 280.8)	40.5 (0 - 285.5)	−0.59 (−5.22 to 4.04)[a]	.80

[a]Difference.

XSL•FO
RenderX

Discussion

Principal Results

This train-the-trainer model for providing technology-enhanced, mono-professional, simulation-based obstetric training to SHOs did not result in changes to maternal and perinatal mortality outcomes. The training program also had no impact on the number of instrumental births, the number of cesareans, or the WAOS.

Strengths and Limitations

A strength of this study was the use of a randomized stepped-wedge trial design, enabling the training for all SHOs in one of the busiest labor wards worldwide. Unlike 1 group pretest-posttest designs commonly used in prior research on obstetric simulation-based training, this approach minimized bias from natural changes in health care outcomes because it could eliminate systematic period effects. Another strength was the use of the train-the-trainer model. Research has shown that training delivered by local trainers results in greater improvements in knowledge and skill acquisition [17]. Furthermore, simulation-based learning is likely to be more effective when tailored to the local context and culture [18]. A third strength of the evaluated training program was the inclusion of teamwork skills. Teamwork skills are increasingly recognized as a critical factor in reducing preventable, substandard care and are viewed as an essential competence for hospital teams [19]. These skills had not been part of previous SHO training programs at Mulago National Referral Hospital.

Limitations of our study should also be considered. First, the intended stepped-wedge design was altered during the study, as clusters requested earlier training. Although this deviation interfered with the planned study design, it was deemed unethical to withhold training further. The change in design did not affect the statistical analysis because the mean features of the stepped wedge design were maintained. Other studies have highlighted ethical concerns with the stepped-wedge design, including justifying the delayed rollout of the intervention to the control group, which is inherent to this design [20,21]. A potential solution in the future could involve using different hospitals as clusters. This approach would also address the challenge of maintaining fixed clusters of individuals during working hours, which was one of the difficulties we encountered. Interaction between trained and nontrained SHOs during shifts in examination and holiday periods may have introduced bias into the results, but our analysis was chosen conservatively, making sure that such biases would affect the treatment effect negatively. Another challenge in the work schedules of the fixed clusters was the repetition of sessions, which led to some sessions not being scheduled according to the protocol. Establishing fixed clusters of multidisciplinary obstetric team members within 1 hospital is anticipated to be even more challenging. Including different hospitals in the design could eliminate these issues and allow all health care providers involved in maternity and neonatal care at a single hospital to be trained within a short period. However, it may be difficult to include hospitals with comparable levels of care and delivery volumes, which should be accounted for in the intracluster correlation coefficient and the statistical analysis.

Comparison With Prior Work

The results of our study on maternal and perinatal outcomes do not align with those of recent studies on simulation-based emergency obstetric training in sub-Saharan Africa. Since the start of this study, 6 studies reported improvements in maternal outcomes, mainly related to postpartum hemorrhage and mortality [22-27], and 7 studies showed improvements in neonatal or perinatal outcomes after simulation-based obstetric training [25,28-33]. One of these studies showed that initial improvements declined over time [29].

When comparing our simulation-based training program to others that were effective, we want to highlight the mono-professional nature of our program. While previous research showed that simulation-based team leader training improved teamwork and communication during clinical resuscitations, our study found that training only SHOs as the leaders during obstetric emergencies did not improve patient outcomes. This aligns with the findings of Siassakos et al [34], who noted that units showing improvements had trained nearly 100% of their staff and implemented training programs within their own units. Additionally, all but 2 of the previously described effective studies were multi-professional training programs [22-31]. An exception to justify a mono-professional training program can be when the focus is on a specific technical task performed by a single health care provider, such as repairing an episiotomy. In such cases, the focus on a specific task allows for deliberate practice, where the trainee improves the task through immediate feedback, problem-solving, evaluation, and repeated performance. However, when the task involves teamwork, the training approach should shift toward a multi-professional model. In conclusion, our results, alongside the literature, suggest that future research should consider multi-professional team training in obstetrics, involving all staff within their units.

Another difference between our training program and others is that ours was a stand-alone program, while simulation-based training as part of an integrated package may be more effective in improving health outcomes [4,8,28]. Integrated packages often include equipment, maternal death reviews, health information system improvements, modified protocols, supportive supervision, mobile mentoring, and peer-assisted learning. Although we included a train-the-trainer model, restructured local protocols, and created posters with flowcharts for obstetric emergencies, the intensity and, ultimately, the training frequency of our intervention may not have been sufficient to impact maternal and neonatal outcomes.

Another notable variance between evaluated simulation-based training programs is the location of training. Our study used an off-site medical simulation center, while on-site training may be more beneficial, as it reaches more staff and generates more suggestions for organizational changes. Sorensen et al [35] found no significant differences in knowledge, patient safety attitude, motivation, or stress between on-site and off-site training, but the on-site group suggested more organizational changes. In low-resource settings, these changes may be more

XSL•FO
RenderX

valuable, although on-site training could be disrupted by clinical situations. Further research comparing on-site versus off-site training in low-resource settings would be valuable.

A further difference in previous studies on simulation-based obstetric training is the definition of mortality ratios. Some studies, including the mortality ratios in the introduction, used maternal mortality per live birth, while others used maternal mortality per number of births [8,22,26,36,37], making comparisons difficult. Additionally, the World Health Organization defines maternal and perinatal mortality ratios with different denominators, complicating statistical analyses of a combined mortality ratio. As a result, we analyzed the combined mortality rate and reported the maternal and perinatal mortality ratios separately. Moreover, the ratios are based on live births, so improving perinatal care can also affect maternal mortality outcomes. A standardized approach to mortality ratios could improve the comparability of future training programs.

A final note should be made regarding improvements in data administration. It took considerable time to manually collect, verify, and process all the data. In some cases, determining the time and cause of death was challenging, potentially leading to the inclusion of more macerated babies and higher perinatal mortality rates. This study's setting in a national underresourced referral hospital may also explain the high perinatal mortality rate. Options to address some data challenges include the development of digital data registration systems and active surveillance of data. While these solutions require initial investments of both time and money, they offer significant potential benefits in terms of efficiency and accuracy. Additionally, digital data registration and continuous data monitoring can be enhanced by the use of dashboards, which provide clinicians with an overview of current practices and can help identify deviations from targets early on [38]. This approach not only benefits the evaluation of obstetric simulation-based training but can also inform ongoing training sessions, contributing to continuous learning and improvements in obstetric training.

Given the complexity of simulation-based obstetric training implementation and evaluation in low-resource settings, future studies should consider conducting implementation and action research. This approach would be valuable for identifying barriers to effective implementation, refining training programs, and ensuring that improvements are successfully integrated into the local health care system.

Conclusions

This train-the-trainer model for providing technology-enhanced, mono-professional, simulation-based training in obstetrics to SHOs did not change maternal and perinatal mortality outcomes in a national referral hospital in a low-resource setting. This study, along with existing literature, suggests that future research should consider conducting and evaluating multi-professional team training in obstetrics, involving all staff within their units.

Acknowledgments

We thank Rob Steinweg, MSc, for all the support from Medsim, a research and training center for medical care in the Netherlands. This study was supported by Máxima Medical Center, Veldhoven, the Netherlands, and Rotary Clubs in the Netherlands and Uganda. They both supported this study by providing training supplies. The Rotary Foundation was not involved in this study's design or interpretation of study results.

Conflicts of Interest

None declared.

Checklist 1
CONSORT checklist. CONSORT: Consolidated Standards of Reporting Trials.
[PDF File, 71 KB - mededu_v11i1e54911_app1.pdf]

References

1. WHO, UNICEF, UNFPA, World Bank Group and the UN. Trends in maternal mortality: 2000 to 2017. World Bank Group. 2019. URL: https://data.worldbank.org/indicator/SH.STA.MMRT?locations=UG [accessed 2025-04-16]
2. Geleto A, Chojenta C, Mussa A, Loxton D. Barriers to access and utilization of emergency obstetric care at health facilities in sub-Saharan Africa-a systematic review protocol. Syst Rev 2018 Apr 16;7(1):60. [doi: 10.1186/s13643-018-0720-y] [Medline: 29661217]
3. Bergh AM, Baloyi S, Pattinson RC. What is the impact of multi-professional emergency obstetric and neonatal care training? Best Pract Res Clin Obstet Gynaecol 2015 Nov;29(8):1028-1043. [doi: 10.1016/j.bpobgyn.2015.03.017] [Medline: 25937554]
4. Ameh CA, Mdegela M, White S, van den Broek N. The effectiveness of training in emergency obstetric care: a systematic literature review. Health Policy Plan 2019 May 1;34(4):257-270. [doi: 10.1093/heapol/czz028] [Medline: 31056670]
5. Chou WK, Ullah N, Arjomandi Rad A, et al. Simulation training for obstetric emergencies in low- and lower-middle income countries: a systematic review. Eur J Obstet Gynecol Reprod Biol 2022 Sep;276:74-81. [doi: 10.1016/j.ejogrb.2022.07.003] [Medline: 35820293]
6. Msemo G, Massawe A, Mmbando D, et al. Newborn mortality and fresh stillbirth rates in Tanzania after helping babies breathe training. Pediatrics 2013 Feb;131(2):e353-e360. [doi: 10.1542/peds.2012-1795] [Medline: 23339223]

XSL•FO
RenderX

7. Andreatta P, Gans-Larty F, Debpuur D, Ofosu A, Perosky J. Evaluation of simulation-based training on the ability of birth attendants to correctly perform bimanual compression as obstetric first aid. Int J Nurs Stud 2011 Oct;48(10):1275-1280. [doi: 10.1016/j.ijnurstu.2011.03.001] [Medline: 21450290]
8. Dumont A, Fournier P, Abrahamowicz M, et al. Quality of care, risk management, and technology in obstetrics to reduce hospital-based maternal mortality in Senegal and Mali (QUARITE): a cluster-randomised trial. Lancet 2013 Jul 13;382(9887):146-157. [doi: 10.1016/S0140-6736(13)60593-0] [Medline: 23721752]
9. Sorensen BL, Rasch V, Massawe S, Nyakina J, Elsass P, Nielsen BB. Advanced life support in obstetrics (ALSO) and post-partum hemorrhage: a prospective intervention study in Tanzania. Acta Obstet Gynecol Scand 2011 Jun;90(6):609-614. [doi: 10.1111/j.1600-0412.2011.01115.x] [Medline: 21388368]
10. van Tetering AAC, van Meurs A, Ntuyo P, et al. Study protocol training for life: a stepped wedge cluster randomized trial about emergency obstetric simulation-based training in a low-income country. BMC Pregnancy Childbirth 2020 Jul 28;20(1):429. [doi: 10.1186/s12884-020-03050-3] [Medline: 32723330]
11. Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC Med Res Methodol 2006 Nov 8;6:54. [doi: 10.1186/1471-2288-6-54] [Medline: 17092344]
12. Nielsen PE, Goldman MB, Mann S, et al. Effects of teamwork training on adverse outcomes and process of care in labor and delivery: a randomized controlled trial. Obstet Gynecol 2007 Jan;109(1):48-55. [doi: 10.1097/01.AOG.0000250900.53126.c2] [Medline: 17197587]
13. van Tetering AAC, Segers MHM, Ntuyo P, et al. Evaluating the instructional design and effect on knowledge, teamwork, and skills of technology-enhanced simulation-based training in obstetrics in Uganda: stepped-wedge cluster randomized trial. JMIR Med Educ 2021 Feb 5;7(1):e17277. [doi: 10.2196/17277] [Medline: 33544086]
14. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials 2007 Feb;28(2):182-191. [doi: 10.1016/j.cct.2006.05.007] [Medline: 16829207]
15. Woertman W, de Hoop E, Moerbeek M, Zuidema SU, Gerritsen DL, Teerenstra S. Stepped wedge designs could reduce the required sample size in cluster randomized trials. J Clin Epidemiol 2013 Jul;66(7):752-758. [doi: 10.1016/j.jclinepi.2013.01.009] [Medline: 23523551]
16. de Hoop E, Woertman W, Teerenstra S. The stepped wedge cluster randomized trial always requires fewer clusters but not always fewer measurements, that is, participants than a parallel cluster randomized trial in a cross-sectional design. J Clin Epidemiol 2013 Dec;66(12):1428. [doi: 10.1016/j.jclinepi.2013.07.008]
17. Chong W, Walsh M, Zannat F, et al. Effectiveness of foreign versus local trainers on skilled birth attendants' knowledge and skills acquisition after simulation-based training in rural Rwanda. Int J Gynecol Obstet 21st FIGO World Congr Gynecol Obstet BC CanadaConference Publ 2015;131(SUPPL. 5):E486.
18. Chung HS, Dieckmann P, Issenberg SB. It is time to consider cultural differences in debriefing. Simul Healthc 2013 Jun;8(3):166-170. [doi: 10.1097/SIH.0b013e318291d9ef] [Medline: 23702587]
19. Gordon M, Darbyshire D, Baker P. Non-technical skills training to enhance patient safety: a systematic review. Med Educ 2012 Nov;46(11):1042-1054. [doi: 10.1111/j.1365-2923.2012.04343.x] [Medline: 23078681]
20. Joag K, Ambrosio G, Kestler E, Weijer C, Hemming K, Van der Graaf R. Ethical issues in the design and conduct of stepped-wedge cluster randomized trials in low-resource settings. Trials 2019 Dec 19;20(Suppl 2):703. [doi: 10.1186/s13063-019-3842-1] [Medline: 31852547]
21. Prost A, Binik A, Abubakar I, et al. Logistic, ethical, and political dimensions of stepped wedge trials: critical review and case studies. Trials 2015 Aug 17;16(1):351. [doi: 10.1186/s13063-015-0837-4] [Medline: 26278521]
22. Pattinson RC, Bergh AM, Ameh C, et al. Reducing maternal deaths by skills-and-drills training in managing obstetric emergencies: a before-and-after observational study. S Afr Med J 2019 Mar 29;109(4):241-245. [doi: 10.7196/SAMJ.2019.v109i4.13578] [Medline: 31084689]
23. Nelissen E, Ersdal H, Mduma E, et al. Clinical performance and patient outcome after simulation-based training in prevention and management of postpartum haemorrhage: an educational intervention study in a low-resource setting. BMC Pregnancy Childbirth 2017 Sep 11;17(1):301. [doi: 10.1186/s12884-017-1481-7] [Medline: 28893211]
24. Hanson C, Atuhairwe S, Lucy Atim J, Marrone G, Morris JL, Kaharuza F. Effects of the helping mothers survive bleeding after birth training on near miss morbidity and mortality in Uganda: a cluster-randomized trial. Int J Gynaecol Obstet 2021 Mar;152(3):386-394. [doi: 10.1002/ijgo.13395] [Medline: 32981091]
25. Evans CL, Bazant E, Atukunda I, et al. Peer-assisted learning after onsite, low-dose, high-frequency training and practice on simulators to prevent and treat postpartum hemorrhage and neonatal asphyxia: a pragmatic trial in 12 districts in Uganda. PLoS ONE 2018;13(12):e0207909. [doi: 10.1371/journal.pone.0207909] [Medline: 30557350]
26. Chang OH, Levy B, Lytle H, et al. Implementation of the Alliance for Innovation on Maternal Health Program to reduce maternal mortality in Malawi. Obstet Gynecol 2019 Mar;133(3):507-514. [doi: 10.1097/AOG.0000000000003108] [Medline: 30741809]
27. Zongo A, Dumont A, Fournier P, Traore M, Kouanda S, Sondo B. Effect of maternal death reviews and training on maternal mortality among cesarean delivery: post-hoc analysis of a cluster-randomized controlled trial. Eur J Obstet Gynecol Reprod Biol 2015 Feb;185:174-180. [doi: 10.1016/j.ejogrb.2014.12.023] [Medline: 25590501]

XSL•FO
RenderX

28. Walker D, Otieno P, Butrick E, et al. Effect of a quality improvement package for intrapartum and immediate newborn care on fresh stillbirth and neonatal mortality among preterm and low-birthweight babies in Kenya and Uganda: a cluster-randomised facility-based trial. Lancet Glob Health 2020 Aug;8(8):e1061-e1070. [doi: 10.1016/S2214-109X(20)30232-1]
29. Rule ARL, Maina E, Cheruiyot D, Mueri P, Simmons JM, Kamath-Rayne BD. Using quality improvement to decrease birth asphyxia rates after "Helping Babies Breathe" training in Kenya. Acta Paediatr 2017 Oct;106(10):1666-1673. [doi: 10.1111/apa.13940] [Medline: 28580692]
30. Mduma ER, Ersdal H, Kvaloy JT, et al. Using statistical process control methods to trace small changes in perinatal mortality after a training program in a low-resource setting. Int J Qual Health Care 2018 May 1;30(4):271-275. [doi: 10.1093/intqhc/mzy003]
31. Mduma E, Ersdal H, Svensen E, Kidanto H, Auestad B, Perlman J. Frequent brief on-site simulation training and reduction in 24-h neonatal mortality--an educational intervention study. Resuscitation 2015 Aug;93:1-7. [doi: 10.1016/j.resuscitation.2015.04.019] [Medline: 25957942]
32. Gomez PP, Nelson AR, Asiedu A, et al. Accelerating newborn survival in Ghana through a low-dose, high-frequency health worker training approach: a cluster randomized trial. BMC Pregnancy Childbirth 2018 Mar 22;18(1):72. [doi: 10.1186/s12884-018-1705-5] [Medline: 29566659]
33. Eblovi D, Kelly P, Afua G, Agyapong S, Dante S, Pellerite M. Retention and use of newborn resuscitation skills following a series of helping babies breathe trainings for midwives in rural Ghana. Glob Health Action 2017;10(1):1387985. [doi: 10.1080/16549716.2017.1387985] [Medline: 29058568]
34. Siassakos D, Crofts JF, Winter C, Weiner CP, Draycott TJ. The active components of effective training in obstetric emergencies. BJOG An Int J Obstet Gynaecol 2009 Jul;116(8):1028-1032. [doi: 10.1111/j.1471-0528.2009.02178.x] [Medline: 19438497]
35. Sørensen JL, van der Vleuten C, Rosthøj S, et al. Simulation-based multiprofessional obstetric anaesthesia training conducted in situ versus off-site leads to similar individual and team outcomes: a randomised educational trial. BMJ Open 2015 Oct 6;5(10):e008344. [doi: 10.1136/bmjopen-2015-008344] [Medline: 26443654]
36. Fransen AF, van de Ven J, Banga FR, Mol BWJ, Oei SG. Multi-professional simulation-based team training in obstetric emergencies for improving patient outcomes and trainees' performance. Cochrane Database Syst Rev 2020 Dec 16;12(12):CD011545. [doi: 10.1002/14651858.CD011545.pub2] [Medline: 33325570]
37. Fransen AF, van de Ven J, Schuit E, van Tetering A, Mol BW, Oei SG. Simulation-based team training for multi-professional obstetric care teams to improve patient outcome: a multicentre, cluster randomised controlled trial. BJOG An Int J Obstet Gynaecol 2017 Mar;124(4):641-650. [doi: 10.1111/1471-0528.14369] [Medline: 27726304]
38. Crofts J, Moyo J, Ndebele W, Mhlanga S, Draycott T, Sibanda T. Adaptation and implementation of local maternity dashboards in a Zimbabwean hospital to drive clinical improvement. Bull World Health Organ 2014 Feb 1;92(2):146-152. [doi: 10.2471/BLT.13.124347] [Medline: 24623908]

Abbreviations

OR: odds ratio
SHO: senior house officer
WAOS: Weighted Adverse Outcome Score

Edited by A Bahattab; submitted 27.11.23; peer-reviewed by CI Sartorão Filho, D Ellwood; revised version received 05.01.25; accepted 30.01.25; published 09.05.25.

Please cite as:
van Tetering AAC, de Vries EL, Ntuyo P, van den Heuvel ER, Fransen AF, van der Hout-van der Jagt MB, Namagembe I, Byamugisha J, Oei SG
Mono-Professional Simulation-Based Obstetric Training in a Low-Resource Setting: Stepped-Wedge Cluster Randomized Trial
JMIR Med Educ 2025;11:e54911
URL: https://mededu.jmir.org/2025/1/e54911
doi:10.2196/54911

© Anne A C van Tetering, Ella L de Vries, Peter Ntuyo, E R van den Heuvel, Annemarie F Fransen, M Beatrijs van der Hout-van der Jagt, Imelda Namagembe, Josaphat Byamugisha, S Guid Oei. Originally published in JMIR Medical Education (https://mededu.jmir.org), 9.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Awareness and Attitude Toward Artificial Intelligence Among Medical Students and Pathology Trainees: Survey Study

Anwar Rjoop[1], MBBS, MD; Mohammad Al-Qudah[1,2], MBBS, MD; Raja Alkhasawneh[3], MBBS, MD; Nesreen Bataineh[4], MBBS, MD; Maram Abdaljaleel[5], MBBS, MD; Moayad A Rjoub[6], MBBS, MD; Mustafa Alkhateeb[7]; Mohammad Abdelraheem[7]; Salem Al-Omari[7]; Omar Bani-Mari[7]; Anas Alkabalan[7]; Saoud Altulaih[7]; Iyad Rjoub[7], MBBS, MD; Rula Alshimi[7]

[1]Department of Pathology and Microbiology, Faculty of Medicine, Jordan University of Science and Technology, Irbid, Jordan
[2]Department of Microbiology, Pathology and Forensic Medicine, Faculty of Medicine, The Hashemite University, Zarqa, Jordan
[3]Department of Pulmonary Medicine, King Hussain Medical Center, Royal Medical Services, Amman, Jordan
[4]Department of Basic Medical Sciences, Faculty of Medicine, Yarmouk University, Irbid, Jordan
[5]Department of Pathology, Microbiology, and Forensic Medicine, School of Medicine, The University of Jordan, Amman, Jordan
[6]Department of General Surgery and Urology, Faculty of Medicine, Jordan University of Science and Technology, Irbid, Jordan
[7]Faculty of Medicine, Jordan University for Science and Technology, Irbid, Jordan

Corresponding Author:
Anwar Rjoop, MBBS, MD
Department of Pathology and Microbiology, Faculty of Medicine, Jordan University of Science and Technology, Irbid, Jordan

Abstract

Background: Artificial intelligence (AI) is set to shape the future of medical practice. The perspective and understanding of medical students are critical for guiding the development of educational curricula and training.

Objective: This study aims to assess and compare medical AI-related attitudes among medical students in general medicine and in one of the visually oriented fields (pathology), along with illuminating their anticipated role of AI in the rapidly evolving landscape of AI-enhanced health care.

Methods: This was a cross-sectional study that used a web-based survey composed of a closed-ended questionnaire. The survey addressed medical students at all educational levels across the 5 public medical schools, along with pathology residents in 4 residency programs in Jordan.

Results: A total of 394 respondents participated (328 medical students and 66 pathology residents). The majority of respondents (272/394, 69%) were already aware of AI and deep learning in medicine, mainly relying on websites for information on AI, while only 14% (56/394) were aware of AI through medical schools. There was a statistically significant difference in awareness among respondents who consider themselves tech experts compared with those who do not (P=.03). More than half of the respondents believed that AI could be used to diagnose diseases automatically (213/394, 54.1% agreement), with medical students agreeing more than pathology residents (P=.04). However, more than one-third expressed fear about recent AI developments (167/394, 42.4% agreed). Two-thirds of respondents disagreed that their medical schools had educated them about AI and its potential use (261/394, 66.2% disagreed), while 46.2% (182/394) expressed interest in learning about AI in medicine. In terms of pathology-specific questions, 75.4% (297/394) agreed that AI could be used to identify pathologies in slide examinations automatically. There was a significant difference between medical students and pathology residents in their agreement (P=.001). Overall, medical students and pathology trainees had similar responses.

Conclusions: AI education should be introduced into medical school curricula to improve medical students' understanding and attitudes. Students agreed that they need to learn about AI's applications, potential hazards, and legal and ethical implications. This is the first study to analyze medical students' views and awareness of AI in Jordan, as well as the first to include pathology residents' perspectives. The findings are consistent with earlier research internationally. In comparison with prior research, these attitudes are similar in low-income and industrialized countries, highlighting the need for a global strategy to introduce AI instruction to medical students everywhere in this era of rapidly expanding technology.

(JMIR Med Educ 2025;11:e62669) doi:10.2196/62669

KEYWORDS

artificial intelligence; AI; deep learning; medical schools; pathology; Jordan; medical education; awareness; attitude; medical students; pathology trainees; national survey study; medical practice; training; web-based survey; survey; questionnaire

XSL•FO
RenderX

Introduction

Artificial intelligence (AI) is the capability of machines to simulate intelligence by exhibiting human-like traits such as understanding, deductive reasoning, and problem-solving abilities [1]. In medicine, many specialties have already used AI in clinical practice, such as oncology for cancer detection and grading [2]. AI, particularly deep learning, has garnered a lot of attention in medical education and pathology in recent years [3,4]. These techniques have been mostly used for visual activities such as consultations, seminar presentations, board exams, and archiving [5]. Even before this achievement, many experts predicted that specialized algorithms capable of reading images as well as or better than human observers would dominate the future of medicine. As a result, residents and undergraduate students are becoming increasingly concerned that pursuing medicine training may be an insecure career path [6]. However, the long-awaited expansion of AI in pathology is still underway, and the area of pathology is changing at a far slower rate than other domains (eg, radiology) [7].

The recent approvals of whole slides imaging scanners by the Food and Drug Administration for primary diagnosis, as well as the approval of the prostate AI algorithm, have cleared the road for the first steps in incorporating this new technology for use in pathologic primary diagnosis. AI solutions can serve as a unique platform for breakthroughs and advancements in anatomical and clinical pathology practice [8].

Commercially available AI systems have recently become the focus of a more concentrated evaluation. However, it is unclear whether medical students and pathology residents are afraid that AI may replace pathologists or other clinicians [9]. There is limited knowledge about medical students' perspectives toward AI and deep learning, particularly in underdeveloped nations. To investigate this topic, we conducted a multicenter national survey of undergraduate medical students and pathology residents to assess their perceptions of AI in medicine in general and pathology in particular, as well as their concern that machines will replace pathologists or other physicians soon.

The perspective and understanding of medical students are critical for guiding the development of educational curricula and training. We investigate major medical AI-related attitudes among medical students in general and in pathology in Jordan, with an emphasis on the components that should be included in the medical curriculum.

Methods

Study Design and Population

A randomized, web-based, cross-sectional study was conducted among medical students and pathology residents in Jordan from 5 public universities—Jordan University of Science and Technology (JUST), The University of Jordan, Yarmouk University, Hashemite University, and Mutah University, including about 19,000 medical students—in addition to 4 pathology residency programs: JUST, King Hussein Cancer Center, Jordan University Hospital, and Royal Medical Services, including about 80 pathology residents. The survey was published over a 10-day period from March 4 to March 14, 2024. The required sample size was estimated using the web-based Raosoft sample size calculator [10]. The proposed sample size was 318, with a 5% margin of error, a 99% confidence, a 30% response rate, and a population size of 20,000. Three hundred ninety-four students completed the questionnaire. Students were encouraged to disseminate the questionnaire among their colleagues to create a snowball sample.

Ethical Considerations

The research has been approved by the institutional review board of JUST, Irbid, Jordan (number: 3/167/2024; date: February 13, 2024). All individuals consented to participate. All methods met with the applicable standards and regulations. Participants provided informed consent at the beginning of the survey and had the ability to opt out at any time. Data were anonymized and no compensation was provided to the participants. The results of this study are original, have not been previously published, are not under review elsewhere, and have received approval from all authors. All the authors have approved final version of the manuscript.

The data were obtained using a self-administered open web-based questionnaire produced with Google Forms. To evaluate the questionnaire's applicability and forward validity, our research team translated the questions into Arabic and conducted a pilot survey with 15 randomly selected participants to examine question comprehension and language clarity. The questionnaire was circulated through medical student groups, social media forums (Facebook, WhatsApp, Telegram, and Instagram), and through announcements in lectures. Participation was entirely voluntary and unrelated to the students' educational curriculum. The students consented to participate by completing the survey. Respondent anonymity was assured by design.

Questionnaire Structure

The questionnaire items were adopted from a previously validated study [11] and amended by 2 expert pathologists (AR and MA) at 2 academic centers to apply the questions to the discipline of pathology. The questionnaire was divided into sections, each of which addressed a different issue (Table S1 in Multimedia Appendix 1). The first section of the questionnaire inquired about demographic data and self-reported technological expertise. The second portion inquired about AI and deep learning applications in medicine. The third part evaluated sources to AI in general. The fourth part focused on AI applications in medicine. The fifth section evaluated emotions and perspectives toward AI and deep learning in medicine and pathology. The sixth part inquired about the expected effects of AI on medical education and specific components that should be implemented in medical education (an open-ended question), followed by a question regarding whether basic AI knowledge should be provided in official medical courses (a Yes/No). Finally, the prospective applications of AI in pathology were considered.

Statistical Analysis

Following the completion of questionnaire submissions, the findings were converted to a comma-separated value file. To

XSL•FO
RenderX

simplify statistical analysis, the categories “disagree entirely” and “rather disagree” were summarized as disagreement, while “rather agree” and “agree entirely” were summarized as agreement. Nominal categorical variables were analyzed using the Pearson chi-square test or Fisher exact test, whereas ordinal data were analyzed using Spearman correlation. The statistical analysis was performed using SPSS version 26.0 (IBM Corp) [12], and *P* values of <.05 were considered statistically significant.

Results

Overview

After 10 days of opening the survey, 394 participants completed the questionnaire (328 medical students and 66 pathology residents). Of these respondents, 49% (193/394) were males and 51% (201/394) were females. The median age is 20 (IQR 20 - 21) years. Most medical students surveyed were in their junior years (1, 2, and 3), accounting for around 85% (279/394) of the sample, while approximately 15% (49/394) were in clinical training years (4, 5, and 6). Most pathology residents surveyed were postgraduate year 1 and postgraduate year 2 (63/66, 96%). Of the total, 44.9% (177/394) of the respondents regarded themselves as technological experts (Table 1).

Table . Demographics and self-reported technical expertise (N=394).

		Participants
I consider myself a tech expert person, n/N (%)		
	Agree entirely	46/394 (12)
	Rather agree	131/394 (33.2)
	Rather disagree	73/394 (19)
	Disagree entirely	16/394 (4)
	__[a]	128/394 (32.5)
Age (years)		
	Median	20
	IQR	20-21
	Minimum/maximum	18/34
Gender, n/N (%)		
	Male	193/394 (49)
	Female	201/394 (51)

[a]Not available (or no response).

Awareness of AI and Deep Learning in Medicine

The vast majority of respondents (272/394, 69%) were previously aware of the medical community’s discussion on AI and deep learning. There was a statistically significant difference in awareness among respondents who consider themselves tech experts compared with those who do not (*P*=.03), but no significant differences were found between males and females, medical students, and pathology residents. Furthermore, 67.5% (266/394) of respondents reported having a basic awareness of the technologies used in these fields (Table 2).

Table . First part of the questionnaire—artificial intelligence and deep learning in medicine (N=394).

Questionnaire items		Yes, n/N (%)	No, n/N (%)	*P* value[a]
“Deep learning” and “artificial intelligence” are currently being broadly discussed in the medical field.				
	Were you already aware of these topics?	272/394 (69)	122/394 (31)	.03/.46
	Do you have a basic understanding of the technologies used in these topics?	266/394 (67.5)	128/394 (32.5)	.89/.94

[a]Tech expert versus non–tech expert/medical students versus pathology residents.

XSL•FO
RenderX

The Sources for the Topic of AI

Websites were the primary sources of knowledge on AI, with 73.4% (289/394) of respondents reporting awareness via this source. In contrast, fewer students heard about AI from friends and colleagues (139/394, 35.3%), webinars (59/394, 15%), and medical school lectures (56/394, 14%). The majority of respondents (310/394, 78.6%) acknowledged that they did not learn about AI through formal AI courses (Table 3).

Table . Second part of the questionnaire—different sources of exposure to artificial intelligence as a topic in general (N=394).

Questionnaire items	Yes, n/N (%)	No, n/N (%)	*P* value[a]
Other applications we use in daily life already use artificial intelligence (eg, speech-/text-recognition). Were you aware of this from?			
Websites	289/394 (73.4)	105/394 (26.6)	<.001/.66
Social friends and colleagues	139/394 (35.3)	255/394 (64.7)	.20/.74
Medical school lectures	56/394 (14)	338/394 (85.8)	.90/.28
Webinars	59/394 (15)	335/394 (85)	.80/.37
Training (eg, courses) in artificial intelligence	22/394 (6)	372/394 (94.4)	.15/.99
No answer	—[b]	—	.001/.85

[a]Tech expert vs non–tech expert/medical students vs pathology residents.

[b]Not applicable (ie, total proportion of no answers: 62/394, 16%).

The Applications of AI in Medicine

More than half of respondents thought that AI might be used to diagnose disease in patients automatically (213/394, 54.1% agreement vs 82/394, 21% disagreement), and medical students agreed more than pathology residents on this topic (*P*=.04). Furthermore, 39.5% (156/394) agreed to the use of AI in automated diagnosis. Moreover, three-quarters agreed that AI might automatically indicate appropriate investigations (318/394, 80.7% agreement vs 19/394, 5% disagreement). Table 4 provides more detailed results.

Table . Third part of the questionnaire—applications for artificial intelligence in medicine (N=394).

Questionnaire items	Agree entirely, n/N (%)	Rather agree, n/N (%)	Rather disagree, n/N (%)	Disagree entirely, n/N (%)	N/A[a], n/N (%)	*P* value[b]
What potential applications for AI in medicine do you see?						
Automated detection of disease	55/394 (14)	158/394 (40.1)	69/394 (16)	13/394 (3)	99/394 (25)	.58/.001
Automated diagnosis of patients	39/394 (10)	117/394 (29.7)	107/394 (27.2)	19/394 (5)	112/394 (28)	.016/.000
Automated indication of appropriate investigations (radiological, laboratory, etc)	113/394 (28.7)	205/394 (52)	17/394 (4)	2/394 (1)	57/394 (15)	<.001/.60

[a]N/A: not applicable.

[b]Tech expert vs non–tech expert/medical students vs pathology residents.

Overall Emotions and Attitudes About AI and Deep Learning in Both Medicine and Pathology

Regarding overall feelings and attitudes toward AI and deep learning in medicine and pathology, most respondents agreed that AI will revolutionize medicine in general (321/394, 81.4% agreement) and pathology in particular (312/394, 79.2% agreement), while a sizable proportion disagreed that human doctors in general (291/394, 73.9% disagreement) and pathologists (248/394, 63% disagreement) could be replaced in the near term. Furthermore, more than one-third of respondents expressed fear about recent AI developments (167/394, 42.4% agreed).

On the other hand, roughly half said that these breakthroughs make pathology or medicine more intriguing to them (184/394, 46.7% and 222/394, 56.4%, respectively), and for those specific questions, tech expert respondents were considerably more likely to say “yes” than non–tech expert respondents (*P*=.002 and *P*=.03, respectively).

Nonetheless, most respondents believed that the adoption of AI would benefit pathology (302/394, 76.7% agreement) and the entire field of medicine (305/394, 77.4% agreement). Notably, two-thirds of respondents disagreed that their medical schools or hospitals had educated them about AI and its uses (261/394, 66.2% disagreed), whereas 46.2% (182/394) expressed an

XSL•FO
RenderX

interest in learning the principles of AI and its applications in medicine. Table S1 in Multimedia Appendix 1 summarizes feelings and attitudes toward AI and deep learning in medicine and pathology.

What Specific Aspects Should Be Implemented in Medical Education?

Students were asked to select from a list of options, each of which allowed for several responses. The majority of respondents (231/394, 58.6%) expressed an interest in learning about AI's applications, potential hazards, and legal and ethical implications. Females were more likely than males to exhibit an interest in learning about AI's possible hazards and legal implications (*P*=.038 and *P*=.001, respectively). In addition, 51.8% (204/394) of the students felt that medical education should cover current AI systems and their technical foundations. However, 64.2% (253/394) of respondents expressed no interest in learning about the classification of AI reliability in medical education (Table 5).

Table . Specific aspects that should be implemented in medical education (multiple responses possible) (N=394)[a].

	Yes, n/N (%)	No, n/N (%)	*P* value[b]
Areas of application	231/394 (58.6)	163/394 (41.4)	.04/.95
Possible risks	231/394 (58.6)	163/394 (41.4)	.25/.83
Technical basics	204/394 (51.8)	190/394 (48.2)	.73/.74
Current AI[c] systems	204/394 (51.8)	190/394 (48.2)	.63/.52
Modes of operation	200/394 (50.8)	194/394 (49.2)	.30/.21
Legal aspects, ethics	198/394 (50.3)	196/394 (49.7)	.094/.35
Potential future developments	195/394 (49.5)	199/394 (51.5)	.45/.36
Classification of AI reliability	141/394 (35.8)	253/394 (64.2)	.77/.23
Basic AI knowledge should be provided in university courses	357/394 (90.6)	37/394 (9.4)	.21/.16

[a]The vast majority of respondents believed that university curricula should cover fundamental AI concepts (357/394, 90.6% said yes).

[b] Tech expert vs non–tech expert/medical students versus pathology residents.

[c]AI: artificial intelligence.

What Potential Applications for AI Do You See in Pathology?

In terms of pathology-specific questions, 3 quadrants (297/394, 75.4%) of respondents agreed that AI could be used to identify pathologies in slides examinations automatically, and more than half (222/394, 56.3%) agreed that AI could be used to diagnose pathologies in slides examinations and indicate appropriate further stains needed. There was a statistically significant difference between medical students, who are more inclined to agree, and pathology residents (*P*=.02/*P*<.001 and *P*<.001/*P*=.61, respectively). In addition, 79.1% (312/394) of respondents felt that AI might be used to automatically identify appropriate special studies and immunohistochemistry stains in slide examinations (Table 6).

Table . Potential applications for artificial intelligence in pathology (N=394).

	Totally agree, n/N (%)	Agree, n/N (%)	Disagree, n/N (%)	Totally disagree, n/N (%)	Neutral, n/N (%)	*P* value[a]
Automated detection of pathologies in slides examinations	91/394 (23)	206/394 (52.3)	26/394 (7)	5/394 (1)	66/394 (17)	.56/.001
Automated diagnosis in slides examinations	58/394 (15)	164/394 (41.6)	50/394 (13)	11/394 (3)	111/394 (28.2)	.02/<.001
Automated indication of appropriate special studies and immunohistochemical stains in slides examinations	103/394 (26.1)	209/394 (53)	11/394 (3)	2/394 (1)	69/394 (18)	<.001/.61

[a]Tech expert vs non–tech expert/medical students vs pathology residents.

XSL•FO
RenderX

Discussion

Principal Findings

Findings revealed that a significant majority of medical students were already aware of the ongoing discussion around AI and deep learning in the medical community. This awareness was significantly higher among those who described themselves as tech savvy. The survey results support the conclusion that the majority of respondents did not learn about AI throughout medical school.

The majority of respondents thought that AI had the potential to alter both medicine in general and pathology in particular. However, a significant proportion raised concerns about the potential displacement of human physicians by AI in the near future. This sentiment is consistent with other research revealing worries among medical students and pathology residents regarding the expanding role of AI in medicine. For example, in Lebanon and Kuwait, there was an agreement that AI would not replace doctors but rather significantly transform health care practices [13,14]. Notably, the majority of medical students who responded were in their junior years (279/394, 85%) of the total sample, which may reflect their interest in the subject compared with senior students. Although this allowed us to compare junior medical students' perceptions with those of postgraduate pathology students, it is regarded as a restriction for evaluating senior medical students' perspectives.

The general agreement was that the use of AI would benefit both pathology and the larger field of medicine [8]. Further comparison with prior work is discussed in the section "Comparison With Prior Work." This study provides useful insights into medical students' perspectives and attitudes regarding AI, which are critical for guiding the development of educational curricula and training.

The rapid growth of AI has the ability to change the face of a variety of medical specialties. Specifically in the visual medical disciplines, such as radiology, pathology, ophthalmology, and dermatology, AI has generated significant interest and will particularly affect the developments of these fields due to the visual nature of their occupations [3]. Deep learning applications are autonomously trained to execute certain tasks in response to the availability of large digital datasets. Visual activities include consultations, seminar presentations, board exams, and archiving [5]. Medical students' perspectives are crucial for shaping medical education, especially in rapidly changing professions. This study collects students' opinions on AI in medicine in order to better understand their requirements and expectations from medical schools, as well as add to the dataset so that data from various countries (both low and high income) may be compared to analyze future medical education plans. Learners in the digital era differ from past generations. They are growing increasingly technologically savvy and socially conscious. Pathology and radiology are 2 visual fields that have seen significant advancements in AI technology. Many studies have been conducted in radiology [11], but to our knowledge, this is the first to analyze the attitudes and awareness of pathology residents.

AI-Related Attitudes of Medical Students in Comparison With Pathology Residency Trainees

To the best of our knowledge, this is the first study to discuss pathology residents' awareness and attitudes toward AI in the field of pathology, in addition to comparing it with medical students' perspectives. In Jordan, digital pathology is a relatively new concept that is primarily used for research purposes. A digital scanner is available at one site in Jordan (at JUST), with a focus on research and consulting purposes.

In general, there was no major difference in the responses between undergraduate students and postgraduate pathology residents. Except on 2 occasions, medical students agreed more than pathology residents that AI can be used to diagnose disease in patients automatically (P=.04). There was also a statistically significant difference between medical students, who are more likely to agree than pathology residents regarding AI's ability to diagnose pathologies in slide examinations and indicate appropriate additional stains needed (P=.02/P<.001 and P<.001/P=.61, respectively). The Food and Drug Administration's recent support of whole-slide imaging scanners for primary diagnosis, together with the approval of prostate AI algorithms, marks the beginning of introducing AI technology into primary diagnostics. AI can provide a unique platform for promoting innovation and breakthroughs in anatomical and clinical pathology practice [9].

Otherwise, there was general agreement between medical students and pathology residents that, for example, the adoption of AI would benefit pathology and the entire field of medicine and that they needed to learn about AI's applications, potential hazards, and legal and ethical implications.

In summary, the findings underscore the imperative to integrate AI education into the medical field to adeptly equip future physicians for AI-augmented health care. Subsequent investigations should concentrate on assessing the efficacy of embedding AI education into medical training programs, both undergraduate and postgraduate, and probing the determinants influencing medical students' attitudes toward AI. AI may also be used in pathology to diagnose cancer, predict survival, modify molecular structures, and forecast treatments. More effort is required to navigate these applications. Furthermore, it is worth investigating the relationship between the extent of digital pathology implementation and AI awareness.

Limitations

Some limitations include the sample strategy's use of social media as well as the web-based approach, which limits randomization and generalization. Also, the majority of medical students who responded were in their junior years (279/394, 85%). Although this allowed us to compare junior medical students' perceptions to those of postgraduate pathology students, it is regarded as a restriction for evaluating senior medical students' perspectives.

Comparison With Prior Work

Compared with published work in this regard, similar results were identified in neighboring countries, including Lebanon [13]. In Kuwait, there was also an agreement that AI would not

XSL•FO
RenderX

replace doctors but rather significantly transform health care practices [14]. In the United Arab Emirates, there was a lack of acquaintance with AI found in a study that called for the introduction of specific education and training in medical schools [15]. In addition, identical outcomes were observed in industrialized countries. In Germany, two-thirds of students (539/838, 64.4%) believed that they were not well informed on AI in medicine, and 57.4% (463/807) thought that AI has beneficial applications in medicine, such as drug research, but less so for clinical use [16]. In the United States, a study showed that 91% (353/387) reported receiving no formal education related to AI [17].

Conclusions

Our study highlights a generally favorable attitude toward AI between medical students and pathology residents. A significant number of participants are already familiar with the ideas of AI and deep learning in medicine while the majority sees potential in AI for automated detection of pathologies and indication of appropriate investigations, which is a warning about replacing physicians and pathologists in the nearest future.

The study found that being a tech expert influenced respondents' awareness and attitudes toward AI. This indicates a potential gap in the current medical education system, with a large proportion of respondents expressing interest in learning more about AI and its applications in medicine.

In pathology, there is a prominent agreement on the potential applications of AI, particularly in the automated detection of pathologies in slide examinations and the automated indication of appropriate special studies and immunohistochemical stains. Medical students show that they are more enthusiastic about the integration of AI in pathology than pathology residents.

While most medical students and pathology residents acknowledge the potential of AI to revolutionize medicine and improve the pathology field, there is a clear need for integrating AI education into medical curricula and addressing concerns about its ethical and legal aspects.

Authors' Contributions

AR contributed to writing—original draft, methodology, formal analysis, data curation, conceptualization, and project administration. MAQ, RA, NB, and M Abdaljaleel contributed to writing—review and editing, supervision, methodology, and conceptualization. MAR, M Alkhateeb, M Abdelraheem, SAO, OBM, AA, SA, IR, and RA contributed to data curation, writing—review and editing, methodology, and conceptualization.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Fourth part of the questionnaire—feelings and attitudes toward artificial intelligence and deep learning in medicine and pathology (N=394).
[DOCX File, 17 KB - mededu_v11i1e62669_app1.docx]

References

1. Enholm IM, Papagiannidis E, Mikalef P, Krogstie J. Artificial intelligence and business value: a literature review. Inf Syst Front 2022 Oct;24(5):1709-1734. [doi: 10.1007/s10796-021-10186-w]
2. Londhe VY, Bhasin B. Artificial intelligence and its potential in oncology. Drug Discov Today 2019 Jan;24(1):228-232. [doi: 10.1016/j.drudis.2018.10.005]
3. Carlos RC, Kahn CE, Halabi S. Data science: big data, machine learning, and artificial intelligence. J Am Coll Radiol 2018;15(3 Pt B):497-498. [doi: 10.1016/j.jacr.2018.01.029]
4. Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. AJR Am J Roentgenol 2017 Apr;208(4):754-760. [doi: 10.2214/AJR.16.17224] [Medline: 28125274]
5. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. Nat Rev Clin Oncol 2019 Nov;16(11):703-715. [doi: 10.1038/s41571-019-0252-y] [Medline: 31399699]
6. @AndrewYNg. Posting with permission. AI people: what would you say to him? I will share my thoughts later. Twitter. 2017 Jul 11. URL: https://mobile.twitter.com/andrewyng/status/884810469575344128?lang=de [accessed 2024-05-25]
7. Reis-Filho JS, Kather JN. Overcoming the challenges to implementation of artificial intelligence in pathology. J Natl Cancer Inst 2023 Jun 8;115(6):608-612. [doi: 10.1093/jnci/djad048] [Medline: 36929936]
8. Shafi S, Parwani AV. Artificial intelligence in diagnostic pathology. Diagn Pathol 2023 Oct 3;18(1):109. [doi: 10.1186/s13000-023-01375-z] [Medline: 37784122]
9. Kulkarni S, Seneviratne N, Baig MS, Khan AHA. Artificial intelligence in medicine: where are we now? Acad Radiol 2020 Jan;27(1):62-70. [doi: 10.1016/j.acra.2019.10.001] [Medline: 31636002]
10. Sample size calculator. Raosoft, Inc. URL: http://www.raosoft.com/samplesize.html [accessed 2025-01-03]

XSL•FO
RenderX

11. Pinto dos Santos D, Giese D, Brodehl S, et al. Medical students' attitude towards artificial intelligence: a multicentre survey. Eur Radiol 2019 Apr;29(4):1640-1646. [doi: 10.1007/s00330-018-5601-1]
12. Arbuckle JL. Amos (Version 26.0) [computer program]. IBM SPSS. 2019. URL: https://www.ibm.com/support/pages/node/878803 [accessed 2025-01-03]
13. Doumat G, Daher D, Ghanem NN, Khater B. Knowledge and attitudes of medical students in Lebanon toward artificial intelligence: a national survey study. Front Artif Intell 2022;5:1015418. [doi: 10.3389/frai.2022.1015418]
14. Buabbas AJ, Miskin B, Alnaqi AA, et al. Investigating students' perceptions towards artificial intelligence in medical education. Healthcare (Basel) 2023 May 1;11(9):1298. [doi: 10.3390/healthcare11091298]
15. Boillat T, Nawaz FA, Rivas H. Readiness to embrace artificial intelligence among medical doctors and students: questionnaire-based study. JMIR Med Educ 2022 Apr 12;8(2):e34973. [doi: 10.2196/34973] [Medline: 35412463]
16. McLennan S, Meyer A, Schreyer K, Buyx A. German medical students´ views regarding artificial intelligence in medicine: a cross-sectional survey. PLOS Digit Health 2022 Oct;1(10):e0000114. [doi: 10.1371/journal.pdig.0000114] [Medline: 36812635]
17. Liu DS, Sawyer J, Luna A, et al. Perceptions of US medical students on artificial intelligence in medicine: mixed methods survey study. JMIR Med Educ 2022 Oct 21;8(4):e38325. [doi: 10.2196/38325] [Medline: 36269641]

Abbreviations

AI: artificial intelligence
JUST: Jordan University of Science and Technology

Edited by B Lesselroth; submitted 28.05.24; peer-reviewed by A Khamees, Y Chong; revised version received 21.06.24; accepted 23.11.24; published 10.01.25.

Please cite as:
Rjoop A, Al-Qudah M, Alkhasawneh R, Bataineh N, Abdaljaleel M, Rjoub MA, Alkhateeb M, Abdelraheem M, Al-Omari S, Bani-Mari O, Alkabalan A, Altulaih S, Rjoub I, Alshimi R
Awareness and Attitude Toward Artificial Intelligence Among Medical Students and Pathology Trainees: Survey Study
JMIR Med Educ 2025;11:e62669
URL: https://mededu.jmir.org/2025/1/e62669
doi:10.2196/62669

© Anwar Rjoop, Mohammad Al-Qudah, Raja Alkhasawneh, Nesreen Bataineh, Maram Abdaljaleel, Moayad A Rjoub, Mustafa Alkhateeb, Mohammad Abdelraheem, Salem Al-Omari, Omar Bani-Mari, Anas Alkabalan, Saoud Altulaih, Iyad Rjoub, Rula Alshimi. Originally published in JMIR Medical Education (https://mededu.jmir.org), 10.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Reviewing Mobile Apps for Teaching Human Anatomy: Search and Quality Evaluation Study

Guadalupe Esmeralda Rivera García[1*], PhD; Miriam Janet Cervantes López[2*], PhD; Juan Carlos Ramírez Vázquez[1*], PhD; Arturo Llanes Castillo[2*], PhD; Jaime Cruz Casados[2*], MAS

[1]Tecnológico Nacional de México, Instituto Tecnológico Superior de Pánuco, Pánuco, Veracruz, Mexico

[2]Facultad de Medicina de Tampico "Dr. Alberto Romo Caballero" de la Universidad Autónoma de Tamaulipas, Tampico, Tamaulipas, Mexico

[*]all authors contributed equally

Corresponding Author:
Guadalupe Esmeralda Rivera García, PhD
Tecnológico Nacional de México, Instituto Tecnológico Superior de Pánuco
Av. Artículo Tercero Constitucional, Colonia Solidaridad
Pánuco, Veracruz, 93990
Mexico
Phone: 52 846 1064254
Email: esmeralda.rivera@itspanuco.edu.mx

Abstract

Background: Mobile apps designed for teaching human anatomy offer a flexible, interactive, and personalized learning platform, enriching the educational experience for both students and health care professionals.

Objective: This study aimed to conduct a systematic review of the human anatomy mobile apps available on Google Play, evaluate their quality, highlight the highest scoring apps, and determine the relationship between objective quality ratings and subjective star ratings.

Methods: The Mobile App Rating Scale (MARS) was used to evaluate the apps. The intraclass correlation coefficient was calculated using a consistency-type 2-factor random model to measure the reliability of the evaluations made by the experts. In addition, Pearson correlations were used to analyze the relationship between MARS quality scores and subjective evaluations of MARS quality item 23.

Results: The mobile apps with the highest overall quality scores according to the MARS (ie, sections A, B, C, and D) were Organos internos 3D (anatomía) (version 4.34), Sistema óseo en 3D (Anatomía) (version 4.32), and VOKA Anatomy Pro (version 4.29). To measure the reliability of the MARS quality evaluations (sections A, B, C, and D), the intraclass correlation coefficient was used, and the result was "excellent." Finally, Pearson correlation results revealed a significant relationship (r=0.989; P<.001) between the quality assessments conducted by health care professionals and the subjective evaluations of item 23.

Conclusions: The average evaluation results of the selected apps indicated a "good" level of quality, and those with the highest ratings could be recommended. However, the lack of scientific backing for these technological tools is evident. It is crucial that research centers and higher education institutions commit to the active development of new mobile health apps, ensuring their accessibility and validation for the general public.

(JMIR Med Educ 2025;11:e64550) doi:10.2196/64550

KEYWORDS

anatomy; Google Play; mobile health; mHealth; Mobile App Rating Scale; MARS

Introduction

Background

At present, there is a wealth of research on mobile apps focused on various aspects and areas of health, such as musculoskeletal injuries [1]; chronic disease management [2]; pediatric disease care [3]; medication management [4]; oral hygiene [5]; asthma [6]; pediatric ear, nose, and throat surgery [7]; low back pain [8]; neurodegenerative disorders [9]; coronary arteries [10]; neurorehabilitation [11]; nutrition, anemia, and preeclampsia [12]; cancer [13,14]; cerebrovascular diseases [15]; childhood obesity [16]; diabetes [17]; tuberculosis [18]; fibromyalgia [19]; dementia [20]; chronic kidney disease [21]; and epilepsy [22],

XSL•FO
RenderX

among others. These mobile health management apps have transformed the way people access and manage information about their well-being, enabling everything from vital signs monitoring to chronic disease management. This advancement in mobile apps not only benefits patients but also opens new possibilities in health education. In particular, mobile apps for teaching human anatomy have become valuable resources that complement the learning and understanding of the structures and functions of the human body. Similar to personal health apps, these tools are designed with interactive features such as 3D models and detailed simulations that enhance the educational experience for medical students and health care professionals.

The justification for this study is based on three important approaches: (1) the technological approach regarding the use of mobile apps for teaching human anatomy, (2) the pedagogical approach, and (3) the quality evaluation approach of the apps. These approaches together provide a solid foundation to justify the importance of the study.

Technological Approach

The first approach involved reviewing and analyzing scientific publications on the use of mobile apps in teaching human anatomy, with the aim of understanding their results. This approach highlights how mobile technology has revolutionized access to knowledge, enabling students to learn in an accessible and practical manner. In teaching human anatomy, mobile apps with 3D models and simulations facilitate immersive and effective learning, complementing and even enhancing traditional methods in health sciences. One of the studies presented the software Road to Birth, developed by the University of Newcastle, which was designed to teach midwifery students at a Midwestern US university about the dynamic concepts of maternal anatomy and physiology during an obstetrics module. The students used Road to Birth, and 66% of them reported an increase in their knowledge, valuing the software as a useful and practical learning resource [23]. Another study used the mobile app AR in Anatomy, developed by the authors; which allows users to dynamically explore various parts of the human body in 3D, enhancing the educational experience [24]. Similarly, apps such as Anat_Hub, developed by faculty and researchers from the Departments of Computer Science and Medical Sciences at the University of the Western Cape; a mobile app with augmented reality (AR) to improve learning about the musculoskeletal system's anatomy, received positive evaluations. User results indicated that the anatomy system could effectively enhance student engagement and retention of anatomical concepts [25]. In addition, another study conducted with undergraduate health sciences students at the University of Cape Town analyzed the impact of an AR mobile app on learning motivation. The study included 78 students, evaluating motivation levels before and after using the app. The results showed that its use increased motivation, improving aspects such as attention, satisfaction, and confidence [26]. In the field of neuroanatomy, the use of mobile AR facilitated the understanding of complex concepts, increasing academic performance and reducing cognitive load among students [27]. Similarly, HuMAR, developed by researchers affiliated with Murdoch University and Universiti Utara Malaysia; an AR-based prototype for learning skeletal structure, demonstrated high satisfaction among students, highlighting the system's usability and functionality [28]. However, although cadavers remain the gold standard in anatomy teaching, there are financial, ethical, and supervisory limitations. Another study compared the effectiveness of virtual reality, AR, and tablet-based devices in teaching cranial anatomy. A total of 59 students participated who were randomly assigned to one of the 3 learning methods. The results suggest that these technologies can effectively complement anatomical teaching [29]. On the other hand, a recent study presented a human anatomy learning system based on AR using a marker on a mobile platform to capture images and merge them with data from an SQLite database. This system allows for interactive visualization of the human body or its organs in 3D. An evaluation conducted with high school and medical students demonstrated that the app facilitates anatomy learning more effectively due to its ability to provide interactive 3D representations through AR [30]. Another example is AEducaAR, an app developed by researchers affiliated with the University of Bologna, which combines AR with a 3D-printed anatomical model to improve anatomy teaching for medical students. Its effectiveness was evaluated with a group of 62 second-year students, comparing its use to traditional learning methods with anatomical atlas books. Although there were no significant differences in objective test results between the two methods, students expressed enthusiasm for AEducaAR in a survey, valuing its potential to motivate learning and enhance the 3D understanding of anatomical structures. This tool could also prepare students to use advanced medical technologies in their future careers [31]. In addition, a relevant study presents 10 mobile apps for teaching human anatomy, where the results indicate that the technological designs studied exhibit a high degree of usability [32]. Another study analyzed 325 anatomy mobile apps and outlined their features to facilitate dissemination in the academic field. It showcases a broad, diverse, and affordable market for human anatomy mobile apps that can complement students' education [33].

Pedagogical Approach

Medical students frequently face challenges in understanding anatomy through the images found in textbooks [34], which are flat and lack interactivity. In contrast, mobile apps for learning human anatomy can serve as a complementary resource for learning this discipline, offering students the opportunity to interact with content more deeply than in a conventional dissection room. Although traditional cadaver-based teaching remains the preferred learning method [35], anatomy education continues to face numerous challenges, including limited practical hours for students and instructors, restricted access, and the high cost of cadavers and artificial models [36]. The use of mobile apps for learning human anatomy offers significant advantages, such as immediate and continuous access to information anytime and anywhere. These apps include interactive 3D designs that allow for a detailed exploration of the human body. Furthermore, they are often updated regularly and are generally more affordable than traditional textbooks, making them accessible to a broader audience.

XSL•FO
RenderX

Quality Evaluation Approach

The third approach focuses on evaluating the quality of human anatomy mobile apps. The importance of evaluating these apps lies in the fact that higher-rated apps can serve as academic support for medical students, health care professionals, and general users. A specific methodology is required to evaluate the quality of health mobile apps. In this context, the Mobile App Rating Scale (MARS) methodology has been used. Various studies have used the MARS to assess health apps targeting a variety of conditions, such as chronic kidney disease and end-stage renal disease [37], chronic lung diseases [38], stress management [39], psoriasis [40], gastrointestinal diseases [41], pain management [42], oral hygiene [43,44], nutrition [45], genetics and genomics [46], food allergies or intolerances [47], deafness and hearing impairment [48], low back pain [49,50], neurological conditions [51], peritoneal dialysis [52], diabetes [53], COVID-19 [54], cancer [55], anticoagulation [56], dementia [57], specialized diets [58], toric intraocular lenses [59], epilepsy [60], depression [61], coronary diseases [62], dyslexia [63], autism spectrum disorder [64], nutrition [65], and pediatric palliative care [66].

This study aimed to (1) identify human anatomy mobile apps available on the Google Play store, which uses the Android operating system, covering 70.87% of the global mobile operating system market [67]; (2) evaluate these apps using the MARS, which considers engagement, functionality, aesthetics, information, subjective quality, and app specificity; (3) present the human anatomy mobile apps with the highest ratings on the MARS; and (4) determine the correlation between the objective MARS quality rating and the subjective MARS rating by health care professionals (ie, item 23).

Methods

Overview

This study was cross-sectional, as it collected data at a single time point without follow-up over time, evaluated the correlation between variables, and provided descriptive data [68-70]. The study was conducted following the guidelines of the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) initiative, which aims to improve the communication of results from observational studies among authors, editors, and readers of scientific publications, focusing primarily on cohort, case-control, and cross-sectional studies [71].

Selection Criteria for Human Anatomy Mobile Apps

From April 1 to May 30, 2024, an extensive search for mobile apps was conducted. The search term used was *anatomy*. The inclusion criteria for mobile apps were as follows: (1) available on the Google Play store; (2) related to human anatomy; (3) available in English or Spanish; (4) user rating ≥4.3 to ensure a minimum level of acceptance and satisfaction among users; (5) free to use—an essential factor in educational contexts where students or universities may face budget constraints; and (6) download count exceeding 100,000.

The exclusion criteria were as follows: (1) duplicate apps, either because of different versions or alternative names but containing the same content; and (2) apps not updated for 2 or more years.

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology was used to select human anatomy–teaching apps. This methodology is used in systematic reviews and meta-analyses to ensure transparency and rigor in the selection and analysis of relevant studies. The phases were identification, selection, eligibility, and inclusion.

During the identification phase, an exhaustive search for mobile apps dedicated to teaching human anatomy was conducted. In the filtering phase, duplicate apps were discarded. In the eligibility phase, the characteristics of these apps were analyzed to discard those that did not meet the previously established inclusion criteria, and an additional exclusion criterion was applied. Finally, in the inclusion phase, the apps that met all eligibility requirements were integrated for analysis and evaluation.

All selected apps were recorded in Excel (version 2016; Microsoft Corporation) with the following characteristics: app name, identification screen, languages, star rating, total downloads, developer, Android version, last update date, and features.

Evaluation of Mobile Apps

Overview

We used the MARS, developed by Stoyanov et al [72], which has been widely used to evaluate the design and usability of mobile health apps. The MARS consists of 3 dimensions. The first dimension is an objective tool based on 4 main components: engagement (section A), functionality (section B), aesthetics (section C), and information quality (section D). The second dimension assesses subjective quality (section E), and the third dimension evaluates the perceived effectiveness (section F).

Each section of the MARS has several items. An item refers to a specific element, question, or unit of evaluation within a questionnaire or survey. Section A comprises 5 items and evaluates whether the app is engaging, interesting, customizable, interactive, and targeted at a specific population. Section B comprises 4 items and focuses on the app's performance, ease of use, navigation, and gesture design. Section C comprises 3 items and examines the app's design, graphics, and visual appeal. Section D comprises 7 items and analyzes the accuracy of information description, objectives, quality and quantity of information, visual information quality, credibility, and scientific evidence base of the evaluated app. The 4 objective sections of MARS (ie, A, B, C, and D) encompass 19 items. The average scores of sections A, B, C, and D represent the overall MARS quality score [72].

Section E comprises 4 items and focuses on the evaluator's personal perception of the app and typically includes items that ask about the likelihood of recommending the app, the probability of the user continuing to use the app, and the overall perception of its quality. Finally, section F comprises 6 items and focuses on how health care professionals perceive the impact

XSL•FO
RenderX

of the app on their knowledge, attitudes, intentions, and behaviors related to health.

Evaluation Instrument Scale

To evaluate each item, a 5-point Likert scale was used, ranging from 1 to 5 (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent). A MARS score of more than 3 points indicates acceptable quality.

Selection of Evaluators

A total of 10 evaluators were selected, all health care professionals from the Faculty of Medicine of Tampico Dr Alberto Romo Caballero at the Universidad Autónoma de Tamaulipas, Tampico, Tamaulipas, Mexico. The inclusion criteria were (1) being a top-performing student in the final year of medical school and (2) having a mobile phone with the Android operating system to download human anatomy teaching apps from Google Play.

Evaluation Process

Before starting the evaluation of anatomy-related apps, it was necessary to train the evaluators in the use of the MARS. For this, the authors of this study convened the 10 selected evaluators at the library of the Faculty of Medicine of Tampico Dr Alberto Romo Caballero at the Universidad Autónoma de Tamaulipas, Tampico, Tamaulipas, Mexico, to present a training video in English by Stoyanov et al [72]. Following the video presentation, a training exercise was suggested for all evaluators using an app called Anatomymaster. This app, also focused on human anatomy, was not included in the study sample, as it did not meet the requirement of having a user rating ≥4.3.

The evaluators downloaded and tested the trial app for at least 10 minutes before completing the MARS web-based questionnaire. If any individual evaluation score varied by at least 2 points, the evaluators discussed it to reach a consensus, ensuring a uniform understanding of each item.

After completing the trial exercise, the 10 selected evaluators evaluated the 18 human anatomy mobile apps during June 2024. Each evaluator was provided with a list of apps, which they downloaded and used for 10 minutes before completing a web-based evaluation instrument designed based on the MARS. Each item from the different sections was rated using a Likert scale (1-5). The collected data were initially recorded in Excel.

Statistical Analysis

Overview

Statistical analyses, including the calculation of the intraclass correlation coefficient (ICC) and Pearson correlation coefficient, were performed using SPSS (version 29.0.2.0; IBM Corp).

Intraclass Correlation Coefficient

The ICC was used using a random effects model of 2 factors with consistency to measure the overall agreement between the quantitative measurements obtained by different evaluators [73]. The ICC ranges from 0 to 1, with 0 indicating a total lack of reliability among evaluators and 1 representing perfect reliability. According to the 95% CI for ICC estimation, values below 0.5 are considered “poor” reliability, those between 0.5 and 0.75 are considered “moderate,” those between 0.75 and 0.9 are considered “good,” and those above 0.90 are rated as “excellent” [74].

The individual ICC was calculated for each of the sections A, B, C, and D. The arithmetic means of each section were used to calculate the ICC for the overall MARS quality score (ie, sections A, B, C, and D). In section A, 900 data points were considered, accounting for 10 evaluators, 5 items, and 18 mobile apps. In section B, 720 data points were used (ie, 10 evaluators, 4 items, and 18 mobile apps). In section C, 540 data points were analyzed (ie, 10 evaluators, 3 items, and 18 mobile apps). Finally, in section D, 1080 data points were considered (ie, 10 evaluators, 6 items, and 18 mobile apps). In this last section, item 19 was excluded because of missing values; therefore, only 6 items were considered instead of 7. For each MARS section (ie, A, B, C, and D), the arithmetic mean and SD were calculated.

Pearson Correlation

The statistical technique of Pearson correlation was used to evaluate the relationship between the MARS quality scores (ie, sections A, B, C, and D) and the subjective item 23 from section E. The software used was SPSS.

Ethical Considerations

This study did not involve experiments on humans, animals, or the collection of sensitive personal data. It focused exclusively on evaluating publicly available mobile applications on Google Play through a structured analysis conducted by health care professionals. No direct interaction with developers or users of the applications took place, and no private or identifiable information was accessed or stored.

This type of research, which does not involve mental health e-communities or sensitive data, does not require institutional review board approval. Furthermore, the activities were carried out following the institutional policies and local guidelines of the Universidad Autónoma de Tamaulipas, Mexico, for observational research and technological product reviews.

This decision aligns with international ethical standards and aims to ensure transparency and accountability in the evaluation of publicly available technological products.

Results

Selection Criteria for Anatomy Mobile Apps

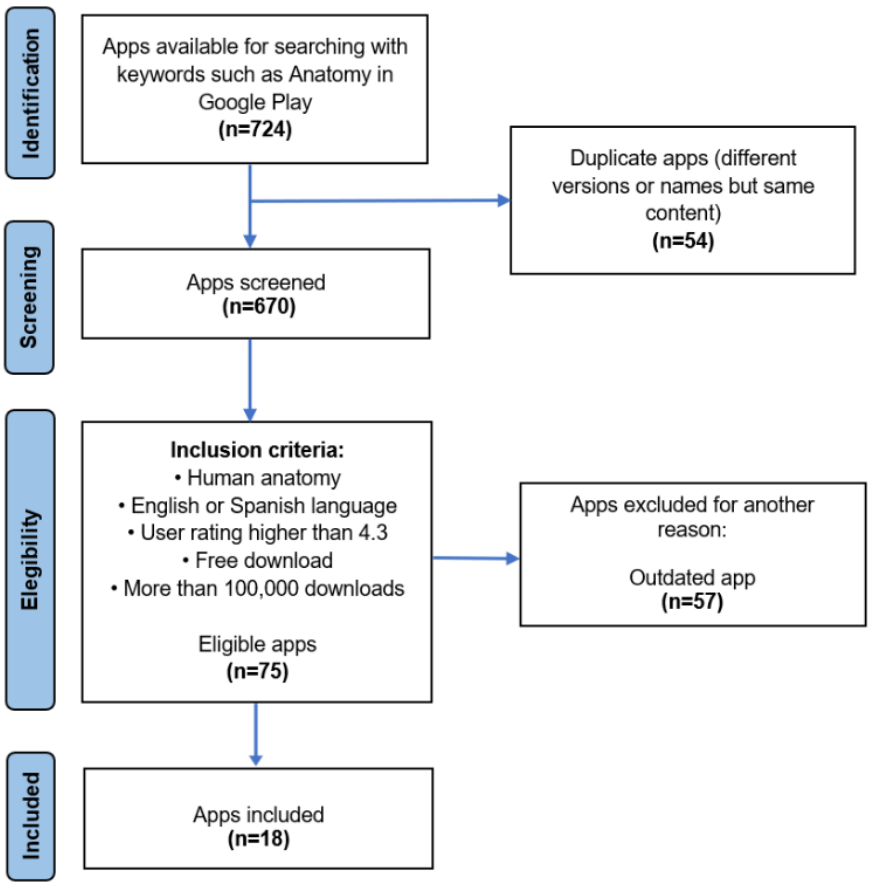
Figure 1 shows the PRISMA diagram applied to the selection of mobile apps; a total of 724 apps were identified in the Google Play store under the search criterion *anatomy*. In total, 54 duplicate apps were eliminated either because they had the same name or different names but the same content, leaving 670 in the screening phase. In the eligibility phase, the inclusion criteria were applied, where only 75 apps met these characteristics. In the same way, 57 apps that were more than 2 years old without receiving updates from the developer were excluded, finally leaving 18 apps in the inclusion phase.

In Table 1, the names of the selected apps, their identification screens, developer names, required Android operating system

XSL•FO
RenderX

version, and the date of the last update are presented. All the listed mobile apps focus on human anatomy, are available in English or Spanish, have a user rating above 4.3, can be downloaded for free, and have more than 100,000 downloads.

Figure 1. Flowchart of the selection process.



XSL•FO
RenderX

Table 1. Characteristics of selected mobile apps (Google Play, 2024).

App name	Identification screen	Developer	Android version	Last update
Anatomy Learning-Anatomía 3D		3D Medical OU	7.0 and later versions	June 16, 2024
Complete Anatomy 2024		3D4Medical from Elsevier	7.0 and later versions	May 28, 2024
Biodigital Human-3D Anatomy		BioDigital	5.0 and later versions	February 28, 2024
Anatomía–Atlas 3D		Catfish Animation Studio	8.0 and later versions	August 21, 2023
Anatomyka-Anatomy 3D		Woodoo Art s.r.o.	5.1 and later versions	November 17, 2023
VOKA 3d Anatomy and Physiology		Factory of innovations and solutions LLC	8.0 and later versions	January 27, 2024
Organos internos 3D (anatomía)		Ing Víctor Michel González Galván	5.1 and later versions	November 1, 2023
Teach Me Anatomy		TeachMeSeries Ltd	7.0 and later versions	May 22, 2024
3D Bones and Organs (Anatomy)		Education Mobile	5.1 and later versions	September 3, 2023
Esqueleto|Anatomía 3D		Catfish Animation Studio	8.0 and later versions	August 21, 2023
Visual Anatomy Lite		Education Mobile	4.4 and later versions	August 9, 2023
Gray's Anatomy-Anatomy Atlas		SEStudio	4.4 and later versions	March 1, 2023
El cuerpo humano en 3D		Mozaik Education	5.0 and later versions	May 30, 2024
e-Anatomy		IMAIOS SAS	5.0 and later versions	June 11, 2024
Sistema muscular 3D (Anatomía)		Ing. Víctor Michel González Galván	5.1 and later versions	November 27, 2023
Sistema óseo en 3D (Anatomía)		Ing. Víctor Michel González Galván	5.1 and later versions	November 6, 2023
Anatomy by Muscle & Motion		Muscle and Motion	5.0 and later versions	April 2, 2024
Flashcards de Daily Anatomy		Kenhub	5.0 and later versions	September 21, 2023

Evaluation of Mobile Apps

MARS Overall Quality Scores

The average overall MARS quality score (ie, sections A, B, C, and D) was rated as "good" (mean 4.02, SD 0.20) [75]. The 3 mobile apps with the highest overall MARS quality scores (ie, averages of sections A, B, C, and D) were Organos internos 3D (anatomía) (mean 4.34, SD 0.29), Sistema óseo en 3D (Anatomía) (mean 4.32, SD 0.28), and VOKA Anatomy Pro (mean 4.29, SD 0.28). In contrast, the apps with the lowest overall MARS quality scores were Anatomy–3D Atlas (mean 3.66, SD 0.27), Complete Anatomy 2024 (mean 3.73, SD 0.32), and Visual Anatomy Lite (mean 3.80, SD 0.28). The average overall MARS quality scores are listed in Table 2.

XSL•FO
RenderX

Table 2. Average Mobile App Rating Scale quality scores (sections A, B, C, and D).

App name	Section A—engagement (mean 3.69, SD 0.20), mean (SD)	Section B—functionality (mean 4.36, SD 0.22), mean (SD)	Section C— aesthetics (mean 4.14, SD 0.21), mean (SD)	Section D—information (mean 3.90, SD 0.23), mean (SD)	Arithmetic average A, B, C, and D (mean 4.02, SD 0.20), mean (SD)
Organos internos 3D (anatomía)	3.96 (0.19)	4.65 (0.19)	4.43 (0.23)	4.30 (0.13)	4.34 (0.29)
Sistema óseo en 3D (Anatomía)	3.98 (0.28)	4.63 (0.19)	4.47 (0.15)	4.22 (0.17)	4.32 (0.28)
VOKA 3 d Anatomy and Physiology	3.94 (0.11)	4.60 (0.24)	4.37 (0.12)	4.23 (0.05)	4.29 (0.28)
Anatomy Learning-Anatomía 3D	3.88 (0.15)	4.55 (0.13)	4.27 (0.31)	4.20 (0.24)	4.22 (0.28)
Flashcards de Daily Anatomy	3.82 (0.28)	4.55 (0.10)	4.30 (0.20)	4.05 (0.10)	4.18 (0.32)
Teach Me Anatomy	3.76 (0.24)	4.48 (0.15)	4.40 (0.20)	3.95 (0.15)	4.15 (0.35)
Anatomyka- Anatomía 3D	3.74 (0.23)	4.50 (0.14)	4.23 (0.42)	4.00 (0.11)	4.12 (0.32)
3D Bones and organs (Anatomy)	3.72 (0.33)	4.53 (0.05)	4.03 (0.12)	3.90 (0.18)	4.04 (0.35)
El cuerpo humano en 3D	3.64 (0.21)	4.38 (0.21)	4.20 (0.26)	3.92 (0.10)	4.03 (0.32)
Sistema muscular 3D (Anatomía)	3.84 (0.27)	4.33 (0.15)	4.13 (0.42)	3.83 (0.05)	4.03 (0.24)
Biodigital Human-3D Anatomy	3.66 (0.31)	4.30 (0.08)	4.10 (0.44)	3.80 (0.06)	3.97 (0.29)
Anatomy by Muscle & Motion	3.66 (0.24)	4.45 (0.13)	4.07 (0.38)	3.60 (0.13)	3.94 (0.40)
e-Anatomy	3.54 (0.36)	4.20 (0.28)	4.07 (0.15)	3.87 (0.08)	3.92 (0.29)
Gray's Anatomy-Anatomy Atlas	3.58 (0.32)	4.18 (0.29)	3.93 (0.06)	3.72 (0.08)	3.85 (0.26)
Esqueleto|Anatomía 3D	3.48 (0.13)	4.10 (0.34)	4.00 (0.10)	3.75 (0.15)	3.83 (0.28)
Visual Anatomy Lite	3.44 (0.05)	4.08 (0.35)	3.97 (0.06)	3.70 (0.06)	3.80 (0.28)
Complete Anatomy 2024	3.34 (0.09)	4.05 (0.31)	3.93 (0.06)	3.58 (0.08)	3.73 (0.32)
Anatomía–Atlas 3D	3.36 (0.05)	4.00 (0.08)	3.70 (0.30)	3.57 (0.23)	3.66 (0.27)

MARS Section Scores (Sections A, B, C, and D)

The mean (SD) for the "engagement" section (ie, section A) was 3.69 (0.20). The 3 top-rated apps in this section were Sistema óseo en 3D (Anatomía) (mean 3.98, SD 0.28), Organos internos 3D (anatomía) (mean 3.96, SD 0.19), and VOKA Anatomy Pro (mean 3.94, SD 0.19). In contrast, the 3 apps with the lowest scores in section A were Complete Anatomy 2024 (mean 3.34, SD 0.09), Anatomy–3D Atlas (mean 3.36, SD 0.05), and Visual Anatomy Lite (mean 3.44, SD 0.05).

For the "functionality" section (ie, section B), the mean (SD) was 4.36 (0.22). The top-rated apps were Organos internos 3D (anatomía) (mean 4.65, SD 0.19), Sistema óseo en 3D (Anatomía) (mean 4.63, SD 0.19), and VOKA Anatomy Pro (mean 4.60, SD 0.24). Conversely, the 3 apps with the lowest scores in this section were Anatomy–3D Atlas (mean 4.00, SD 0.08), Complete Anatomy 2024 (mean 4.05, SD 0.31), and Visual Anatomy Lite (mean 4.08, SD 0.35).

Regarding the "aesthetics" section (ie, section C), the mean (SD) was 4.14 (0.21). The highest-rated apps were Sistema óseo en 3D (Anatomía) (mean 4.47, SD 0.15), Organos internos 3D (anatomía) (mean 4.43, SD 0.23), and Teach Me Anatomy (mean 4.40, SD 0.20). In contrast, the lowest-scoring apps in this section were Anatomy–3D Atlas (mean 3.70, SD 0.30), Complete Anatomy 2024 (mean 3.93, SD 0.06), and Gray's Anatomy-Anatomy Atlas (mean 3.93, SD 0.06).

For the "information quality" section (ie, section D), the mean (SD) was 3.90 (0.23). The highest-rated apps were Organos internos 3D (anatomía) (mean 4.30, SD 0.13), VOKA Anatomy Pro (mean 4.23, SD 0.05), and Sistema óseo en 3D (Anatomía) (mean 4.22, SD 0.17). In contrast, the lowest-scoring apps in this section were Anatomy–3D Atlas (mean 3.57, SD 0.23), Complete Anatomy 2024 (mean 3.58, SD 0.08), and Anatomy by Muscle and Motion (mean 3.60, SD 0.13; Table 2).

Subjective Quality Evaluation (Section E) and Perceived Effectiveness (Section F)

The general mean (SD) for the "subjective quality" section (ie, section E) was 3.63 (0.22). The 3 top-rated mobile apps in this section were VOKA Anatomy Pro (mean 3.95, SD 0.10), Organos internos 3D (anatomía) (mean 3.93, SD 0.17), and Sistema óseo en 3D (Anatomía) (mean 3.88, SD 0.22). Conversely, the 3 apps with the lowest scores in this section were Anatomy–Atlas 3D (mean 3.25, SD 0.10), Complete Anatomy 2024 (mean 3.28, SD 0.15), and Esqueleto|Anatomía 3D (mean 3.35, SD 0.19). The average scores for section E are listed in Table 3.

Regarding the "perceived effectiveness" section (ie, section F), the recorded mean (SD) was 3.65 (0.18). The 3 top-rated apps with the highest scores were Organos internos 3D (anatomía) (mean 3.93, SD 0.10), VOKA Anatomy Pro (mean 3.90, SD 0.11), and Sistema óseo en 3D (Anatomía) (mean 3.87, SD 0.15). In contrast, the 3 apps with the lowest scores in this

XSL•FO
RenderX

section were Complete Anatomy 2024 (mean 3.37, SD 0.20), Anatomy–Atlas 3D (mean 3.40, SD 0.13), and Visual Anatomy Lite (mean 3.45, SD 0.15). The average scores for section F are listed in Table 4.

The section with the highest score was "functionality" (ie, section B), with a mean (SD) of 4.36 (0.22), followed by "aesthetics" (ie, section C), which scored a mean (SD) of 4.14 (0.21). In the third place was "information quality" (ie, section D), with a mean (SD) of 3.90 (0.22), followed by "engagement" (ie, section A), with a mean (SD) of 3.69 (0.20). The fifth position corresponded to "perceived effectiveness" (ie, section F), with a mean (SD) of 3.65 (0.18), while the sixth and final position was occupied by "subjective quality" (ie, section E), with a mean (SD) of 3.63 (0.22). It is notable that the app-specific score (ie, section F) was higher than the subjective quality score (ie, section E), although the latter was lower than the overall MARS quality score (mean 4.02, SD 0.20). This is demonstrated in Figure 2.

Table 3. Average score for "subjective quality" (section E).

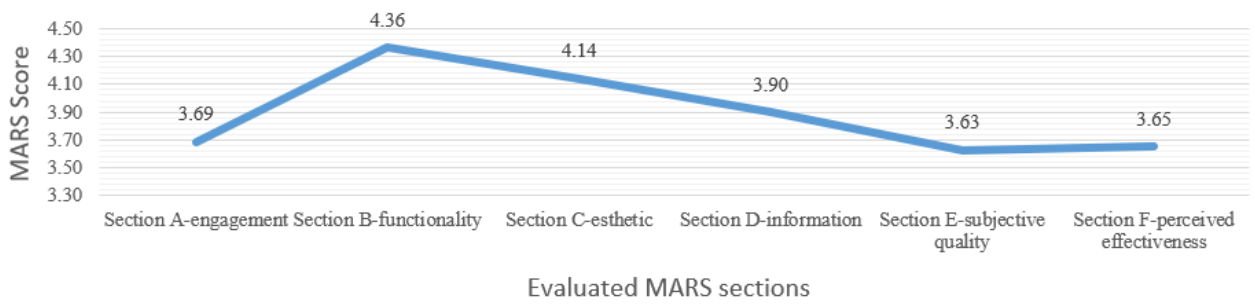
App name	Section E—subjective quality (mean 3.63, SD 0.22), mean (SD)
VOKA 3d Anatomy and Physiology	3.95 (0.10)
Organos internos 3D (anatomía)	3.93 (0.17)
Sistema óseo en 3D (Anatomía)	3.88 (0.22)
Anatomy Learning-Anatomía 3D	3.85 (0.10)
Sistema muscular 3D (Anatomía)	3.83 (0.17)
Flashcards de Daily Anatomy	3.78 (0.10)
Teach Me Anatomy	3.75 (0.10)
Anatomyka-Anatomía 3D	3.73 (0.15)
Anatomy by Muscle & Motion	3.63 (0.15)
El cuerpo humano en 3D	3.60 (0.14)
3D Bones and organs (Anatomy)	3.58 (0.17)
Biodigital Human-3D Anatomy	3.55 (0.19)
Gray's Anatomy-Anatomy Atlas	3.53 (0.13)
e-Anatomy	3.48 (0.17)
Visual Anatomy Lite	3.40 (0.14)
Esqueleto|Anatomía 3D	3.35 (0.19)
Complete Anatomy 2024	3.28 (0.15)
Anatomía–Atlas 3D	3.25 (0.10)

XSL•FO
RenderX

Table 4. Average score for "perceived effectiveness" (section F).

App name	Section F—perceived effectiveness (mean 3.65, SD 0.18), mean (SD)
Organos internos 3D (anatomía)	3.93 (0.10)
VOKA 3d Anatomy and Physiology	3.90 (0.11)
Sistema óseo en 3D (Anatomía)	3.87 (0.15)
Anatomy Learning-Anatomía 3D	3.82 (0.21)
Flashcards de Daily Anatomy	3.80 (0.14)
Sistema muscular 3D (Anatomía)	3.78 (0.12)
3D Bones and organs (Anatomy)	3.73 (0.21)
Anatomyka-Anatomía 3D	3.72 (0.08)
Teach Me Anatomy	3.70 (0.17)
Anatomy by Muscle & Motion	3.60 (0.11)
El cuerpo humano en 3D	3.58 (0.08)
Biodigital Human-3D Anatomy	3.57 (0.15)
Gray's Anatomy-Anatomy Atlas	3.55 (0.22)
Esqueleto|Anatomía 3D	3.48 (0.35)
e-Anatomy	3.47 (0.15)
Visual Anatomy Lite	3.45 (0.15)
Anatomía–Atlas 3D	3.40 (0.13)
Complete Anatomy 2024	3.37 (0.20)

Figure 2. Average scores by the Mobile App Rating Scale (MARS) sections.



Average Scores by Section and Item

The following section details the items that received the highest and lowest scores in the various evaluated sections. In section A, item 5 concerning the target population received a mean score of 3.87 (SD 0.24), whereas item 4 related to interactivity scored a mean value of 3.52 (SD 0.20). In section B, item 6 on performance received the highest score with a mean value of 4.57 (0.14), followed by item 8 related to navigation with a mean score of 4.24 (SD 0.23). In section C, item 11 concerning graphics received a mean score of 4.37 (SD 0.28), and item 12 on visual appeal scored a mean value of 4.02 (SD 0.26). In section D, item 17 on the quality of visual information received a mean score of 3.98 (SD 0.24), whereas item 18 on credibility scored a mean value of 3.83 (SD 0.31). In section E, item 23 on overall quality received a mean score of 3.79 (SD 0.20), and item 22 on willingness to pay for the app scored a mean value of 3.52 (SD 0.24). Finally, in section F, the item with the highest overall mean score was "attitudes," with a mean value of 3.77 (SD 0.23), whereas the item "help seeking" received the lowest score (mean 3.54, SD 0.23). The mean scores with respect to the section and items are listed in Table 5.

XSL•FO
RenderX

Table 5. Mean scores by section and item.

Section and item	Scores, mean (SD)
Section A—engagement (mean 3.69, SD 0.20)	
Item 1. Entertainment	3.67 (0.35)
Item 2. Interests	3.78 (0.28)
Item 3. Customization	3.59 (0.20)
Item 4. Interactivity	3.52 (0.20)
Item 5. Target population	3.87 (0.24)
Section B—functionality (mean 4.36, SD 0.22)	
Item 6. Performance	4.57 (0.14)
Item 7. Ease of use	4.37 (0.37)
Item 8. Navigation	4.24 (0.23)
Item 9. Gestural design of the app	4.27 (0.23)
Section C—aesthetics (mean 4.14, SD 0.21)	
Item 10. Design	4.04 (0.19)
Item 11. Graphics	4,37 (0.28)
Item 12. Visual appeal	4.02 (0.26)
Section D—quality of information (mean 3.90, SD 0.23)	
Item 13. Accuracy of information description	3.86 (0.23)
Item 14. Objectives	3.95 (0.25)
Item 15. Quality of information	3.87 (0.24)
Item 16. Amount of information	3.91 (0.26)
Item 17. Quality of visual information	3.98 (0.24)
Item 18. Credibility	3.83 (0.31)
Item 19. Evidence base	__[a]
Section E—subjective quality (mean 3.63, SD 0.22)	
Item 20. Would you recommend this app?	3.58 (0.26)
Item 21. How many times would you use this app?	3.61 (0.23)
Item 22. Would you pay for this app?	3.52 (0.24)
Item 23. General qualifications	3.79 (0.20)
Section F—perceived effectiveness (mean 3.65, SD 0.18)	
Awareness	3.67 (0.27)
Knowledge	3.64 (0.22)
Attitudes	3.77 (0.23)
Intention to change	3.57 (0.22)
Help seeking	3.54 (0.23)
Behavior change	3.71 (0.17)

[a]No apps presented explicit scientific support in the descriptions and comments.

MARS Overall Quality Scores and Star Rating (Item 23)

The overall MARS quality scores were higher than the scores for item 23 (ie, subjective quality). Similarly, the overall MARS star ratings (ie, item 23) were lower than the star ratings in the Google Play store (Table 6).

XSL•FO
RenderX

Table 6. Overall MARSa quality scores, overall star ratings for item 23, and star ratings in the Google Play store.

App name	Health professionals		Users	
	MARS overall quality score (mean 4.02)	MARS (item 23; mean 3.79)	Star rating in the Google Play store (mean 4.63)	Downloads
Organos internos 3D (anatomía)	4.34	4.10	4.90	>5,000,000
Sistema óseo en 3D (Anatomía)	4.32	4.10	4.90	>1,000,000
VOKA 3d Anatomy and physiology	4.29	4.10	4.80	>100,000
Anatomy Learning-Anatomía 3D	4.22	4.00	4.80	>10,000,000
Flashcards de Daily Anatomy	4.18	3.90	4.80	>500,000
Teach Me Anatomy	4.15	3.90	4.70	>1,000,000
Anatomyka-Anatomía 3D	4.12	3.90	4.70	>500,000
3D Bones and organs (Anatomy)	4.04	3.80	4.70	>1,000,000
El cuerpo humano en 3D	4.03	3.80	4.60	>1,000,000
Sistema muscular 3D (Anatomía)	4.03	3.80	4.80	>1,000,000
Biodigital Human-3D Anatomy	3.97	3.70	4.60	>500,000
Anatomy by Muscle & Motion	3.94	3.70	4.60	>500,000
e-Anatomy	3.92	3.70	4.50	>1,000,000
Gray's Anatomy-Anatomy Atlas	3.85	3.70	4.60	>1,000,000
Esqueleto|Anatomía 3D	3.83	3.60	4.40	>1,000,000
Visual Anatomy Lite	3.80	3.60	4.40	>1,000,000
Complete Anatomy 2024	3.73	3.50	4.30	>1,000,000
Anatomía–Atlas 3D	3.66	3.40	4.30	>1,000,000

[a]MARS: Mobile App Rating Scale.

Statistical Analysis

ICC (Assessment Reliability)

The average reliability measures of the evaluation ranged from "good" to "excellent." In the engagement section (ie, section A), an ICC of 0.892 (95% CI 0.807-0.952) was obtained. In the functionality section (ie, section B), the ICC was 0.901 (95% CI 0.822-0.956). In the aesthetics section (ie, section C), an ICC of 0.866 (95% CI 0.758-0.941) was recorded. In the information quality section (ie, section D), the ICC was 0.890 (95% CI 0.804-0.951). In the subjective quality section (ie, section E), an ICC of 0.862 (95% CI 0.751-0.939) was obtained. Finally, in the app specificity section (ie, section F), an ICC of 0.868 (95% CI 0.764-0.941) was recorded. Similarly, the reliability of the overall MARS quality evaluation (ie, average of sections A, B, C, and D) was classified as "excellent," with an ICC of 0.912 (95% CI 0.820-0.963).

Pearson Correlation

For the calculation of Pearson correlation coefficient, the average MARS quality scores and the scores for subjective item 23 from section E presented earlier in Table 6 were considered. The result showed an excellent correlation (r=0.989, P<.001; Table 7). Also, the 95% CI for this correlation was 0.971 to 0.996, based on the Fisher r-to-z transformation.

XSL•FO
RenderX

Table 7. Pearson correlation results.

Correlations	MARS overall quality score	MARS (item 23)
MARS overall quality score		
Pearson correlation	1	0.989[a]
P value (bilateral)	__[b]	<.001
N	19	19
MARS (item 23)		
Pearson correlation	0.989[a]	1
P value (bilateral)	<.001	—
N	19	19

[a]The correlation is significant at the .01 level (2 sided).

[b]Not applicable.

Discussion

Overview

The primary objective of this study was to identify and assess the quality of mobile apps related to human anatomy available on Google Play using the MARS. This scale focuses on the usability and accessibility of mobile health apps, considering aspects such as engagement, functionality, aesthetics, information quality, subjective quality, and app specificity. The MARS organizes the evaluations of the apps into 3 different dimensions. The first dimension includes sections A, B, C, and D and focuses on the evaluation of the objective technical items. The evaluations in the second and third dimensions are subjective and are divided into 2 sections: section E, which considers the evaluator's personal appreciation, and section F, which focuses on the perceived effectiveness. These 3 dimensions are crucial because, while mobile health apps must meet functionality and design standards, the evaluator's perception and the app's impact are determinants for its adoption. In addition, the MARS sections cannot be considered in isolation, as they are interrelated and influence each other.

Principal Findings

In the first dimension of the MARS, the best-rated section was "functionality" with a mean score of 4.36, followed by "aesthetics" with a mean score of 4.14, "information quality" with a mean score of 3.90, and, finally, "engagement" with a mean score of 3.69, which was the least valued. Although the apps generally received a good average score, it is crucial to examine the relatively low ratings in fundamental aspects such as engagement (mean score 3.69) and look for solutions. In order to strengthen the engagement section (ie, section A), which is made up of items 1 to 5 of the MARS (ie, entertainment, interest, personalization, interactivity, and target population), specific recommendations can be applied for each item. In item 1 (ie, entertainment), it is suggested to integrate elements such as gamification, challenges, achievements, rewards, or progressive levels and use good quality graphics, attractive colors, animations, and multimedia content (eg, videos and music) to make the user experience more attractive. In item 2 (ie, interest), it is recommended to include new content and personalized reminders. For item 3 (ie, personalization), it is recommended that users be able to adjust themes, difficulty levels, colors, or display modes according to their preferences, as well as the use of artificial intelligence to offer suggestions based on the user's preferences. Regarding item 4 (ie, interactivity), it is suggested to incorporate interactive content such as questionnaires and practical activities that require active participation, as well as real-time communication to forums, live chats, or social interactions to encourage collaboration between users and provide immediate feedback to correct errors or recognize achievements. Finally, in item 5 (ie, target population), it is recommended to carry out previous studies on the characteristics and needs of the target population, such as age, educational level, and cultural context, and ensuring that the content, graphics, and design are consistent with the population for which the app was designed.

The average general quality score according to the MARS (ie, sections A, B, C, and D) was good (mean score 4.02), supported by excellent reliability with an ICC of 0.912 and a 95% CI of 0.820 to 0.963. The mobile apps that excelled in overall quality according to the MARS (ie, sections A, B, C, and D) were Organos internos 3D (anatomía) with a mean score of 4.34, Sistema óseo en 3D (Anatomía) with a mean score of 4.32, and VOKA Anatomy Pro with a mean score of 4.29. These results indicate that these apps, having received high scores and offering high-quality content, can be recommended for users interested in learning human anatomy. In the second and third dimensions of the MARS, corresponding to sections E and F, where the impact on the user is more significant, the lowest average scores were recorded: subjective quality with a mean score of 3.63 and app specificity with a mean score of 3.65. These ratings were even lower than the general quality score of the MARS, which was 4.02.

These results underscore the importance of conducting a thorough analysis of all the 3 dimensions of the MARS; otherwise, apps that are technically well developed might be overvalued, whereas those that receive better subjective ratings from users could be overlooked. This indicates that developers of human anatomy mobile apps should not only address aspects of functionality, aesthetics, engagement, and information but also actively consider user perception and the impact of their

XSL•FO
RenderX

apps. There are various practical uses of the study's results, such as a more appropriate selection of mobile apps in the student context or in medical practice, where those that obtained a higher score in the MARS evaluation are chosen, which provide greater reliability and comfort in use.

Another relevant point of discussion is that the MARS, specifically item 19 (ie, section D), which addresses "information quality," assesses whether mobile apps have scientific foundations that support their usefulness. However, the apps evaluated in this study did not present explicit scientific support in the descriptions and comments provided by the developers, which is why they lack a rating in item 19 of the MARS evaluation.

Therefore, it is crucial that research centers and universities get involved in the development of mobile health apps so that they are supported by scientific research and can be hosted in app stores to make them accessible to the general public. Collaboration among software developers, health professionals, researchers, and academics in the creation and review of educational materials for a medical mobile app would generate greater confidence in its use. In addition, conducting validation studies in real learning environments also plays an important role in assessing the quality and effectiveness of apps through various methodologies, such as the MARS framework discussed in this study.

Limitations

The main limitations are the exclusion of paid apps, apps in languages other than English or Spanish, and apps with a star ratings less than 4.3. In addition, the search was limited to apps present in the Google Play store. Although these criteria may seem restrictive, English is the predominant language in global medical education, ensuring that the evaluated apps covered a substantial portion of the app market. However, the exclusions may limit the scope, particularly by omitting paid apps, which in certain cases may offer higher-quality content that could facilitate and enhance the learning of anatomy. To address these limitations in future research, inclusion criteria could be expanded to incorporate human anatomy mobile apps available in other languages or those that are paid, creating a broader repertoire for analysis. This approach would also enable comparative studies, such as exploring potential differences between free apps and those requiring a license or payment.

Conclusions

This study provides a comprehensive and detailed analysis of apps available for teaching human anatomy, aimed at health care professionals, medical students, and interested users. For example, students and health professionals can both use a human anatomy mobile app before orthopedic surgery to consult a 3D model of the leg of a patient with a femur fracture. This would allow them to more accurately understand the location of bones, blood vessels, and muscles in the affected region, contributing to greater success in the procedure.

Overall, the evaluated apps demonstrated high quality, particularly excelling in functionality and aesthetic design. However, some apps need to improve aspects such as user engagement (ie, section A) and the quality of the information provided (ie, section D). Among the highest-rated apps according to the MARS are Organos internos 3D (anatomía), Sistema óseo en 3D (Anatomía), and VOKA Anatomy Pro.

The subjective MARS score (ie, item 23) was 3.79, in contrast to the average rating of 4.63 given by users on the Google Play store. This suggests that evaluators provided lower ratings, whereas users tend to overrate the apps. This discrepancy may stem from the fact that evaluators typically adhere to more rigorous and objective criteria, systematically assessing technical, functional, and usability aspects. Professional evaluators are often more critical regarding technical implementation and practical utility.

In contrast, users base their ratings on personal and subjective experiences, scoring according to their expectations and the level of satisfaction experienced while using the app. Both perspectives offer valuable feedback: on the one hand, an objective evaluation of quality, and on the other hand, a subjective evaluation of user satisfaction. This difference in ratings does not negatively impact the overall MARS evaluation of the apps. Instead, it provides a perspective where both developers and potential users can identify strengths and areas for improvement from complementary approaches.

This study highlights the evolving role of mobile apps as transformative tools in medical education by offering innovative solutions for accessibility and interactivity in learning. Mobile apps use advanced features such as 3D models, simulations, and dynamic interfaces, and these tools overcome the limitations of traditional methods of teaching human anatomy, such as the scarcity of cadavers and high costs of dissection laboratories. In addition, they facilitate personalized learning of topics and selection of difficulty levels. They allow continuous access, allowing students to practice and reinforce their knowledge anytime, anywhere.

To maximize the impact of mobile apps in medical education, we suggest strategies focused on design and functionality, such as the incorporation of gamification elements, challenges, and rewards to increase user motivation, as well as strengthening interactivity through real-time feedback, collaborative learning tools, and interactive clinical cases. It is essential to align the content of the apps with medical education curricula to ensure their relevance and applicability. Similarly, we recommend combining their use with traditional methods, such as face-to-face classes and laboratory practice, to offer a comprehensive learning experience. Training teachers to integrate these tools into their teaching methodologies is also essential. Finally, to guarantee both the scientific rigor and the accessibility of these mobile apps, we propose collaboration with universities and research centers to develop content based on solid scientific evidence.

XSL•FO
RenderX

Data Availability

The data used in this study are available upon request from the corresponding author.

Conflicts of Interest

None declared.

References

1. Fan K, Zhao Y. Mobile health technology: a novel tool in chronic disease management. Intell Med 2022 Feb;2(1):41-47 [FREE Full text] [doi: 10.1016/j.imed.2021.06.003]
2. Opipari-Arrigan L, Dykes DMH, Saeed SA, Thakkar S, Burns L, Chini BA, et al. Technology-enabled health care collaboration in pediatric chronic illness: pre-post interventional study for feasibility, acceptability, and clinical impact of an electronic health record-linked platform for patient-clinician partnership. JMIR Mhealth Uhealth 2020 Nov 26;8(11):e11968 [FREE Full text] [doi: 10.2196/11968] [Medline: 33242014]
3. Tabi K, Randhawa A, Choi F, Mithani Z, Albers F, Schnieder M, et al. Mobile apps for medication management: review and analysis. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e13608 [FREE Full text] [doi: 10.2196/13608] [Medline: 31512580]
4. Ryan S, Ní Chasaide N, O' Hanrahan S, Corcoran D, Caulfield B, Argent R. mHealth apps for musculoskeletal rehabilitation: systematic search in app stores and content analysis. JMIR Rehabil Assist Technol 2022 Aug 01;9(3):e34355 [FREE Full text] [doi: 10.2196/34355] [Medline: 35916688]
5. Zahid T, Alyafi R, Bantan N, Alzahrani R, Elfirt E. Comparison of effectiveness of mobile app versus conventional educational lectures on oral hygiene knowledge and behavior of high school students in Saudi Arabia. Patient Prefer Adherence 2020;14:1901-1909 [FREE Full text] [doi: 10.2147/PPA.S270215] [Medline: 33116434]
6. Caponnetto P, Prezzavento G, Casu M, Polosa R, Quattropani M. Psychological factors, digital health technologies, and best asthma management as three fundamental components in modern care: a narrative review. Appl Sci 2024 Apr 16;14(8):3365 [FREE Full text] [doi: 10.3390/app14083365]
7. Dobrina R, Starec A, Brunelli L, Orzan E, De Vita C, Bicego L, et al. Applying the participatory slow design approach to a mHealth application for family caregivers in pediatric ear, nose, and throat surgery. Healthcare (Basel) 2024 Feb 08;12(4):442 [FREE Full text] [doi: 10.3390/healthcare12040442] [Medline: 38391818]
8. López-Marcos JJ, Díaz-Arribas MJ, Valera-Calero JA, Navarro-Santana MJ, Izquierdo-García J, Ortiz-Gutiérrez RM, et al. The added value of face-to-face supervision to a therapeutic exercise-based app in the management of patients with chronic low back pain: a randomized clinical trial. Sensors 2024 Jan 16;24(2):567 [FREE Full text] [doi: 10.3390/s24020567] [Medline: 38257659]
9. Giannopoulou P, Vrahatis A, Papalaskari MA, Vlamos P. The RODI mHealth app insight: machine-learning-driven identification of digital indicators for neurodegenerative disorder detection. Healthcare (Basel) 2023 Nov 19;11(22):2985 [FREE Full text] [doi: 10.3390/healthcare11222985] [Medline: 37998477]
10. Boszko M, Krzowski B, Peller M, Hoffman P, Żurawska N, Skoczylas K, et al. Impact of AfterAMI mobile app on quality of life, depression, stress and anxiety in patients with coronary artery disease: open label, randomized trial. Life (Basel) 2023 Oct 05;13(10):2015 [FREE Full text] [doi: 10.3390/life13102015] [Medline: 37895396]
11. Sánchez-Rodríguez MT, Pinzón-Bernal MY, Jiménez-Antona C, Laguarta-Val S, Sánchez-Herrera-Baeza P, Fernández-González P, et al. Designing an informative app for neurorehabilitation: a feasibility and satisfaction study by physiotherapists. Healthcare (Basel) 2023 Sep 14;11(18):2549 [FREE Full text] [doi: 10.3390/healthcare11182549] [Medline: 37761746]
12. Choudhury A, Shahsavar Y, Sarkar K, Choudhury M, Nimbarte A. Exploring perceptions and needs of mobile health interventions for nutrition, anemia, and preeclampsia among pregnant women in underprivileged Indian communities: a cross-sectional survey. Nutrients 2023 Aug 24;15(17):3699 [FREE Full text] [doi: 10.3390/nu15173699] [Medline: 37686731]
13. Martínez C, Martínez A, Uclés V, Jenkins M, Badilla A. Aplicaciones móviles para la rehabilitación de pacientes con cáncer de mama. Iberian Mag Inf Syst Technol 2022;50:310-321 [FREE Full text]
14. Wang L, Langlais CS, Kenfield SA, Chan JM, Graff RE, Allen IE, et al. mHealth interventions to promote a healthy diet and physical activity among cancer survivors: a systematic review of randomized controlled trials. Cancers (Basel) 2022 Aug 06;14(15):3816 [FREE Full text] [doi: 10.3390/cancers14153816] [Medline: 35954479]
15. Elbagoury B, Vladareanu L, Vlădăreanu V, Salem A, Travediu AM, Roushdy M. A Hybrid Stacked CNN and residual feedback GMDH-LSTM deep learning model for stroke prediction applied on mobile AI smart hospital platform. Sensors (Basel) 2023 Mar 27;23(7):350 [FREE Full text] [doi: 10.3390/s23073500] [Medline: 37050561]
16. Zarkogianni K, Chatzidaki E, Polychronaki N, Kalafatis E, Nicolaides N, Voutetakis A, et al. The ENDORSE feasibility study: exploring the use of m-Health, artificial intelligence and serious games for the management of childhood obesity. Nutrients 2023 Mar 17;15(6):1451 [FREE Full text] [doi: 10.3390/nu15061451] [Medline: 36986180]
17. Ali Sherazi B, Laeer S, Krutisch S, Dabidian A, Schlottau S, Obarcanin E. Functions of mHealth diabetes apps that enable the provision of pharmaceutical care: criteria development and evaluation of popular apps. Int J Environ Res Public Health 2022 Dec 21;20(1):64 [FREE Full text] [doi: 10.3390/ijerph20010064] [Medline: 36612402]

XSL•FO
RenderX

18. Needamangalam Balaji J, Prakash S, Park Y, Baek J, Shin J, Rajaguru V, et al. A scoping review on accentuating the pragmatism in the implication of mobile health (mHealth) technology for tuberculosis management in India. J Pers Med 2022 Sep 28;12(10):1599 [FREE Full text] [doi: 10.3390/jpm12101599] [Medline: 36294738]
19. Miró J, Lleixà-Daga M, de la Vega R, Llorens-Vernet P, Jensen M. A mobile application to help self-manage pain severity, anxiety, and depressive symptoms in patients with fibromyalgia syndrome: a pilot study. Int J Environ Res Public Health 2022 Sep 23;19(19):12026 [FREE Full text] [doi: 10.3390/ijerph191912026] [Medline: 36231327]
20. Rashid N, Chen X, Mohamad Marzuki MF, Takshe AA, Okasha A, Maarof F, et al. Development and usability assessment of a mobile app (demensia KITA) to support dementia caregivers in Malaysia: a study protocol. Int J Environ Res Public Health 2022 Sep 20;19(19):11880 [FREE Full text] [doi: 10.3390/ijerph191911880] [Medline: 36231181]
21. Teong LF, Khor BH, Radion Purba K, Gafor A, Goh BL, Bee BC, et al. A mobile app for triangulating strategies in phosphate education targeting patients with chronic kidney disease in Malaysia: development, validation, and patient acceptance. Healthcare (Basel) 2022 Mar 14;10(3):535 [FREE Full text] [doi: 10.3390/healthcare10030535] [Medline: 35327013]
22. Abreu M, Carmo A, Franco A, Parreira S, Vidal B, Costa M, et al. Mobile applications for epilepsy: where are we? Where should we go? A systematic review. Signals 2022 Feb 03;3(1):40-65 [FREE Full text] [doi: 10.3390/signals3010005]
23. Hutchcraft ML, Wallon RC, Fealy SM, Jones D, Galvez R. Evaluation of the road to birth software to support obstetric problem-based learning education with a cohort of pre-clinical medical students. Multimodal Technol Interact 2023;7(8):84 [FREE Full text] [doi: 10.3390/mti7080084]
24. Patil S, Rao A, Pardeshi N, Gavhane M, Waghole D. Augmented reality in anatomy. Int J Eng Res Technol 2021;10(4):55 [FREE Full text]
25. Boomgaard A, Fritz K, Isafiade O, Kotze R, Ekpo O, Smith M, et al. A novel immersive anatomy education system (Anat_Hub): redefining blended learning for the musculoskeletal system. Appl Sci 2022 Jun 03;12(11):5694 [FREE Full text] [doi: 10.3390/app12115694]
26. Khan T, Johnston J, Ophoff J. The impact of an augmented reality application on learning motivation of students. Adv Hum Comput Interact 2019 Feb 03;2019:1-14 [FREE Full text] [doi: 10.1155/2019/7208494]
27. Küçük S, Kapakin S, Göktaş Y. Learning anatomy via mobile augmented reality: effects on achievement and cognitive load. Anat Sci Educ 2016 Oct;9(5):411-421 [FREE Full text] [doi: 10.1002/ase.1603] [Medline: 26950521]
28. Jamali SS, Shiratuddin MF, Wong K, Oskam CL. Utilising mobile-augmented reality for learning human anatomy. Procedia Soc Behav Sci 2015 Jul;197:659-668 [FREE Full text] [doi: 10.1016/j.sbspro.2015.07.054]
29. Moro C, Štromberga Z, Raikos A, Stirling A. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. Anat Sci Educ 2017 Nov;10(6):549-559 [FREE Full text] [doi: 10.1002/ase.1696] [Medline: 28419750]
30. Kurniawan M, Suharjito, Diana, Witjaksono G. Human anatomy learning systems using augmented reality on mobile application. Procedia Comput Sci 2018;135:80-88 [FREE Full text] [doi: 10.1016/j.procs.2018.08.152]
31. Cercenelli L, De Stefano A, Billi A, Ruggeri A, Marcelli E, Marchetti C, et al. AEducaAR, anatomical education in augmented reality: a pilot experience of an innovative educational tool combining AR technology and 3D printing. Int J Environ Res Public Health 2022 Jan 18;19(3):1024 [FREE Full text] [doi: 10.3390/ijerph19031024] [Medline: 35162049]
32. Martínez G, Mir F, Garcia L. Caracterización de las aplicaciones móviles para la enseñanza y el aprendizaje de la anatomía humana. Enseñanza de las ciencias. Núm. Extra 0; 2017. URL: https://ddd.uab.cat/record/184395 [accessed 2024-04-27]
33. Montaner Sanchis A, Gumbau Puchol V, Villalba Ferrer F, Eleuterio Cerveró G. Mobile learning en la anatomía humana: estudio del mercado de aplicaciones. Educ Médica 2022 Mar;23(2):100726 [FREE Full text] [doi: 10.1016/j.edumed.2022.100726]
34. Wainman B, Wolak L, Pukas G, Zheng E, Norman G. The superiority of three-dimensional physical models to two-dimensional computer presentations in anatomy learning. Med Educ 2018 Nov;52(11):1138-1146 [FREE Full text] [doi: 10.1111/medu.13683] [Medline: 30345680]
35. Havens K, Saulovich N, Saric K. A case report about anatomy applications for a physical therapy hybrid online curriculum. J Med Libr Assoc 2020 Apr;108(2):295-303 [FREE Full text] [doi: 10.5195/jmla.2020.825] [Medline: 32256241]
36. Lopes I, Teixeira B, Cortez P, de Silva G, de Sousa Neto A, de Sousa Leal N. Use of human cadavers in teaching of human anatomy in Brazilian medical faculties. Acta Sci Biol Sci 2017 May 03;39(1):1 [FREE Full text] [doi: 10.4025/actascibiolsci.v39i1.33860]
37. Siddique A, Krebs M, Alvarez S, Greenspan I, Patel A, Kinsolving J, et al. Mobile apps for the care management of chronic kidney and end-stage renal diseases: systematic search in app stores and evaluation. JMIR Mhealth Uhealth 2019 Sep 04;7(9):e12604 [FREE Full text] [doi: 10.2196/12604] [Medline: 31486408]
38. Quach S, Michaelchuk W, Benoit A, Oliveira A, Packham TL, Goldstein R, et al. Mobile heath applications for self-management in chronic lung disease: a systematic review. Netw Model Anal Health Inform Bioinform 2023;12(1):25 [FREE Full text] [doi: 10.1007/s13721-023-00419-0] [Medline: 37305790]
39. Paganini S, Meier E, Terhorst Y, Wurst R, Hohberg V, Schultchen D, et al. Stress management apps: systematic search and multidimensional assessment of quality and characteristics. JMIR Mhealth Uhealth 2023 Aug 29;11:e42415 [FREE Full text] [doi: 10.2196/42415] [Medline: 37642999]

XSL•FO
RenderX

40. Lull C, von Ahnen JA, Gross G, Olsavszky V, Knitza J, Leipe J, et al. German mobile apps for patients with psoriasis: systematic search and evaluation. JMIR Mhealth Uhealth 2022 May 26;10(5):e34017 [FREE Full text] [doi: 10.2196/34017] [Medline: 35617014]
41. Messner E, Sturm N, Terhorst Y, Sander LB, Schultchen D, Portenhauser A, et al. Mobile apps for the management of gastrointestinal diseases: systematic search and evaluation within app stores. J Med Internet Res 2022 Oct 05;24(10):e37497 [FREE Full text] [doi: 10.2196/37497] [Medline: 36197717]
42. Salazar A, de Sola H, Failde I, Moral-Munoz J. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. JMIR Mhealth Uhealth 2018 Oct 25;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]
43. Sharif M, Alkadhimi A. Patient focused oral hygiene apps: an assessment of quality (using MARS) and knowledge content. Br Dent J 2019 Sep 13;227(5):383-386 [FREE Full text] [doi: 10.1038/S41415-019-0665-0]
44. Kanoute A, Carrouel F, Gare J, Dieng S, Dieng A, Diop M, et al. Evaluation of oral hygiene-related mobile apps for children in sub-Saharan Africa. Int J Environ Res Public Health 2022 Oct 01;19(19):12565 [FREE Full text] [doi: 10.3390/ijerph191912565] [Medline: 36231862]
45. Martinon P, Saliasi I, Bourgeois D, Smentek C, Dussart C, Fraticelli L, et al. Nutrition-related mobile apps in the French app stores: assessment of functionality and quality. JMIR Mhealth Uhealth 2022 Mar 14;10(3):e35879 [FREE Full text] [doi: 10.2196/35879] [Medline: 35285817]
46. Talwar D, Tseng T, Foster M, Xu L, Chen L. Genetics/genomics education for nongenetic health professionals: a systematic literature review. Genet Med 2017 Jul;19(7):725-732 [FREE Full text] [doi: 10.1038/gim.2016.156] [Medline: 27763635]
47. Mandracchia F, Llauradó E, Tarro L, Valls R, Solà R. Mobile phone apps for food allergies or intolerances in app stores: systematic search and quality assessment using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2020 Sep 16;8(9):e18339 [FREE Full text] [doi: 10.2196/18339] [Medline: 32936078]
48. Romero R, Kates F, Hart M, Ojeda A, Meirom I, Hardy S. Quality of deaf and hard-of-hearing mobile apps: evaluation using the Mobile App Rating Scale (MARS) with additional criteria from a content expert. JMIR Mhealth Uhealth 2019 Oct 30;7(10):e14198 [FREE Full text] [doi: 10.2196/14198] [Medline: 31670695]
49. Escriche-Escuder A, De-Torres I, Roldán-Jiménez C, Martín-Martín J, Muro-Culebras A, González-Sánchez M, et al. Assessment of the quality of mobile applications (apps) for management of low back pain using the Mobile App Rating Scale (MARS). Int J Environ Res Public Health 2020 Dec 09;17(24):9209 [FREE Full text] [doi: 10.3390/ijerph17249209] [Medline: 33317134]
50. Scala L, Giglioni G, Bertazzoni L, Bonetti F. The efficacy of the smartphone app for the self-management of low back pain: a systematic review and assessment of their quality through the Mobile Application Rating Scale (MARS) in Italy. Life (Basel) 2024 Jun 13;14(6):760 [FREE Full text] [doi: 10.3390/life14060760] [Medline: 38929744]
51. Rendell R, Pinheiro M, Wang B, McKay F, Ewen A, Carnegie C, et al. Digital apps to improve mobility in adults with neurological conditions: a health app-focused systematic review. Healthcare (Basel) 2024 Apr 30;12(9):929 [FREE Full text] [doi: 10.3390/healthcare12090929] [Medline: 38727486]
52. Chao SM, Wang ML, Fang YW, Lin ML, Chen SF. Mobile apps for patients with peritoneal dialysis: systematic app search and evaluation. Healthcare (Basel) 2024 Mar 25;12(7):719 [FREE Full text] [doi: 10.3390/healthcare12070719] [Medline: 38610142]
53. Lee J, Kim Y. Evaluation of mobile applications for patients with diabetes mellitus: a scoping review. Healthcare (Basel) 2024 Jan 31;12(3):368 [FREE Full text] [doi: 10.3390/healthcare12030368] [Medline: 38338253]
54. Holl F, Schobel J, Swoboda W. Mobile apps for COVID-19: a systematic review of reviews. Healthcare (Basel) 2024 Jan 08;12(2):139 [FREE Full text] [doi: 10.3390/healthcare12020139] [Medline: 38255029]
55. Wasserman S, Ould Brahim L, Attiya A, Belzile E, Lambert S. An evaluation of interactive mHealth applications for adults living with cancer. Curr Oncol 2023 Jul 25;30(8):7151-7166 [FREE Full text] [doi: 10.3390/curroncol30080518] [Medline: 37622999]
56. Wang SW, Chiou CC, Su CH, Wu CC, Tsai SC, Lin TK, et al. Measuring mobile phone application usability for anticoagulation from the perspective of patients, caregivers, and healthcare professionals. Int J Environ Res Public Health 2022 Aug 16;19(16):10136 [FREE Full text] [doi: 10.3390/ijerph191610136] [Medline: 36011765]
57. Kuo HL, Chang CH, Ma WF. A survey of mobile apps for the care management of patients with dementia. Healthcare (Basel) 2022 Jun 23;10(7):1173 [FREE Full text] [doi: 10.3390/healthcare10071173] [Medline: 35885700]
58. McAleese D, Linardakis M, Papadaki A. Quality and presence of behaviour change techniques in mobile apps for the Mediterranean diet: a content analysis of android Google Play and Apple app store apps. Nutrients 2022 Mar 18;14(6):1290 [FREE Full text] [doi: 10.3390/nu14061290] [Medline: 35334947]
59. Scantling-Birch Y, Naveed H, Mukhija R, Nanavaty MA. A review of smartphone apps used for Toric intraocular lens calculation and alignment. Vision (Basel) 2022 Feb 18;6(1):13 [FREE Full text] [doi: 10.3390/vision6010013] [Medline: 35225972]
60. Mohammadzadeh N, Khenarinezhad S, Ghazanfarisavadkoohi E, Safari MS, Pahlevanynejad S. Evaluation of m-Health applications use in epilepsy: a systematic review. Iran J Public Health 2021 Mar;50(3):459-469 [FREE Full text] [doi: 10.18502/ijph.v50i3.5586] [Medline: 34178793]

XSL•FO
RenderX

61. Martín-Martín J, Muro-Culebras A, Roldán-Jiménez C, Escriche-Escuder A, De-Torres I, González-Sánchez M, et al. Evaluation of Android and Apple Store depression applications based on mobile application rating scale. Int J Environ Res Public Health 2021 Nov 27;18(23):12505 [FREE Full text] [doi: 10.3390/ijerph182312505] [Medline: 34886232]
62. Mack C, Terhorst Y, Stephan M, Baumeister H, Stach M, Messner EM, et al. "Help in a heartbeat?": A systematic evaluation of mobile health applications (apps) for coronary heart disease. Int J Environ Res Public Health 2021 Sep 30;18(19):10323 [FREE Full text] [doi: 10.3390/ijerph181910323] [Medline: 34639623]
63. Larco A, Carrillo J, Chicaiza N, Yanez C, Luján-Mora S. Moving beyond limitations: designing the Helpdys app for children with dyslexia in rural areas. Sustainability 2021 Jun 24;13(13):7081 [FREE Full text] [doi: 10.3390/su13137081]
64. Lian X, Sunar M. Mobile augmented reality technologies for autism spectrum disorder interventions: a systematic literature review. Appl Sci 2021 May 17;11(10):4550 [FREE Full text] [doi: 10.3390/app11104550]
65. Choi J, Chung C, Woo H. Diet-related mobile apps to promote healthy eating and proper nutrition: a content analysis and quality assessment. Int J Environ Res Public Health 2021 Mar 28;18(7):3496 [FREE Full text] [doi: 10.3390/ijerph18073496] [Medline: 33800531]
66. Weekly T, Walker N, Beck J, Akers S, Weaver M. A review of apps for calming, relaxation, and mindfulness interventions for pediatric palliative care patients. Children (Basel) 2018 Jan 26;5(2):16 [FREE Full text] [doi: 10.3390/children5020016] [Medline: 29373515]
67. Mobile operating system market share worldwide. StatCounter. URL: https://gs.statcounter.com/os-market-share/mobile/worldwide [accessed 2024-04-29]
68. Manterola C, Zavando D, Cartes-Velásquez R, Otzen T, Sanhueza A. Initial validation of a scale to measure methodological quality in prognosis studies. The MInCir proposal. Int J Morphol 2018 Jun;36(2):762-767 [FREE Full text] [doi: 10.4067/S0717-95022018000200762]
69. Wang X, Cheng Z. Cross-sectional studies: strengths, weaknesses, and recommendations. Chest 2020 Jul;158(1S):S65-S71 [FREE Full text] [doi: 10.1016/j.chest.2020.03.012] [Medline: 32658654]
70. Manterola C, Hernández-Leal M, Otzen T, Espinosa M, Grande L. Estudios de Corte Transversal. Un Diseño de Investigación a Considerar en Ciencias Morfológicas. Int J Morphol 2023 Feb;41(1):146-155 [FREE Full text] [doi: 10.4067/S0717-95022023000100146]
71. von Elm E, Altman D, Egger M, Pocock S, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007 Oct 20;335(7624):806-808 [FREE Full text] [doi: 10.1136/bmj.39335.541782.AD] [Medline: 17947786]
72. Stoyanov S, Hides L, Kavanagh D, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
73. Martínez JA, Pérez PS. Intraclass correlation coefficient. Semergen 2023;49(23):101907 [FREE Full text] [doi: 10.1007/springerreference_205392]
74. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
75. Ahmadi S, Klingelhöfer D, Erbe C, Holzgreve F, Groneberg DA, Ohlendorf D. Oral health: global research performance under changing regional health burdens. Int J Environ Res Public Health 2021 May 27;18(11):5743 [FREE Full text] [doi: 10.3390/ijerph18115743] [Medline: 34071884]

Abbreviations

AR: augmented reality
ICC: intraclass correlation coefficient
MARS: Mobile App Rating Scale
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by AH Sapci; submitted 19.07.24; peer-reviewed by D Patel, T Gladman; comments to author 05.10.24; revised version received 18.11.24; accepted 03.01.25; published 14.02.25.

Please cite as:
Rivera García GE, Cervantes López MJ, Ramírez Vázquez JC, Llanes Castillo A, Cruz Casados J
Reviewing Mobile Apps for Teaching Human Anatomy: Search and Quality Evaluation Study
JMIR Med Educ 2025;11:e64550
URL: https://mededu.jmir.org/2025/1/e64550
doi:10.2196/64550
PMID:

XSL•FO
RenderX

©Guadalupe Esmeralda Rivera García, Miriam Janet Cervantes López, Juan Carlos Ramírez Vázquez, Arturo Llanes Castillo, Jaime Cruz Casados. Originally published in JMIR Medical Education (https://mededu.jmir.org), 14.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Making Medical Education Courses Visible: Theory-Based Development of a National Database

Andi Gashi[1,2], MMed; Monika Brodmann Maeder[2,3], PD, Dr med, MME; Eva K. Hennel[2], PhD, Dr med, MME

[1]Medizinische Fakultät, University of Bern, Bern, Switzerland

[2]Forschung und Entwicklung, Schweizerisches Institut für ärztliche Weiter- und Fortbildung SIWF/ISFM, Elfenstrasse 18, Bern, Switzerland

[3]Universitätsklinik für Notfallmedizin, Inselspital Bern, Universitätsspital Bern, University of Bern, Bern, Switzerland

Corresponding Author:
Andi Gashi, MMed
Medizinische Fakultät, University of Bern, Bern, Switzerland

Abstract

Background: Medical education has undergone professionalization during the last decades, and internationally, educators are trained in specific medical education courses also known as "train the trainer" courses. As these courses have developed organically based on local needs, the lack of a general structure and terminology can confuse and hinder educators' information and development. The first aim of this study was to conduct a national search, analyze the findings, and provide a presentation of medical education courses based on international theoretical frameworks to support Swiss course providers and educators searching for courses. The second aim was to provide a blueprint for such a procedure to be used by the international audience.

Objective: In this study, we devised a scholarly approach to sorting and presenting medical education courses to make their content accessible to medical educators. This approach is presented in detailed steps and our openly available exemplary database to make it serve as a blueprint for other settings.

Methods: Following our constructivist paradigm, we examined content from medical education courses using a theory-informed inductive data approach. Switzerland served as an example, covering 4 languages and different approaches to medical education. Data were gathered through an online search and a nationwide survey with course providers. The acquired data and a concurrently developed keyword system to standardize course terminology are presented using Obsidian, a software that shows data networks.

Results: Our iterative search included several strategies (web search, survey, provider enquiry, and snowballing) and yielded 69 courses in 4 languages, with varying terminology, target audiences, and providers. The database of courses is interactive and openly accessible. An open-access template database structure is also available.

Conclusions: This study proposes a novel method for sorting and visualizing medical education courses and the competencies they cover to provide an easy-to-use database, helping medical educators' practical and scholarly development. Notably, our analysis identified a specific emphasis on undergraduate teaching settings, potentially indicating a gap in postgraduate educational offerings. This aspect could be pivotal for future curriculum development and resource allocation. Our method might guide other countries and health care professions, offering a straightforward means of cataloging and making information about medical education courses widely available and promotable.

(JMIR Med Educ 2025;11:e62838) doi:10.2196/62838

KEYWORDS

curriculum mapping; faculty development; competencies; database; medical education

Introduction

Ensuring high-quality health care necessitates the presence of well-trained medical educators [1]. Internationally, this has led to the development of frameworks that define their role and the creation of educational pathways for certification in many predominantly Anglophone countries. The terms "medical education" and "medical educator" are used with varying definitions in various settings. For this study, we define "medical educators" as the diverse group of health care professionals who teach, but regarding their learners, we focused on undergraduate medical students and physicians only. Hence, we include medical educators who are active in multiple settings: from clinical supervision and classroom instruction to the development and implementation of curricula. Their roles span the continuum of medical education from undergraduate through postgraduate training and continuing professional development.

Recent studies have identified significant tensions faced by medical educators, including lack of defined career structures, insufficient recognition of teaching roles, and challenges in developing educational identities [2,3]. While formal training programs alone cannot fully address these structural issues, they

XSL•FO
RenderX

serve as a crucial stepping stone in professionalizing medical education.

Globally, this coincides with the effort to provide these educators with a defined career path and support for their teaching activities. Notably, the Royal College of Physicians and Surgeons in Canada has introduced a "Clinician Educator Diploma," rooted in the findings of studies by Sherbino et al [4,5]. Similarly, the University of Michigan Medical School in the United States offers a Master of Health Professions Education [6-8], and in the United Kingdom, a variety of programs align with the "Professional Standards for medical, dental, and veterinary educators" established by the Academy of Medical Educators [9]. These initiatives exemplify the efforts to emphasize the value of quality medical education and facilitate access to quality training for educators.

Drawing on Wenger's Communities of Practice theory [10], these structured pathways and frameworks serve not only as career development tools but also as boundary objects that connect different medical education communities and facilitate knowledge sharing. When educators can easily identify and access training opportunities, they can more effectively participate in and contribute to their professional community. The visibility and accessibility of educational resources play a crucial role in fostering community development and knowledge exchange across different regions and language groups, ultimately supporting the growth of medical education as a professional field.

Lack of Career Pathways and an Opaque Landscape

Nevertheless, many regions internationally do not provide a career pathway for medical educators, nor do they use specific frameworks yet [3,11,12]. While some countries offer incentives such as continuous medical education (CME) credits for educational training, this is not standardized. Similarly in Switzerland, formal requirements for teaching qualifications are not yet standardized; there are some existing incentives for medical educators to pursue didactic training. Many didactic courses, particularly those offered or recognized by the Swiss Institute for Postgraduate Medical Education (SIWF), provide CME credits, independent from whether they concern undergraduate, postgraduate, or CME. Additionally, some specialty training programs, such as Internal Medicine, have incorporated mandatory didactic courses into their curriculum requirements [13]. However, these incentives remain fragmented and vary across specialties and institutions.

Still, in everyday reality, senior staff members are expected to educate while juggling everyday clinical tasks with providing education for students and junior doctors, often lacking dedicated time for teaching activities, as described in literature from Australia, Canada, and the United Kingdom [14-17] and in a 2008 Association for Medical Education in Europe guide [18]. Similarly, in Switzerland, where this study took place, there are few examples of clinics hiring specific teaching staff or at least dedicating specific hours in employment contracts to teaching. This practice is still not the norm and is rarely seen outside of specific pilot programs [19]. All of this leads to a diverse, somewhat undefined, and opaque medical education landscape and hinders high-quality medical education.

Aims and Research Questions

The aims of this study are twofold. The first aim of this study was to conduct a national search, analyze the findings, and provide a presentation of medical education courses based on international theoretical frameworks to support mainly Swiss course providers and educators searching for courses.

The second aim was to provide a blueprint for such a procedure to be used by the international audience.

Our research questions are (1) How do we conduct the search for courses and their content? and (2) How can we present the courses in an accessible way that is translatable to other regions?

Methods

Research Paradigm and Use of Theory

Guided by a constructivist research paradigm, which implies that no objective reality exists but that knowledge is created by social interactions [20], this study adopted a subjectivist inductive approach, using a theory-informed inductive data analysis method, described by Varpio et al [21].

This approach allowed a dynamic interplay between intermediate results and potential theoretical frameworks during data collection until we found the most fitting theoretical lens. This reflective and adaptive approach also ensured the relevance of the study results to the roles and competencies of medical educators as they were reflected in the Swiss data.

After a literature review of possible frameworks [4,5,7,9,22-29] and thorough deliberation by the research team, we selected the framework by Sidhu et al [28] as the best fit to support data analysis. Our decision to adopt this framework was strategic, not only for its clear enough lens through which to categorize and analyze course content, due to its comprehensive integration of 67 texts on educator competency domains, but also because it encompassed multiple health care professions, not just physicians. It comprehensively synthesizes educator competencies and identifies 6 distinct domains: "Teaching and Facilitating Learning," "Designing and Planning Learning," "Assessment of Learning," "Educational Research and Scholarship," "Educational Leadership and Management," and "Educational Environment, Quality, and Safety." This approach provides a robust, inclusive structure for understanding and evaluating educator competencies across different health professions.

This study lays the groundwork for a national strategy in Switzerland to enhance the quality of medical education. The first step was establishing a comprehensive database of existing educational offerings. This database facilitates an iterative and reciprocal examination of the course landscape, identifying gaps in the current educational landscape and recognizing requirements that may diverge from international frameworks. This approach ensures that the subsequent framework development and the certification of educators and courses are tailored to the unique needs and priorities of the Swiss medical education system. Additionally, it also allows us to identify and highlight areas lacking coverage by comparing with existing provisions.

XSL•FO
RenderX

Selection and Eligibility Criteria for Course Inclusion and Exclusion

Our study employed a systematic approach to searching for courses, with carefully defined inclusion and exclusion criteria.

The inclusion criteria encompassed the geographical scope, that is courses offered in Switzerland or online courses offered by Swiss institutions; the target group that is courses aimed at educators teaching physicians or medical students; and the content focus, that is courses with a primary focus on teaching skills.

The study's exclusion criterion was individual (1-to-1) offerings.

The limitation of the geographical scope ensured feasibility and was chosen with regard to the first study aim of building a national database. Our target group of educators who teach medical students and physicians, rather than including educators for learners in all health care professions, was driven by several methodological and practical considerations. First, our research team's expertise lies specifically in physician education, allowing us to conduct more nuanced and informed analysis within this domain. Second, our primary data collection method through the Joint Commission of Swiss Medical Schools (SMIFK/CIMS) naturally oriented our study toward medical education targeting physician settings.

With the formulation of the content focus in teaching skills, we deliberately excluded courses with generalized skills that could apply to any professional setting. These "soft skills" courses—such as generic presentation techniques, leadership pitching, or broadly applicable communication strategies—were set aside to concentrate on educational content specifically tailored to medical teaching contexts.

These criteria helped to refine the search and survey strategies, allowing for a targeted collection of data that is pertinent to the goals of this study.

Data Collection

First Phase: Online Search

We initiated our investigation with an extensive online search (May to July 2023), employing various search strategies to explore clinic, faculty, university, and specialists' association websites to find information on courses. We approached the search from an estimated end-user perspective, simulating how a medical professional seeking to expand their teaching skills might investigate educational opportunities. This meant using primarily search engines and direct institutional websites (including every medical faculty n=11, university hospital n=9, cantonal and larger regional hospital websites n=18, medical associations n=45, and the official SIWF website). Our search extended across Switzerland's linguistic diversity, using keywords in the country's 3 official languages (German, French, and Italian) and English to ensure no regional offerings were overlooked. Exemplary search strings, detailed in Multimedia Appendix 1, were designed to capture the multifaceted nature of medical education across different linguistic and regional contexts. Additionally, we used a pragmatic snowball-like method where we let discovered sources lead to additional course offerings. While not claiming absolute comprehensiveness, this method aimed to provide an overview of medical education courses in the Swiss context.

Second Phase: Widening the Search

Subsequently (July to December 2023), we sent out an e-mail targeting representatives within the Joint Commission of the Swiss Medical Schools (SMIFK/CIMS), a commission that includes deans and other stakeholders from all Swiss medical faculties. (see Multimedia Appendix 1 for survey questions). The survey did not systematically track response rates or potential selection biases. To widen our scope and better represent non-university medical institutions, we distributed a modified version of the survey to Human Resources departments of 25 larger hospitals in Switzerland in December 2023.

Through this dual approach of detailed online searches and surveys, we sought to capture a comprehensive snapshot of medical education offerings in Switzerland. It allowed us to gather insights not only from academic institutions but also from larger health care providers, thus offering a comprehensive view of the medical education landscape in Switzerland.

Course data were collected using a spreadsheet attached to the surveys, asking the respondents to fill in information about their courses and refer us to additional information if applicable. The specific metadata sought is presented in Table 1. Missing data were filled out by the authors, either by referring to online information or by consulting with the respective course contact persons directly. This step was important to enrich and verify the information found via the online search.

XSL•FO
RenderX

Table . Course metadata Items.

Item	Description
Title	Course title
Course ID	ID number applied to the course by the researchers
Provider	Institution providing the course
Contact information	Contact information for further inquiries (contact person and, if available, e-mail)
URL	Link to the course website (if available)
Location	Onsite or blended or hybrid or online
Duration	Duration of course as specified by the provider
Language	Course language (German, French, Italian, and English)
Costs	Costs of the course in Swiss Francs (CHF)
Certification	If the course yields a certification of attendance, title, or similar
Accreditation	Number of CME (continuous medical education)—credits or ECTS—credits for a study program
Target audience	For example, Attending Physicians in student care, all health care staff, bedside tutors, etc

Data Analysis

In the data collection phase, we observed an interplay between the course offerings and our theoretical framework, Sidhu et al [28]. It is important to note, however, that despite our openness to various health care professions at the start of the project, we ultimately focused our analysis exclusively on courses aimed at educators teaching physicians or medical students. This decision was made to maintain the scope of our study within manageable bounds, ensuring our research remained focused and relevant to our primary audience.

Our intention to mirror the actual landscape of Swiss medical education in its present state guided the database construction. Therefore, the selection of keywords for sorting courses and the strategy for their visualization were meticulously developed to emphasize prevalent topics and reveal the intricacies of course content, ensuring our methodology resonates with the real-world context of medical education. We also aimed to maintain the original terminology and language of course descriptions where possible to respect Switzerland's multilingual landscape and to accurately reflect the providers' offerings.

Building the Course Database

We chose Obsidian [30] as our platform for several practical reasons: its intuitive interface and its free access model for educational use made it accessible for users without extensive technical expertise, aligning with our goal for easy usability and adaptability. Obsidian allows for direct, interactive engagement with the data; this feature not only facilitated a more nuanced understanding of the data but also enabled us to refine and validate our keyword mapping and course organization strategies. Furthermore, Obsidian offered an integrated solution for both analyzing and publishing the data online, using Obsidian Publish [30], simplifying the transition from data collection to dissemination. This integration was particularly advantageous for providing an interactive, accessible resource for medical educators and ensuring the database's utility extended beyond our research team.

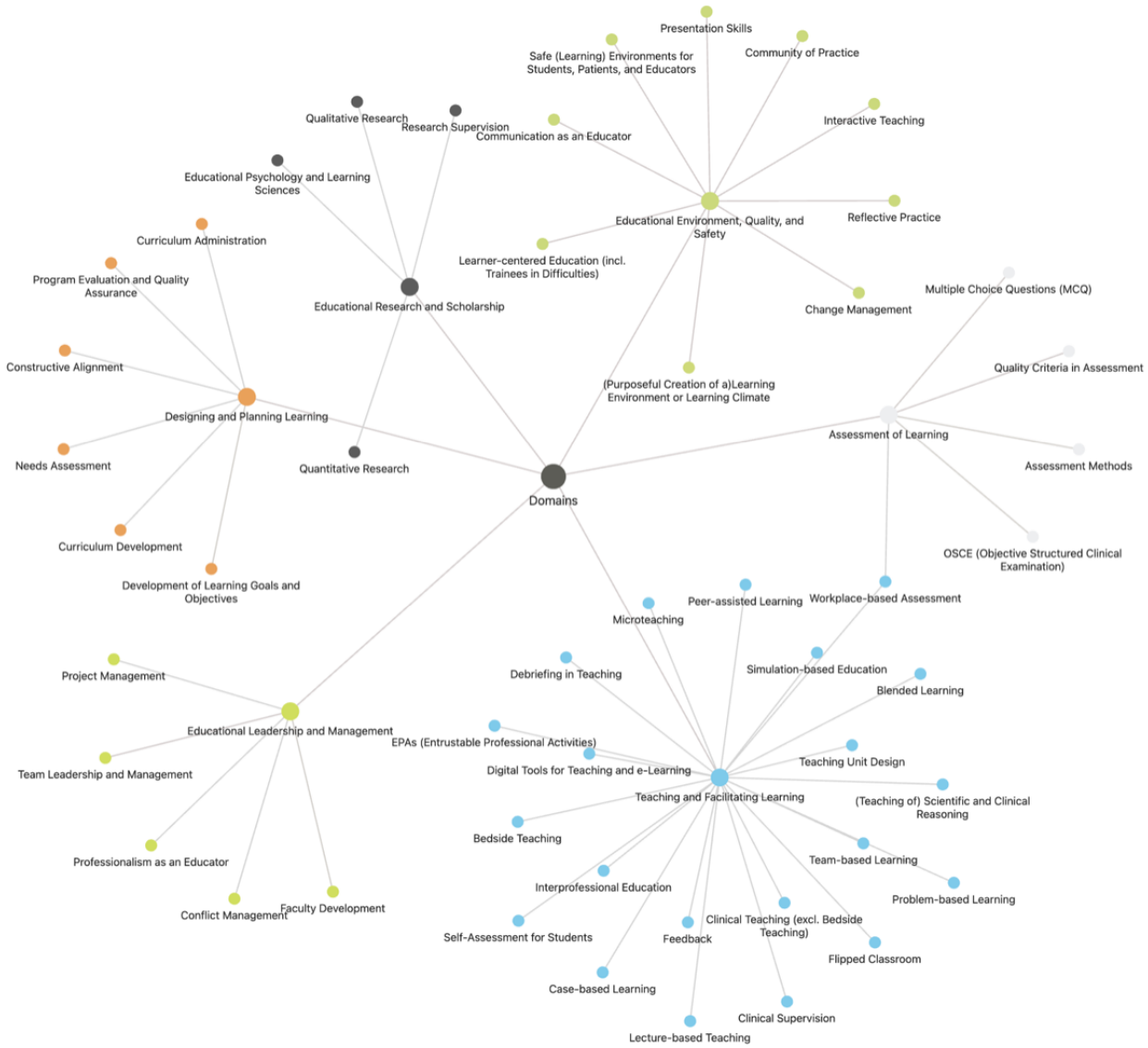
The spreadsheet data were then transferred by hand into an Obsidian database, with each course receiving its own Markdown file. Each file included a metadata section on the course and expanded sections on course descriptions, learning objectives and target audience, if available. All course data is provided in the original language and, if not already in English, was translated into English using DeepL, a free online translation tool [31].

Course Keywording, Standardization, and Refinement of Terminology

AG initially reviewed the compiled comprehensive course data to refine and validate our approach to organizing Swiss medical education course data, identifying preliminary keywords reflective of the content's breadth. Subsequently, AG and EH collaborated to align these keywords with the competency domains and definitions provided by Sidhu et al [28], ensuring a robust mapping grounded in established educational frameworks (see Multimedia Appendix 2 for a visual representation of the mapping process). The interplay between the domains provided by Sidhu et al. and the keywords as displayed in Obsidian.md is visualized in Figure 1.

XSL•FO
RenderX

Figure 1. Obsidian visualization of domains and keywords.



AG initially reviewed the compiled comprehensive course data to refine and validate our approach to organizing Swiss medical education course data, identifying preliminary keywords reflective of the content's breadth. Subsequently, AG and EH collaborated to align these keywords with the competency domains and definitions provided by Sidhu et al [28], ensuring a robust mapping grounded in established educational frameworks (see Multimedia Appendix 2 for a visual representation of the mapping process). The interplay between the domains provided by Sidhu et al and the keywords as displayed in Obsidian.md is visualised in Figure 1 .

This thorough process not only enhanced the accessibility and navigability of course data but also underscored our commitment to precision and educational integrity in documenting the landscape of medical education in Switzerland.

In our methodology, we iteratively developed a set of keywords to authentically represent various aspects of Switzerland's medical education landscape. Each keyword, crucial for clarity and precision, was documented and defined in separate Markdown files within Obsidian. Notably, integrating these definitions directly into Obsidian's Markdown documents enabled the keywords to serve as interactive reference points. By simply hovering over them, users can access a preview of the definition, improving the coherence and navigability of our documentation.

Database Publication

In the development of our study, we aimed to create a comprehensive and interactive database encapsulating the entirety of Switzerland's medical education landscape, specifically designed for ease of use and utility by medical educators. To make Switzerland's medical education course offerings easily navigable and useful for the academic community, we used Obsidian Publish [30] for the database's implementation. This platform was selected to ensure that the database was as interactive and accessible as possible.

To strengthen the trustworthiness of our search strategies we directly contacted each course provider to verify the representation of their course content and confirm the currency and correctness of the information.

XSL•FO
RenderX

Additionally, we prepared to release a template on GitHub to ensure our methodology was transparent and could be replicated in other contexts. This template provides the structure and template files necessary for creating a similar database, focusing on the organization and presentation of data. By offering these template files, we intended to facilitate the adoption of our approach by researchers or educators interested in developing educational landscape databases for different settings. The combination of using Obsidian Publish for the database and GitHub for sharing the template underscores our commitment to accessibility, transparency, and the potential for our work to be adapted and applied broadly.

Reflexivity

Acknowledging the importance of reflexivity for our constructivist approach, we critically examined the influence of our roles and backgrounds on this study. Each researcher brought distinct perspectives to conceptualizing a medical educator's role: AG started this project shortly after finishing his studies and provided insights on studying medicine, role definitions, and teaching cultures from a German-speaking and Italian-speaking Swiss region. MB, with a robust background as a practicing internal and emergency medicine doctor and a Master of Medical Education from the University of Bern obtained in 2006, brought insights both from her tenure as president of the SIWF and her experience in teaching emergency medicine courses. EH transitioned from clinical practice and medical education in Germany (Master of Medical Education in 2020) into medical education research in Switzerland (PhD in Medical Education), enriching our project with her comparative view and interest in theoretical concepts. To further enhance the validity of our keyword selection and study design, we engaged with 3 medical educators from Germany and Switzerland who have vast experience in curriculum development. Their feedback was instrumental in refining our course keywords, ensuring our framework resonated with the complexities and nuances of medical education across different linguistic and educational landscapes.

Study Setting

The present study was carried out in Switzerland, with a focus on identifying and analyzing medical education courses available within the country. The data collection phase spanned from May 2023 to December 2023. Subsequent data processing occurred concurrently and extended until March 2024.

Ethical Considerations

As no participants were involved, no ethical approval was needed. Everybody who supported this study by providing data, did so voluntarily without any incentives or conflicts of interest.

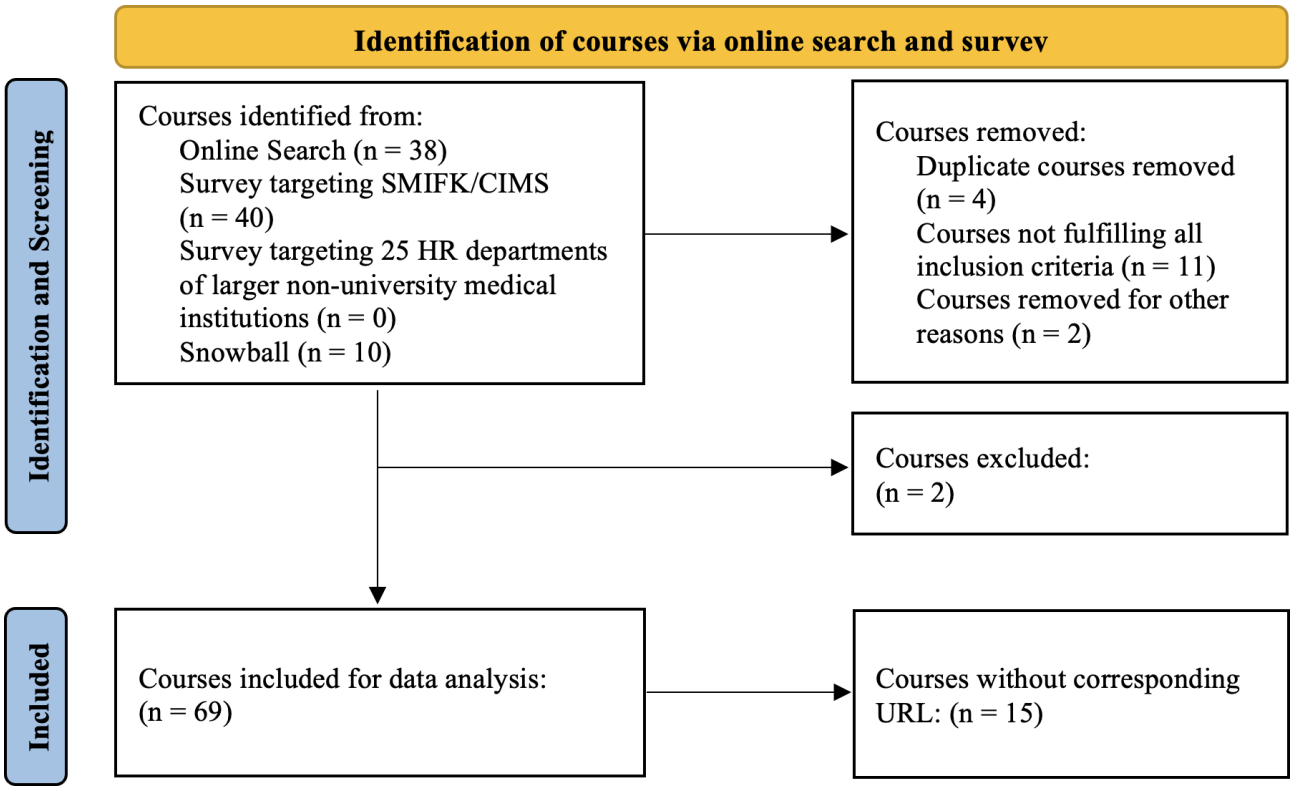
Results

Courses and Keywords

The initial online search, including the snowball method-like approach, resulted in 38 courses. The survey targeting the SMIFK/CIMS members resulted in 40 additional courses. The survey targeting Human Resources departments of 25 larger, non-university medical institutions in Switzerland yielded no additional new courses. In total, our search method yielded 69 eligible courses, as visible in Figure 2. Of the 69 total eligible courses identified, 36 were found through online searches alone, while 33 were discovered exclusively through survey responses.

XSL•FO
RenderX

Figure 2. Flowchart of the course identification process. One course that omitted whether physicians were part of its target audience expressed the wish upon further inquiry not to be included in this study for now but to be re-evaluated in future similar endeavors due to the current reorganization of the course offerings. Another course provider told us upon request for further information that our information on their course provided as a survey result was outdated and that they could not share further details about their course. Adapted from Page et al [32].



To ensure the reliability and currency of our findings, we checked back our data with the respective course providers. Of the 69 courses, 54 had active URLs, while 15 courses were documented without direct web links. These courses were primarily identified through institutional contacts and verified through direct communication in order to maintain the integrity of our comprehensive search strategy.

In our findings, undergraduate teaching courses outnumbered those aimed at postgraduate education (55 out of 69 courses aimed at undergraduate training), while most courses were offered by medical faculties (50 out of 69). The majority of courses were offered in French. In total, 11 courses were offered either additionally or exclusively in English. Most courses were offered onsite, meaning in an offline, face-to-face setting. The descriptive statistics derived from the course metadata, summarizing language use, provider types, course duration, and pricing, are summarized in Table 2.

XSL•FO
RenderX

Table . Aggregated course statistics.

Description	Count
Total number of courses	69
Average cost (CHF)	880
Courses offered free of charge	28
Courses with unknown costs	10
Course location	
Onsite	50
Hybrid	7
Blended	4
Online	6
Unspecified	2
Course provider	
Swiss Institute for Postgraduate Medical Education (SIWF)	8
Medical faculties	50
Specialists' associations	3
University hospitals	5
Other	3
Course language[a]	
German	36
French	40
Italian	7
English	11
Romansh	0
Target setting[b]	
Undergraduate teaching	55
Postgraduate teaching	28
Duration, h, average (range)	42 (1.5–8000)

[a]The count for languages may exceed the total number of courses because some courses are offered in more than one language.

[b]The count may exceed the total number of courses because some courses target both teaching settings.

In matching the course content across the 6 educational competency domains delineated by Sidhu et al, we identified 52 keywords. The distribution of these keywords was uneven, with 42% (22/52) of the keywords falling into the "Teaching and Facilitating Learning" domain. In contrast, the domains of "Educational Leadership and Management" and "Assessment of Learning" were represented to a lesser extent, with only 5 keywords each, while "Educational Environment, Quality, and Safety" had 10 keywords. The "Educational Research and Scholarship" domain showed the least coverage with 4 keywords. Individual courses varied in their keyword coverage, with an average of 4 keywords per course (median: 3); one comprehensive course covered 33 keywords across domains. The distribution of keywords across domains is presented in Table 3, with a graphical overview represented in Figure 1.

XSL•FO
RenderX

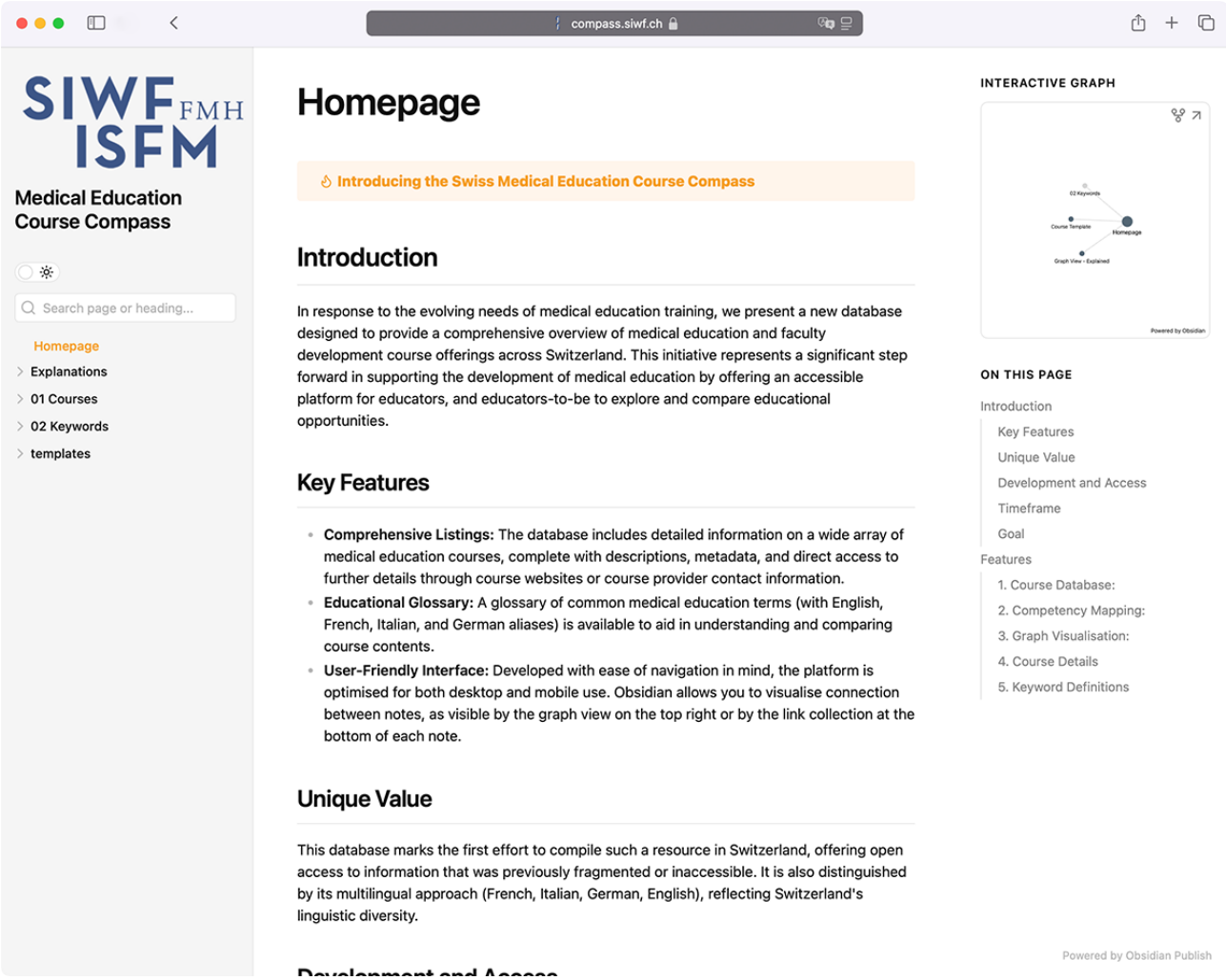
Table . Domains and keyword distribution among courses.

Keywords	Keywords per domain
Assessment of Learning	5
Designing and Planning Learning	6
Educational Environment, Quality, and Safety	10
Educational Leadership and Management	5
Educational Research and Scholarship	4
Teaching and Facilitating Learning	22
Keywords covered per Course (average, median, max)	4, 3, 33
Sum of keywords	52

Database and Template Publication

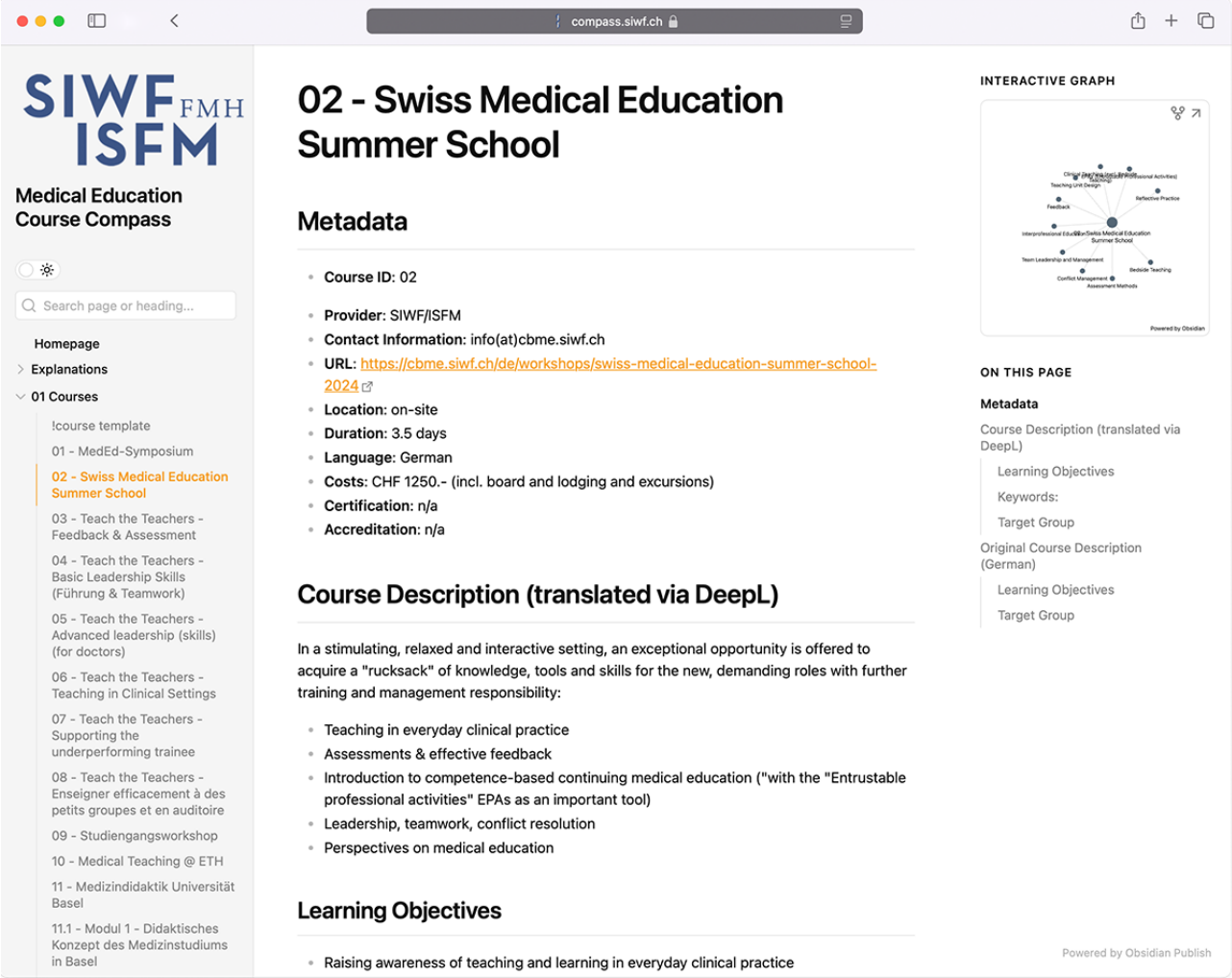
Using Obsidian publish, the database containing the Swiss Course Data were published online [33] (Screenshots of the landing page and an exemplary course are provided with Figures 3 and 4, please refer to Multimedia Appendix 3 for a video example of navigating the database). We also published an exemplary, ready-to-use database template with a Markdown folder and file structure as a GitHub-Repository, together with instructions explaining the structure and use of the repository to facilitate adoption by other users [34].

Figure 3. Screenshot of the publicly accessible database landing page.



XSL•FO
RenderX

Figure 4. Exemplary screenshot of one course.



Discussion

Principal Findings

In line with the 2 aims of this study, our results cover 2 aspects. First, we searched for, analyzed, and presented the Swiss courses. We found that 69 courses are provided, of which 55 target teachers of undergraduate learners and 28 target teachers of postgraduate learners. The courses were offered in several languages (52% German, 58% French, 10% Italian, 16% English) and several formats (72% onsite, 10% in a hybrid manner, 6% as a blended-learning format, and 9% online). Course content covered mainly the “Teaching and Facilitating Learning” domain, which contained 42% (22/52) of all identified keywords and was also the most prominent in course offerings, with the keyword ’feedback’ being featured most frequently, appearing in 40.6% (28/69) of all courses.

Second, we created an internationally transferable strategy based on an international framework, using free-for-educational-use software, with published search strings to be adapted by others, and providing all metadata open access.

This study addressed the need for an approachable method of sorting and displaying courses by providing a comprehensive, accessible database of medical education courses in Switzerland. This initiative is particularly relevant in contexts where medical education training lacks clear structure and uniformity, especially across different linguistic and cultural regions.

While we acknowledge that our search strategy may not have uncovered every course, we believe this data set sufficient for our study’s objectives and that it reflects the real-world challenges faced by aspiring medical educators in navigating the educational landscape.

Challenges in Navigating the Medical Education System

Several pervasive challenges exist in medical education. These include a discernible lack of dedicated teaching time [35], insufficient, non-transparent, or misdirected funding [36], and an absence of recognition for educators within the medical field [37]. Such issues create a universal backdrop against which specific national education systems can be analyzed.

Our research provides a structured analysis of medical education courses in the Swiss system, illuminating both publicly accessible offerings and those that emerge through professional networks. Significantly, nearly half of the identified courses (33 out of 69) were discovered only through contact with course providers rather than public online searches, revealing a critical accessibility gap. This finding suggests that course discovery often depends heavily on existing professional networks and insider knowledge, potentially creating barriers for newcomers

XSL•FO
RenderX

to the field who lack such connections. This study is the first of its kind to investigate medical didactics courses in a structured manner in Switzerland, providing an overview and allowing new perspectives.

Our study uncovers a complex challenge within the Swiss medical education system: navigating the educational landscape shaped by 4 national languages—German, French, Italian, and Romansh. Among these, German, French, and Italian are official languages. In this multilingual setting, English emerges as a crucial lingua franca in international research and academia and as a common medium within Switzerland, facilitating communication across linguistic barriers. This is reflected in the educational offerings, with 11 courses being conducted partially or entirely in English. Such linguistic plurality further aggravates the existing variability in didactic methods and medical education and research terminology [38,39], presenting additional challenges to the creation of an educational framework. This situation is further complicated by issues of information accessibility—with 15 courses not having available URLs, it renders it difficult for potential participants to find accurate, up-to-date information about these educational opportunities online.

Implications for Medical Educator Career Pathways

Building on the initial findings, our work, as the first comprehensive compilation of available didactic courses within the Swiss medical education system, can potentially enable more transparency in the field. By illuminating the current educational offerings, our research could facilitate faculty development and inform career choices for medical educators, as they now have a more straightforward overview of the resources and opportunities available to them. This transparency is a critical step towards enhancing career pathways, as it allows for a strategic approach to personal and professional development within the context of medical education.

However, it is essential to note that our study revealed a predominance of courses aimed at teaching in undergraduate (medical school) settings. While this finding initially suggests a potential undersupply of educational opportunities targeting postgraduate teaching, it notably reflects Switzerland's institutional structure, where universities are primarily responsible for undergraduate medical education but not residency training or CME. This structural arrangement, where different institutions oversee different stages of medical education, may contribute to the observed imbalance in course offerings. The relative scarcity of training for postgraduate settings might indicate a need for a broader array of courses catering to the ongoing development of medical professionals beyond their initial degrees. Addressing this gap could lead to a more robust and comprehensive educational framework, ensuring that medical educators are well-equipped to foster the next generation of medical practitioners at all stages of their professional journey.

Comparison to International Models

Our comparative analysis with international models highlights notable differences in structuring medical education career pathways. Particularly in Anglophone countries, such as Australia, Canada, the United Kingdom, and the United States, well-established career pathways and educator frameworks are actively promoted and used, unlike in Switzerland, where such structured frameworks are absent [3,11,12].

When aligning our findings with the integrative framework by Sidhu et al [28], which outlines 6 educator competency domains, we observed that Swiss medical education predominantly emphasizes the domain of "Teaching and Facilitating Learning." This domain is focused on enhancing learning through suitable methods and resources, including assessment for learning. This heavy focus may inadvertently lead to a neglect of the other 5 domains ("Educational Leadership and Management," "Educational Environment, Quality, and Safety", "Designing and Planning Learning," "Assessment of Learning," and "Educational Research and Scholarship"), suggesting a possible imbalance in the educational emphasis. Our keyword map corroborates this emphasis, revealing a detailed and nuanced depiction of the "Teaching and Facilitating Learning" domain. Our mapping strategy aimed to reflect the specificity and depth of course offerings without excessive summarization, thus indicating their relative prominence by the frequency with which they appear in our data set.

The predominance of the "Teaching and Facilitating Learning" domain could stem from its direct applicability in educational settings and the relative ease of teaching and assessing these skills. This focus on tangible teaching methods and resources, which offer clear benefits and outcomes, could make it a natural focal point for educators aiming to directly impact student achievement. However, this emphasis may inadvertently lead to the undervaluation of other critical educator competencies, such as "Educational Leadership and Management" or "Educational Research and Scholarship." These domains encompass more abstract competencies that resemble attitudes or overarching professional dispositions rather than concrete skills, presenting challenges for direct instruction due to their less immediately visible impacts and harder-to-quantify qualities.

The challenge of effectively imparting the more abstract domains within medical education and the observable predilection for addressing more accessible teaching topics can be analogized to the tendency to assess readily quantifiable competencies [40]. This parallel might reflect an educational predilection for what can be straightforwardly taught and measured, perhaps at the expense of more profound, more complex competencies that are less amenable to conventional assessment methodologies. Such a tendency may not fully encapsulate the multifaceted nature of medical educator competencies, underscoring a potential disjunction between educational priorities and the comprehensive skill set required for clinical excellence.

The marked predominance of teaching methods and learning facilitation within the Swiss context, as opposed to a more balanced distribution across the 6 domains, is an intriguing phenomenon. We propose that this area warrants further inquiry. Investigating why there is such a strong focus on these teaching competencies within Switzerland, especially compared to the broader scope seen internationally, could yield insights that

XSL•FO
RenderX

inform future developments in medical education. This analysis might ultimately contribute to enhancing educator frameworks and diversifying professional development opportunities.

Future Directions and Potential Impact

The insights garnered from this study lay the groundwork for developing a unified and expansive framework for medical education in Switzerland, emphasizing the need for a formal recognition and certification process for medical education courses and medical educators. By enhancing awareness of available courses and clarifying the expectations for a medical educator, this initiative could significantly improve the quality of medical education. Increased awareness is crucial for establishing clear competencies in the first place, and it also serves to acknowledge and validate existing efforts while encouraging the development of future educational opportunities. The overview of courses also offers the basis for future discussions, which have already begun. Currently, the database does not give information on the quality of courses, as this could not reliably be assessed. For a future version of the database, methods to assess course quality and display this transparently are developed.

In addition, the overview on courses in comparison with the framework has shown some blind spots of the current training. Courses on “Assessment of Learning,” “Educational Leadership and Management,” and “Educational Research and Scholarship” seem to be rare. As a follow-up to this study, the responsible groups (SMIFK and SIWF) will discuss these gaps and take this as a base for future course offerings.

Furthermore, this focus on a structured framework could resonate on an international level, offering a model for contexts where clear career pathways for medical educators are similarly undefined, thus having the potential to inform and shape international standards in medical education.

While this study focused exclusively on courses on medical education (physician-adjacent) settings to maintain a manageable scope, we deliberately chose Sidhu et al’s framework for its interprofessional applicability across health care education. Combined with our intentionally transparent and replicable methodology, this provides a strong foundation for other health care professions to adapt our approach to mapping their own educational landscapes. The template we developed could serve as a starting point for similar analyses in other health professions’ education systems.

Limitations

Our study encountered different challenges. The research relied primarily on internet-searchable courses, potentially overlooking non-digital or unadvertised educational offerings. Additionally, our snowball search approaches introduce potential selection biases.

As we focused on assembling an overview instead of the details of courses, the depth of content validation regarding, for example, course quality, pedagogical methodologies, and educational effectiveness is limited.

Future research should focus on course content and course quality and investigate the usability and comprehensiveness of our database, including an accessibility audit of course information, a comprehensive comparison with expert-identified course offerings, and systematic verification of course details. We are planning to investigate the implications of the use of the database, such as changes in content offered and new strategies to acknowledge medical educator careers.

Conclusions

The methodology showcased in our study serves as a valuable template for international adaptation, offering insights into how diverse educational systems can be evaluated, organized, and presented effectively. By innovatively mapping the educational landscape with the example of Switzerland, we have provided a replicable approach that could aid other countries, even with similar linguistic and cultural challenges. The creation of our database amplifies the visibility and transparency of medical education courses available, facilitating better decision-making for educators. This enhanced access to information allows medical professionals to chart their career progressions more effectively, emphasizing the benefits of more transparent navigation, especially in environments without well-defined educational pathways.

Through this pioneering work, we aim to contribute to the global conversation on medical education, fostering a more interconnected and accessible educational landscape that benefits educators and students alike.

Acknowledgments

We thank Dr phil-nat. Elke Bayha, Dr rer nat Felix Joachimski, MME, and Dr med vet Melanie Simon, MME, for their critical review and helpful comments on developing the keyword map. We also thank all course providers and survey participants for their collaboration on this project and for so generously answering our questions on course details.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Document outlining our Search Strategy for the online search.
[DOCX File, 23 KB - mededu_v11i1e62838_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Screenshot of Miro Board used for mapping between domains and keywords which helped facilitate discussions between authors.
[PNG File, 1372 KB - mededu_v11i1e62838_app2.png]

Multimedia Appendix 3
Video example of navigating the publicly accessible online database
[MP4 File, 204005 KB - mededu_v11i1e62838_app3.mp4]

References

1. Norman G. Medical education: past, present and future. Perspect Med Educ 2012 Mar;1(1):6-14. [doi: 10.1007/s40037-012-0002-7] [Medline: 23316454]
2. Sethi A, Ajjawi R, McAleer S, Schofield S. Exploring the tensions of being and becoming a medical educator. BMC Med Educ 2017 Mar 23;17(1):62. [doi: 10.1186/s12909-017-0894-3] [Medline: 28335820]
3. Bartle E, Thistlethwaite J. Becoming a medical educator: motivation, socialisation and navigation. BMC Med Educ 2014 May 31;14(1):110. [doi: 10.1186/1472-6920-14-110] [Medline: 24885740]
4. Sherbino J, Frank JR, Snell L. Defining the key roles and competencies of the clinician–educator of the 21st century. Acad Med 2014 May;89(5):783-789. [doi: 10.1097/ACM.0000000000000217] [Medline: 24667507]
5. Sherbino J, Snell L, Dath D, Dojeiji S, Abbott C, Frank JR. A national clinician-educator program: A model of an effective community of practice. Med Educ Online 2010 Dec 6;15(1):5356. [doi: 10.3402/meo.v15i0.5356] [Medline: 21151594]
6. Fitzgerald JT, Burkhardt JC, Kasten SJ, et al. Assessment challenges in competency-based education: a case study in health professions education. Med Teach 2016 May;38(5):482-490. [doi: 10.3109/0142159X.2015.1047754] [Medline: 26052881]
7. Gruppen LD, Burkhardt JC, Fitzgerald JT, et al. Competency-based education: programme design and challenges to implementation. Med Educ 2016 May;50(5):532-539. [doi: 10.1111/medu.12977] [Medline: 27072442]
8. Master of HPE of university of michigan medical school. Learning Health Sciences. 2017 Mar 17. URL: https://medicine.umich.edu/dept/lhs/education/degree-programs/master-health-professions-education-mhpe/curriculum [accessed 2023-11-03]
9. Academy of Medical Educators. Professional Standards for Medical, Dental and Veterinary Educators, 4th edition: Academy of Medical Educators; 2021. URL: https://www.medicaleducators.org/write/MediaManager/Documents/AoME_Professional_Standards_4th_edition_1.0_(web_full_single_page_spreads).pdf [accessed 2025-04-09]
10. Wenger E. Communities of practice and social learning systems: the career of a concept. In: Blackmore C, editor. Social Learning Systems and Communities of Practice: Springer; 2010:179-198. [doi: 10.1007/978-1-84996-133-2_11]
11. Walport M. Medically-and dentally-qualified academic staff: recommendations for training the researchers and educators of the future. : Academic Careers Sub-committee of Modernising Medical Careers and the UK Clinical Research Collaboration; 2005 URL: https://www.ukcrc.org/wp-content/uploads/2014/03/Medically_and_Dentally-qualified_Academic_Staff_Report.pdf [accessed 2025-04-09]
12. Church H, Brown MEL. Rise of the Med-Ed-ists: achieving a critical mass of non-practicing clinicians within medical education. Med Educ 2022 Dec;56(12):1160-1162. [doi: 10.1111/medu.14940] [Medline: 36148497]
13. SGAIM - Schweizerische Gesellschaft für Allgemeine Innere Medizin. Fachärztin oder facharzt für allgemeine innere medizin - weiterbildungsprogramm vom. 2022 Jan 1. URL: https://www.sgaim.ch/fileadmin/user_upload/Weiterbildung/aim_wbp_d_1_.pdf [accessed 2025-01-16]
14. Jelinek GA, Weiland TJ, Mackinlay C. Supervision and feedback for junior medical staff in Australian emergency departments: findings from the emergency medicine capacity assessment study. BMC Med Educ 2010 Nov 2;10(1):74. [doi: 10.1186/1472-6920-10-74] [Medline: 21044342]
15. Steinert Y, McLeod PJ, Boillat M, Meterissian S, Elizov M, Macdonald ME. Faculty development: a “field of dreams”? Med Educ 2009 Jan;43(1):42-49. [doi: 10.1111/j.1365-2923.2008.03246.x] [Medline: 19140996]
16. Fish RM, Gawne SJ, Machin L. Balancing medical education with service in the workplace: a qualitative case study. JWL 2022 Jan 27;34(2):176-187. [doi: 10.1108/JWL-05-2021-0064]
17. Piquette D, Moulton CA, LeBlanc VR. Balancing care and teaching during clinical activities: 2 contexts, 2 strategies. J Crit Care 2015 Aug;30(4):678-684. [doi: 10.1016/j.jcrc.2015.03.002] [Medline: 25776896]
18. Ramani S, Leinster S. AMEE Guide no. 34: teaching in the clinical environment. Med Teach 2008;30(4):347-364. [doi: 10.1080/01421590802061613] [Medline: 18569655]
19. Hohl F. 42-Stunden-Woche mit grossem Plus. Schweiz Ärzteztg . [doi: 10.4414/saez.2023.1267925043]
20. Ng SL, Baker L, Cristancho S, Kennedy TJ, Lingard L. Qualitative research in medical education: methodologies and methods. In: Swanwick T, Forrest K, O’Brien BC, editors. Understanding Medical Education, 1st edition: Wiley, Vol. 2018:427-441. [doi: 10.1002/9781119373780]
21. Varpio L, Paradis E, Uijtdehaage S, Young M. The distinctions between theory, theoretical framework, and conceptual framework. Acad Med 2020 Jul;95(7):989-994. [doi: 10.1097/ACM.0000000000003075] [Medline: 31725464]
22. Crosby RJ. AMEE Guide No 20: the good teacher is more than a lecturer - the twelve roles of the teacher. Med Teach 2000 Jan;22(4):334-347. [doi: 10.1080/014215900409429]

XSL•FO
RenderX

23. Hesketh EA, Bagnall G, Buckley EG, et al. A framework for developing excellence as A clinical educator. Med Educ 2001 Jun;35(6):555-564. [doi: 10.1046/j.1365-2923.2001.00920.x] [Medline: 11380858]
24. Srinivasan M, Li STT, Meyers FJ, et al. "Teaching as a Competency": competencies for medical educators. Acad Med 2011 Oct;86(10):1211-1220. [doi: 10.1097/ACM.0b013e31822c5b9a] [Medline: 21869655]
25. Parson L, Childs B, Elzie P. Using competency-based curriculum design to create a Health Professions Education Certificate Program that Meets the needs of students, administrators, faculty, and patients. Health Professions Education 2018 Sep;4(3):207-217. [doi: 10.1016/j.hpe.2018.03.008]
26. Walsh A, Koppula S, Antao V, et al. Preparing teachers for competency-based medical education: Fundamental teaching activities. Med Teach 2018 Jan;40(1):80-85. [doi: 10.1080/0142159X.2017.1394998] [Medline: 29113520]
27. Fallis D, Irwin S, Cervero R, Durning S. Frameworks to guide faculty development for health professions education: a scoping review. J Contin Educ Health Prof 2022 Jul 1;42(3):180-189. [doi: 10.1097/CEH.0000000000000376] [Medline: 34459440]
28. Sidhu NS, Allen KJ, Civil N, et al. Competency domains of educators in medical, nursing, and health sciences education: an integrative review. Med Teach 2023 Feb;45(2):219-228. [doi: 10.1080/0142159X.2022.2126758] [Medline: 36179761]
29. Barrier J. Conférence Internationale des Doyens des Facultés de Médecine d' Expression Française. Recommandations pour la formation Pédagogique des enseignants de Médecine. 1998. URL: https://docplayer.fr/19159965-Recommandations-pour-la-formation-pedagogique-des-enseignants-de-medecine.html [accessed 2023-11-03]
30. Li S, Xu E. Dynalist Inc. Obsidian.md. URL: https://obsidian.md/ [accessed 2023-11-02]
31. DeepL SE. DeepL Translator. URL: https://www.deepl.com/translator [accessed 2024-02-04]
32. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
33. Homepage - Medical Education Course Compass. URL: https://compass.siwf.ch/Homepage [accessed 2024-05-20]
34. Gashi A. Andi-gashi/mededcoursecompass. 2024 Apr 1. URL: https://github.com/andi-gashi/mededcoursecompass [accessed 2024-05-05]
35. Aldeen AZ, Gisondi MA. Bedside teaching in the emergency department. Acad Emerg Med 2006 Aug;13(8):860-866. [doi: 10.1197/j.aem.2006.03.557] [Medline: 16766739]
36. Dacre J, Walsh K. Funding of medical education: the need for transparency. Clin Med (Northfield) 2013 Dec;13(6):573-575. [doi: 10.7861/clinmedicine.13-6-573]
37. Fantaye AW, Kitto S, Hendry P, et al. Attributes of excellent clinician teachers and barriers to recognizing and rewarding clinician teachers' performances and achievements: a narrative review. Can Med Educ J 2022 May;13(2):57-72. [doi: 10.36834/cmej.73241] [Medline: 35572019]
38. Wojtczak A. Medical education terminology. Med Teach 2002 Jul;24(4):357-357. [doi: 10.1080/01421590220145699] [Medline: 12193314]
39. Finn GM, Charmer B, Burton OE, et al. How to define clinical education research terminology: a glossary. Clin Teach 2023 Aug;20(4):e13605. [doi: 10.1111/tct.13605] [Medline: 37503773]
40. Foster N. 21st Century competencies: challenges in education and assessment Innov Assess Meas Support Complex Ski. URL: https://www.oecd.org/en/publications/innovating-assessments-to-measure-and-support-complex-skills_e5f3e341-en/full-report.html [accessed 2025-04-09]

Abbreviations

CME: continuous medical education
SIWF: Swiss Institute for Postgraduate Medical Education
SMIFK/CIMS: Joint Commission of the Swiss Medical Schools

Edited by B Lesselroth; submitted 12.06.24; peer-reviewed by AMH Chen, D Jackson; revised version received 19.01.25; accepted 25.02.25; published 16.04.25.

Please cite as:
Gashi A, Brodmann Maeder M, Hennel EK
Making Medical Education Courses Visible: Theory-Based Development of a National Database
JMIR Med Educ 2025;11:e62838
URL: https://mededu.jmir.org/2025/1/e62838
doi:10.2196/62838

© Andi Gashi, Monika Brodmann Maeder, Eva K Hennel. Originally published in JMIR Medical Education (https://mededu.jmir.org), 16.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution

XSL•FO
RenderX

License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Large Language Models in Biochemistry Education: Comparative Evaluation of Performance

Olena Bolgova[1], MD, PhD; Inna Shypilova[2], MD, PhD; Volodymyr Mavrych[1], MD, PhD

[1]College of Medicine, Alfaisal University, Al Takhassousi St, Riyadh, Saudi Arabia

[2]School of Medicine, St Mathews University, George Town, Cayman Islands

Corresponding Author:
Volodymyr Mavrych, MD, PhD
College of Medicine, Alfaisal University, Al Takhassousi St, Riyadh, Saudi Arabia

Abstract

Background: Recent advancements in artificial intelligence (AI), particularly in large language models (LLMs), have started a new era of innovation across various fields, with medicine at the forefront of this technological revolution. Many studies indicated that at the current level of development, LLMs can pass different board exams. However, the ability to answer specific subject-related questions requires validation.

Objective: The objective of this study was to conduct a comprehensive analysis comparing the performance of advanced LLM chatbots—Claude (Anthropic), GPT-4 (OpenAI), Gemini (Google), and Copilot (Microsoft)—against the academic results of medical students in the medical biochemistry course.

Methods: We used 200 USMLE (United States Medical Licensing Examination)–style multiple-choice questions (MCQs) selected from the course exam database. They encompassed various complexity levels and were distributed across 23 distinctive topics. The questions with tables and images were not included in the study. The results of 5 successive attempts by Claude 3.5 Sonnet, GPT-4 - 1106, Gemini 1.5 Flash, and Copilot to answer this questionnaire set were evaluated based on accuracy in August 2024. Statistica 13.5.0.17 (TIBCO Software Inc) was used to analyze the data's basic statistics. Considering the binary nature of the data, the chi-square test was used to compare results among the different chatbots, with a statistical significance level of $P<.05$.

Results: On average, the selected chatbots correctly answered 81.1% (SD 12.8%) of the questions, surpassing the students' performance by 8.3% ($P=.02$). In this study, Claude showed the best performance in biochemistry MCQs, correctly answering 92.5% (185/200) of questions, followed by GPT-4 (170/200, 85%), Gemini (157/200, 78.5%), and Copilot (128/200, 64%). The chatbots demonstrated the best results in the following 4 topics: eicosanoids (mean 100%, SD 0%), bioenergetics and electron transport chain (mean 96.4%, SD 7.2%), hexose monophosphate pathway (mean 91.7%, SD 16.7%), and ketone bodies (mean 93.8%, SD 12.5%). The Pearson chi-square test indicated a statistically significant association between the answers of all 4 chatbots ($P<.001$ to $P<.04$).

Conclusions: Our study suggests that different AI models may have unique strengths in specific medical fields, which could be leveraged for targeted support in biochemistry courses. This performance highlights the potential of AI in medical education and assessment.

(JMIR Med Educ 2025;11:e67244) doi:10.2196/67244

KEYWORDS

ChatGPT; Claude; Gemini; Copilot; biochemistry; LLM; medical education; artificial intelligence; NLP; natural language processing; machine learning; large language model; AI; ML; comprehensive analysis; medical students; GPT-4; questionnaire; medical course; bioenergetics

Introduction

Recent breakthroughs in artificial intelligence (AI), especially in large language models (LLMs), have started a new era of innovation across diverse fields, with medicine leading the charge in this technological revolution. The integration of AI into various medical disciplines such as oncology, radiology, and pathology has demonstrated its advancing clinical uses and its potential to revolutionize health care delivery [1-3]. As new LLMs continue to emerge and evolve, AI is poised to fundamentally reshape our understanding and approach to medicine, offering unprecedented opportunities for improved patient care, diagnostics, and medical education [4].

While academic interest in AI has surged in recent years, integrating AI technologies in educational settings, particularly medicine, has been uneven and fraught with challenges. Among

XSL•FO
RenderX

many AI tools available, ChatGPT has emerged as a potential game-changer in medical education [5,6]. This sophisticated language model, powered by advanced neural networks, demonstrates a remarkable ability to interpret prompts and generate human-like responses, making it difficult to distinguish from human-produced language.

LLM's underlying transformer architecture enables it to excel in natural language understanding, continuously processing and adapting to new information. This adaptability, combined with its vast knowledge base, presents promising opportunities for enhancing teaching and learning methodologies in medical education [7]. AI-powered tools such as ChatGPT may be particularly effective in addressing persistent challenges in student engagement, offering interactive and personalized learning experiences that traditional teaching methods often struggle to provide [8].

OpenAI's GPT-4 and GPT-3.5, Google's Gemini, and Anthropic's Claude have emerged as frontrunners, offering unique capabilities and potential medical education and practice applications. As of 2024, the AI landscape in health care has become increasingly diverse, with over 20 LLMs available for public use. Among them, 4 are the most promising.

Anthropic developed Claude, an AI assistant known for its strong natural language understanding and generation capabilities. It has been trained on a wide range of data and is designed to be helpful, harmless, and honest. Claude has shown particular strength in tasks requiring nuanced understanding and ethical reasoning [9].

Created by OpenAI, GPT-4 is the latest GPT series iteration. It represents a significant advancement over its predecessor, GPT-3, with improved language understanding, generation, and reasoning capabilities. GPT-4 has demonstrated impressive performance across various domains, including coding, creative writing, and analytical tasks [10].

Developed by Google AI, Gemini is a multimodal AI model capable of understanding and generating text, images, and other forms of data. It comes in different sizes and is optimized for various tasks and computational requirements. Gemini has shown strong performance in complex reasoning tasks and can understand context across different modalities [11].

Created by GitHub in collaboration with OpenAI, Copilot is an AI pair programmer designed to assist developers by suggesting code completions and entire functions. It is now an integral part of Microsoft Windows. While primarily focused on coding tasks, Copilot's underlying language model has shown capabilities in understanding and generating natural language [12].

One primary method for assessing the capabilities of LLMs in knowledge-based fields, including medicine, is their performance on multiple-choice tests [13-16]. The release of GPT-4 by OpenAI in 2023 marked a significant milestone, demonstrating impressive test-taking abilities across various domains [17]. Similarly, Claude 2 from Anthropic, released in June 2023, has gained attention for its ability to process larger input spaces (up to 100,000 tokens), potentially allowing for a more comprehensive analysis of medical texts and case studies [8].

The high accuracy demonstrated by ChatGPT-4 in answering multiple-choice questions (MCQs) compared to medical students' performance is particularly noteworthy. It suggests that AI could be an effective study aid, helping students review and reinforce their knowledge across various medical subjects. However, it is essential to view AI as a complementary tool rather than a replacement for MCQs that have transformed from their conventional use as assessment tools to become a versatile educational approach in medical curricula. MCQs stimulate students' cognitive abilities and promote active interaction with study materials. By using advanced generative AI-driven language models to address MCQs in medical physiology and other subjects, educators may provide students with an innovative and engaging learning experience, potentially enhancing their grasp of essential medical concepts, traditional teaching methods, or human expertise [18,19].

Recent studies have begun to compare the performance of different AI models in medical education contexts. For instance, Claude, an LLM developed by Anthropic, has shown promising results in solving medical MCQs. Some studies have indicated that Claude demonstrated a high frequency of right answers and explanations compared to ChatGPT-3.5 [8,20]. These comparative studies are crucial in understanding the strengths and limitations of different AI models in medical education. They help educators and researchers identify the most suitable tools for specific learning objectives and contexts within medical curricula.

Despite the promising results, it is important to note the variability in AI performance across different studies and question types. For example, while some studies reported high accuracy rates for ChatGPT in physiology tests [5,8], others found lower performance rates, particularly as the complexity and difficulty of questions increased [21,22]. This variability underscores the need for careful consideration when integrating AI tools into medical education. Educators must be aware of these tools' strengths and limitations and ensure they are used appropriately to complement, rather than replace, traditional teaching methods.

It is important for educational strategies to prioritize the integration of LLMs into the curriculum as a vital aspect of the learning process. This integration should enable students to cultivate critical thinking and analytical skills, particularly in understanding the constraints of AI. LLMs have the potential to offer students in-depth knowledge and diverse viewpoints, facilitating a more thorough comprehension of intricate medical concepts [23]. By using the output of LLMs and working alongside educators to draw upon their existing knowledge, students can actively participate in the learning process. This collaborative approach allows for the refinement of their understanding and insights. The future of medical education depends on the seamless integration of human expertise with AI-powered tools [3,19,23].

The aim of this study was to conduct a comprehensive analysis comparing the performance of advanced LLM chatbots—GPT-4, Claude, Copilot, and Gemini—against the academic results of

XSL•FO
RenderX

medical students in biochemistry. The research objectives were to evaluate the following hypotheses:

- The AI chatbots will perform similarly to medical students on factual recall and basic concept application questions in biochemistry but may show differences in performance on complex problem-solving scenarios.
- There will be significant variation in performance among the different AI models, with newer models (GPT-4 and Claude) potentially showing higher accuracy compared to earlier versions.
- The AI-driven LLMs' performance will vary across different biochemistry topics, with potentially stronger performance in areas requiring systematic pathway analysis and weaker performance in topics requiring integration of clinical context.

Methods

Study Design

This study focused on a comparative analysis of the capabilities of different AI-driven LLMs in the medical biochemistry course. The research included an examination of 4 chatbots currently available to the public: Claude (Anthropic), GPT-4 (OpenAI), Gemini (Google), and Copilot (Microsoft).

A total of 200 scenario-based MCQs with 4 options and a single correct answer were randomly chosen from the medical biochemistry course's examination database for medical students and validated by 2 independent experts. The study did not include questions with images and tables. The selected questions encompassed various levels of complexity. They were distributed across 23 distinctive categories: structural proteins and associated diseases, globular proteins and hemoglobin, red blood cells (RBCs) and anemia, structure and function of amino acids, structure and function of proteins, bioenergetics and electron transport chain, enzymes, glycolysis and gluconeogenesis, glycogen, signaling mechanisms, pyruvate dehydrogenase and Krebs cycle, cholesterol metabolism, eicosanoids, fatty acid metabolism, fructose and galactose metabolism, hexose monophosphate pathway, ketone bodies, lipoproteins, lysosomal storage diseases, amino acid metabolism, fast and fed state, heme metabolism, and nitrogen metabolism.

Data Collection

For the testing phase, each selected chatbot was required to answer a set of 200 questions, and their performance was evaluated against the responses provided by medical students for the same set of questions. Claude 3.5 Sonnet, GPT-4 - 1106, Gemini 1.5 Flash, and Copilot proficiency in responding to MCQs was assessed in the last 2 weeks of August 2024. An OpenAI paid subscription was obtained to get GPT-4 access.

Each chatbot was given the prompt "generate the list of correct answers for the following MCQs" and provided with a first set of 50 questions; following with the same prompt and 3 more sets of 50 MCQs each, totally there were 200 MCQs in the questionnaire. After that, this procedure was repeated 5 times (no time period between the attempts was assigned). The results of 5 successive attempts by each chatbot to answer this questionnaire set were meticulously recorded in a Microsoft Excel spreadsheet and evaluated based on accuracy. A total of 4000 answers from LLMs were analyzed.

Five random answers were generated and analyzed for the same MCQ set using the RAND() function in Excel (Microsoft 365) to compare chatbot results with random guessing.

Data Analysis

The answers provided by each LLM were recorded and input into the Excel spreadsheet (Microsoft 365). The data from each (1-5) attempt was matched with the answer key and compared with all previous attempts, finding the percentage of repeated and correct answers among them. After that, a detailed item analysis was performed for each chatbot concerning different question categories.

Statistica 13.5.0.17 (TIBCO Software) was used to analyze the data's basic statistics. Considering the data's binary nature, the chi-square test was used to compare results among the different chatbots.

Results

Overview

According to our data, on average, 4 selected chatbots accurately answered 81.1% (SD 12%) of 200 MCQs from the medical biochemistry course. This result was 8.3% (P=.02) above the students' average (mean 72.8%, SD 12.7%) and almost 4 times better than randomly generated responses (mean 22%, SD 2.9%) for the same questions.

There was a significant variation in correct responses among the chatbots. The best result was recorded for Claude (92.5%, SD 0%), followed by GPT-4 (mean 85.1%, SD 1%) and Gemini (mean 78.5%, SD 0%), which were better than the students' average. Copilot showed the lowest result (mean 64%, SD 0%; Figure 1).

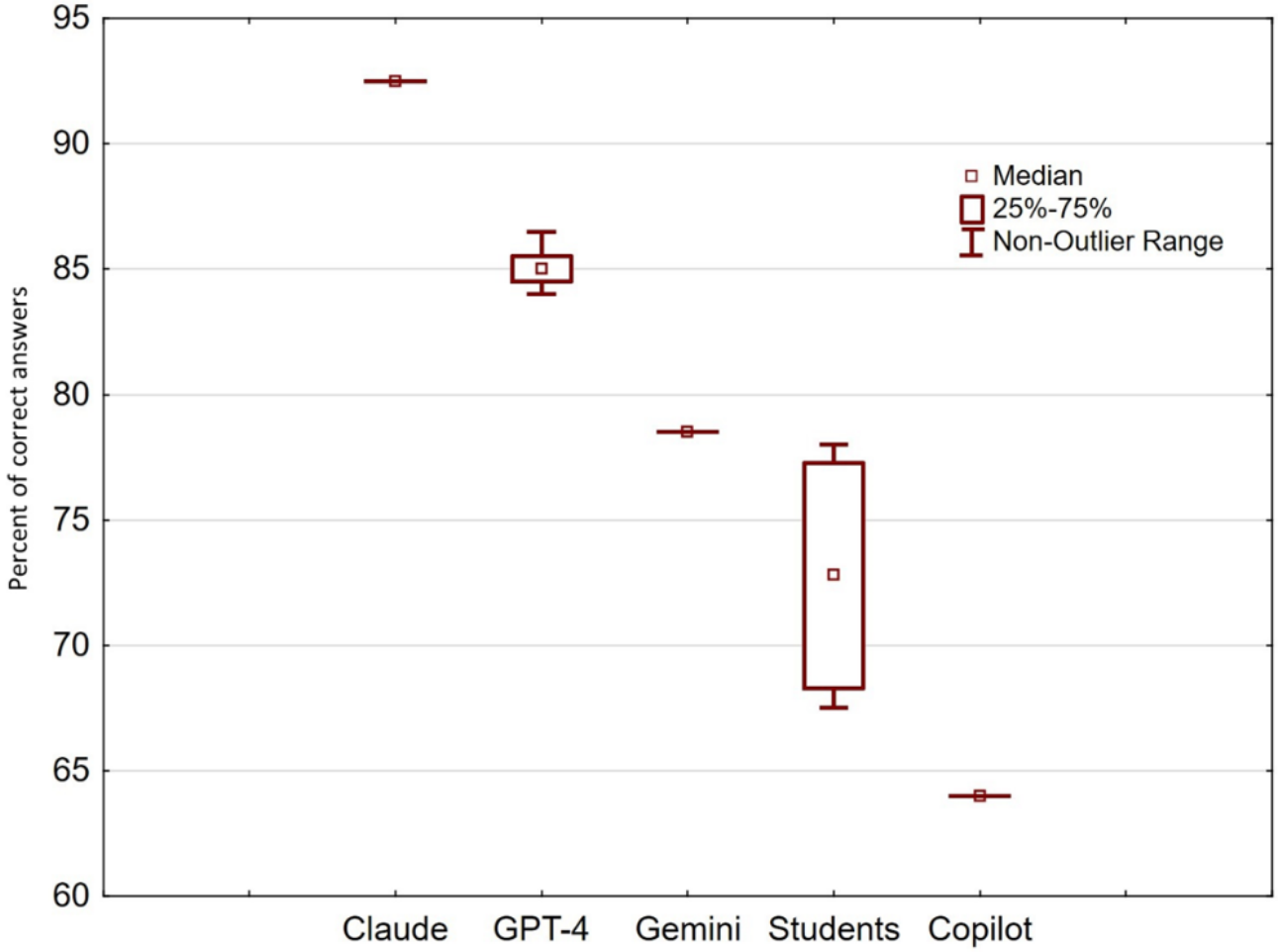
Interestingly, all chatbots answered 104 (52%) of the 200 questions correctly in all attempts. General item analysis revealed that eicosanoids, bioenergetics and electron transport chain, hexose monophosphate pathway, and ketone bodies were the 4 best topics, with the mean (SD) results for all chatbots being 100% (0%), 96.4% (7.2%), 91.7% (16.7%), and 93.8% (12.5%), respectively.

In contrast, the lowest results were recorded for globular proteins and hemoglobin (mean 58.4%, SD 26.4%), lipoproteins (mean 64.6%, SD 20.3%), and fructose and galactose metabolism questions (mean 65.8%, SD 29.9%).

After that, each chatbot's results for all 23 topics were evaluated (Figure 2).

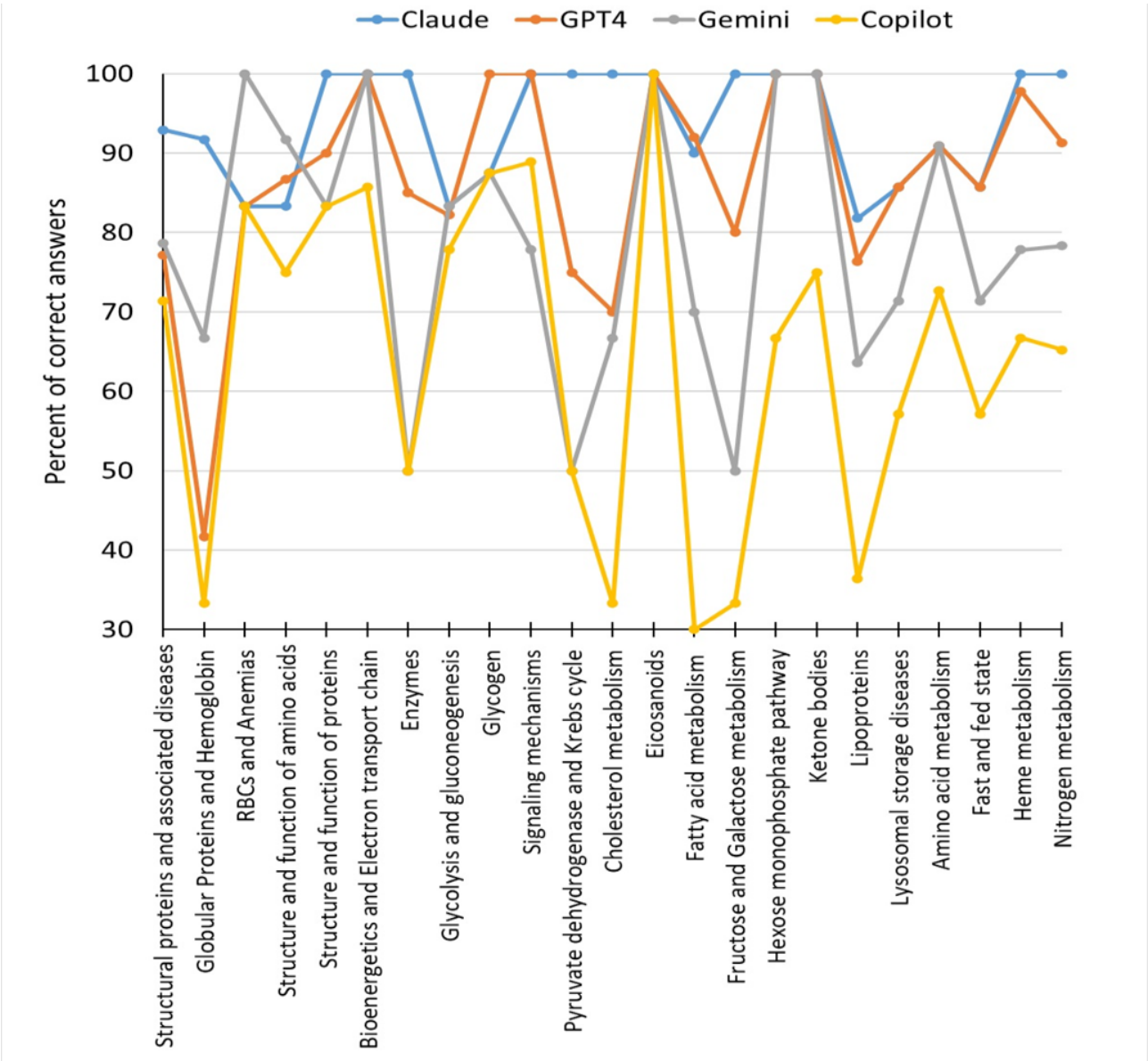
XSL•FO
RenderX

Figure 1. Percentile of correct answers from different chatbots and students on 200 multiple-choice questions from the medical biochemistry course.



XSL•FO
RenderX

Figure 2. Evaluation of chatbot performance in different topics of the medical biochemistry course. RBC: red blood cell.



Claude

Claude, offered by Anthropic, provided 92.5% (185/200) correct answers to the set of biochemistry MCQs. The answers in this second and all subsequent attempts were identical to the first. As the chatbot claims, its knowledge base has not changed between attempts, and it applies the same reasoning to answer each question. It was the best result among the 5 chatbots, 19.7% better than the student average and 70.5% superior to random guessing. The item analysis suggested that Claude correctly answered all questions (100%) from the following 12 categories: structure and function of proteins, bioenergetics and electron transport chain, enzymes, signaling mechanisms, pyruvate dehydrogenase and Krebs cycle, cholesterol metabolism, eicosanoids, fructose and galactose metabolism, hexose monophosphate pathway, ketone bodies, heme metabolism, and nitrogen metabolism. The lowest result (81.8%) was recorded for the lipoproteins. For the rest of the topics, the percentile of correct answers was 83.3% - 91.7%.

Claude did not solve 15 (7.5%) out of 200 MCQs from the entire questionnaire set. These were comprehensive questions about RBCs, hemoglobin, enzymes, biotin deficiency, and lipoproteins.

GPT-4

The results of 5 successive ChatGPT-4 (OpenAI) attempts to answer the set of 200 biochemistry MCQs showed 85.1% (SD 1%) correct answers on average. The best result of its 5 attempts was 86.5%, 13.7% better than the average for medical students and 64.5% above random guessing. The fourth attempt was the most successful; the mean results of the other 4 attempts were close to 85% (range 84% - 85.5%). The coincidence generated by GPT-4 answers with the previous attempts was 91.5% - 94.5%, and the coincidence of correct answers among them was in the range 81% - 83.6%.

Of the 200 questions, 158 (79%) were answered correctly across all 5 attempts and considered a solid knowledge area for GPT-4. Most of these MCQs were recall questions, but some were complex and required critical thinking. The item analysis indicated that the best 6 categories with 100% correct answers

XSL•FO
RenderX

were bioenergetics and electron transport chain, glycogen, signaling mechanisms, eicosanoids, hexose monophosphate pathway, and ketone bodies. The lowest result was recorded for globular proteins and hemoglobin questions—only 41.7% of the correct answers. For the rest of the topics, the percentile of correct answers was 77.1% - 97.8%.

GPT-4 did not answer 17 (8.5%) of the 200 MCQs from the entire questionnaire set in any 1 out of all 5 attempts. These were more comprehensive questions about defective proteins, oxygen saturation, anemia, amino acids, glycogen, glycolysis and gluconeogenesis, and lipoproteins.

Gemini

Google recently introduced Gemini as a successor to Bard. The results of 5 attempts by Gemini to answer the set of 200 biochemistry MCQs showed 157 (78.5%) correct answers, 5.7% above the average for medical students and 56.5% above the random answers. Unlike Bard, 5 successive attempts from Gemini were similar; the same answers were received.

The item analysis of these 157 correct answers shows that Gemini did the best (100% accurate) for questions in the following 5 categories: RBCs and anemia, bioenergetics and electron transport chain, eicosanoids, hexose monophosphate pathway, and ketone bodies. Most of these MCQs were recall questions. The lowest 50% results were recorded for the following 3 categories: enzymes, pyruvate dehydrogenase and Krebs cycle, and fructose and galactose metabolism. Gemini's responses in other topics were in the 63.6% - 91.7% interval. Gemini did not answer 43 (21.5%) of the 200 MCQs from the entire questionnaire set, which were comprehensive questions mostly about proteins, enzymes, the Krebs cycle, fatty acids, fructose, and galactose metabolism.

Copilot

Microsoft's Copilot can accept only up to 2000 characters in the prompt, so only 2 to 7 MCQs can be answered at a time, which is inconvenient to work with. The results received on the first try were not different from 4 successive attempts, so there was zero variation among all 5 attempts. Copilot generated 128 (64%) accurate answers for the same set of 200 MCQs from the biochemistry course, 8.8% lower than the average medical student but 42% better than random guessing.

The item analysis of these 126 correct answers indicated that these MCQs were mostly recall questions. The best result was shown for the eicosanoids category (100%), and the lowest was for fatty acid metabolism (only 30% of correct answers). Copilot's responses in other topics vary from 33.3% to 88.9%. Copilot did not answer 72 (36%) of the 200 MCQs from the questionnaire set. These questions concerned proteins, hemoglobin, amino acids, enzymes, fatty acids, pyruvate dehydrogenase, Krebs cycle, and fast and fed state.

Pearson Chi-Square Test Results

Table 1 shows the results of the Pearson chi-square test, which we used due to the binary nature of the data to compare the performance of the different AI-driven chatbots against each other.

The null hypothesis was rejected because the *P* value for all chatbots was less than α (*P*=.05), so there is a statistically significant association between the answers of all 4 chatbots.

Table . Pearson chi-square test results to compare the performance of Claude, GPT-4, Gemini, and Copilot against each other.

Large language models	Chi-square (*df*)	*P* value
Claude × GPT-4	19.7 (1)	<.001
Claude × Gemini	6.1 (1)	.01
Claude × Copilot	4.1 (1)	.04
GPT-4 × Gemini	33.1 (1)	<.001
GPT-4 × Copilot	15.9 (1)	<.001
Gemini × Copilot	23.5 (1)	<.001

Discussion

Principal Findings

Medical education is rapidly evolving, with AI playing an increasingly significant role. In this context, evaluating AI efficacy and relevancy to results is crucial, particularly given the precision and depth of understanding required in medical practice. AI-driven LLMs such as ChatGPT, Claude, Copilot, and Gemini have been compared against medical students in various studies, revealing both the strengths and limitations of AI in medical education. These comparisons show how AI can enhance human learning while also highlighting areas where it may not measure up. Research into AI's role in medical training has uncovered intriguing possibilities and important constraints [1,5,7].

MCQs form a cornerstone of assessment in medical education. Analyzing these questions is vital as it allows educators to assess their effectiveness in testing higher-order thinking and clinical reasoning skills, ensuring that assessments accurately reflect the competencies required for medical practice [18]. While LLMs have demonstrated impressive capabilities in answering queries and simulating scenarios, the depth and breadth of their understanding, particularly concerning MCQs in medical exams, still requires thorough evaluation [19].

The comparative analysis of LLMs and medical students in biochemistry assessment reveals several intriguing patterns that both confirm and challenge our initial hypotheses. While we anticipated comparable performance between AI models and medical students, the results demonstrated that LLMs not only matched but significantly exceeded student performance, with

XSL•FO
RenderX

an 8.3% higher average score (*P*=.02) across 200 medical biochemistry questions. This finding particularly supports our hypothesis regarding factual recall and concept application, though with a more pronounced advantage for AI systems than initially predicted. The observed variation in performance among different LLM platforms—ranging from Claude's exceptional 92.5% (185/200) accuracy to Copilot's more modest 64% (128/200)—aligns with our hypothesis about performance differences between AI models, suggesting that architectural and training differences significantly impact their capabilities in specialized medical knowledge domains.

Comparison to Literature

Recent studies have shown that LLMs, specifically GPT-4, often outperform medical students on MCQ items in board and licensing exams. This finding underscores the significance of MCQs in medical licensing exams, extensively used in crucial assessments worldwide. Examples include the Peruvian National Licensing Medical Examination, the United States Medical Licensing Examination (USMLE), the United Kingdom Medical Licensing Assessment (UKMLA), and the Australian Medical Council (AMC) Exam [20,24-26]. The widespread use of MCQs is attributed to their effectiveness in evaluating higher-order skills through complex clinical scenarios, analysis, and problem-solving. These questions assess students' ability to integrate information, reflecting real-world challenges and shaping competent professionals. It is well correlated with the results of our study, which have shown that the selected 4 chatbots answered correctly to 81.1% (SD 12%) of the 200 questions from the medical biochemistry course, which is 8.3% above the students' average.

Another comprehensive study compared the results of 4 LLMs across 163 questions from sample NBME (National Board of Medical Examiners) clinical subject exams. The results were striking: GPT-4 achieved a perfect score of 100% (163/163), significantly outperforming GPT-3.5, Claude, and Bard. GPT-3.5 scored 82.2% (134/163), Claude 84.7% (138/163), and Bard 75.5% (123/163). The statistical superiority of GPT-4 was evident, with no significant differences observed among the other 3 models [27]. Interestingly, while GPT-4 excelled across all subject exams, the different models demonstrated variable strengths. GPT-3.5 performed best in family medicine and obstetrics and gynecology, Claude in surgery, and Bard in surgery and neurology. The surgery exam yielded the highest average score across all models, while family medicine had the lowest. GPT-4's exceptional performance may be attributed to its extensive training data, which exceeded 45 terabytes by September 2021, despite not being specifically fine-tuned for medical data [10].

Our data contradict this clinical study and suggest that GPT-4 did well with 85% (170/200) of correct answers but is not currently the most proficient chatbot for biochemistry questions. The best result was recorded for Claude, with an impressive 92.5% (185/200) of the correct answers. Gemini took third place with 78.5% (157/200) of correct answers, which is still above the student's average of 72.8% (SD 5.2%) for the same questions. The lowest result was recorded for Copilot (128/200, 64%).

These findings highlight the potential of LLMs in medical education and practice. Their ability to tackle complex medical questions opens doors to innovative clinical decision support, research, and education applications. However, it is worth noting that GPT-4, the only LLM in this study not available for free, could be less accessible to a broad range of students, potentially limiting its widespread use in educational settings.

Several studies have evaluated ChatGPT's performance in biochemistry. One study examined GPT-3.5's potential as a self-study adjunct for medical students in biochemistry, using 200 questions. ChatGPT provided correct answers to 58% (116/200) of the biochemistry questions. While this performance allowed it to pass the university's medical biochemistry exam, the study suggests there is room for improvement in GPT-3.5 as a comprehensive and reliable self-learning tool [28].

Another study focused on ChatGPT's ability to address higher-order questions in medical biochemistry. Using GPT-3.5, researchers conducted a web-based cross-sectional study presenting 200 randomly selected, complex reasoning questions from an institutional question bank, classified according to CBME (Competency-Based Medical Education) curriculum modules. Two expert biochemistry academicians evaluated responses on a 0 - 5 scale. The AI achieved a median score of 4 (IQR 3.5-4.5), which was comparable to a hypothetical value of 4 (*P*=.16) but significantly lower than the maximum of 5 (*P*=.001). These results suggest that GPT-3.5 shows promise as an effective tool for addressing complex questions in medical biochemistry, demonstrating its potential in handling higher-order thinking tasks in this field [29].

Our research confirms that GPT-4 has significant improvements and is superior to GPT-3.5. Our data suggest that GPT-4 responded correctly to 84% - 86.5% of MCQs, and 79% answered correctly across all 5 attempts.

Implications of Findings

The implications of AI's performance in medical education extend beyond mere test-taking abilities. LLMs can answer complex medical questions that raise important questions about the future of medical education and topics in which LLMs demonstrate proficiency, so they may be used to assist students. The detailed analysis of MCQs in our study revealed that questions from 4 topics are well answered by all chatbots: eicosanoids, bioenergetics, electron transport chain, and ketone bodies. In contrast, the lowest results were recorded for globular proteins and hemoglobin, lipoproteins, and fructose and galactose metabolism questions. However, there was a significant difference in the 4 LLMs performances. Claude showed the most impressive results and answered all questions (100%) from 12 categories: structure and function of proteins, bioenergetics and electron transport chain, enzymes, signaling mechanisms, pyruvate dehydrogenase and Krebs cycle, cholesterol metabolism, eicosanoids, fructose and galactose metabolism, hexose monophosphate pathway, ketone bodies, heme, and nitrogen metabolism.

In conclusion, the rapid advancements in AI technology, particularly in medical education, present opportunities and challenges. While LLMs have shown impressive capabilities

XSL•FO
RenderX

in answering medical exam questions, it is crucial to remember that medical education encompasses more than just knowledge acquisition. Clinical skills, empathy, ethical decision-making, and the ability to navigate complex health care systems are all integral parts of medical training that current AI models may not fully capture.

As we progress, we must continuously evaluate AI's role in medical education, ensuring that it complements rather than replaces human expertise. Our findings also have important implications for assessment strategies in medical education. The ability of LLMs to surf the net and do better than medical students on MCQ-based evaluation is an assault on the traditional ways of measuring medical performance and calls for a better understanding of how medical knowledge and skills should be assessed. While such results provide ideas on how to develop a curriculum and manage educational resources, they also highlight the need to ensure that the value of AI in measuring certain aspects of medical training, such as clinical reasoning, interaction with patients, and even decision-making ethics, is always respected. This underscores the need for medical education to continue emphasizing the development of comprehensive clinical skills beyond what can be measured through standardized testing.

Future Directions

Future research in this field should pursue several key routes to better understand and implement AI technologies in medical education. Long-term studies are needed to evaluate the impact of LLM integration on student learning outcomes, particularly focusing on how AI-assisted learning affects knowledge retention, clinical reasoning development, and overall academic performance. These studies should incorporate diverse assessment methods beyond MCQs, including case-based scenarios, open-ended questions, and practical clinical applications of biochemistry knowledge across different medical disciplines to understand whether the observed performance patterns are consistent.

Strengths and Limitations

This study represents one of the first comprehensive comparisons between multiple leading LLMs and medical students in the specific context of medical biochemistry education. The large sample size of 200 questions provided a robust dataset for analysis, covering a broad spectrum of biochemistry topics typically encountered in medical education. The inclusion of multiple LLM platforms (GPT-4, Claude, Copilot, and Gemini) allowed for a nuanced comparison of AI capabilities across different models, providing valuable insights into their relative strengths and potential applications in medical education.

Several limitations should be considered when interpreting these results. This study's findings on different chatbot proficiencies are limited to MCQs from the biochemistry course, which may not represent other medical questions or contexts. In addition, the sample size of 200 questions, excluding questions with images or tables, may not capture the full range of difficulty levels or content areas.

LLMs receive regular updates, which result from training on inputs and tuning so that they may provide different answers depending on the testing date. Another limitation is that GPT-4, which performed well, is not freely available, potentially limiting its applicability in widespread educational settings.

Conclusions

LLMs such as ChatGPT, Claude, Copilot, and Gemini have impressive capabilities in answering MCQs, often outperforming medical students. In this study, the selected chatbots outperformed students' results. These findings highlight the potential of AI in medical education and assessment. Different LLMs exhibit varying strengths in different topics of medical biochemistry courses. In this study, Claude showed the best performance, followed by GPT-4, Gemini, and Copilot. This variability suggests that different AI models may have unique strengths in specific medical fields, which could be leveraged for targeted educational support. The strong performance of LLMs in answering complex medical questions raises important considerations for the future of medical education. While AI demonstrates proficiency in knowledge-based assessments, it is crucial to remember that medical training encompasses more than just information recall. Clinical reasoning, empathy, ethical decision-making, and navigating health care systems remain essential components that current AI models may need to capture fully.

Acknowledgments

The authors thank the Dean of the College of Medicine at Alfaisal University, Prof Khaled Al-Kattan, and the Head of the Department of Anatomy and Genetics, Prof Paul Ganguly, for their support in this research.

Conflicts of Interest

None declared.

References

1. Liu PR, Lu L, Zhang JY, Huo TT, Liu SX, Ye ZW. Application of artificial intelligence in medicine: an overview. Curr Med Sci 2021 Dec;41(6):1105-1115. [doi: 10.1007/s11596-021-2474-3] [Medline: 34874486]
2. Garcia-Vidal C, Sanjuan G, Puerta-Alcalde P, Moreno-García E, Soriano A. Artificial intelligence to support clinical decision-making processes. EBioMedicine 2019 Aug;46(27-29):27-29. [doi: 10.1016/j.ebiom.2019.07.019] [Medline: 31303500]

XSL•FO
RenderX

3. Ellahham S. Artificial intelligence: the future for diabetes care. Am J Med 2020 Aug;133(8):895-900. [doi: 10.1016/j.amjmed.2020.03.033] [Medline: 32325045]
4. Singhal K, Azizi S, Tu T, et al. Publisher correction: large language models encode clinical knowledge. Nature New Biol 2023 Aug;620(7973):E19. [doi: 10.1038/s41586-023-06455-0] [Medline: 37500979]
5. Meo SA, Al-Masri AA, Alotaibi M, Meo MZS, Meo MOS. ChatGPT knowledge evaluation in basic and clinical medical sciences: multiple choice question examination-based performance. Healthcare (Basel) 2023 Jul 17;11(14):2046. [doi: 10.3390/healthcare11142046] [Medline: 37510487]
6. Bolgova O, Shypilova I, Sankova L, Mavrych V. How well did ChatGPT perform in answering questions on different topics in gross anatomy? Eur J Med Health Sci 2023;5(6):94-100. [doi: 10.24018/ejmed.2023.5.6.1989]
7. Roos J, Kasapovic A, Jansen T, Kaczmarczyk R. Artificial intelligence in medical education: comparative analysis of ChatGPT, Bing, and medical students in Germany. JMIR Med Educ 2023 Sep 4;9:e46482. [doi: 10.2196/46482] [Medline: 37665620]
8. Agarwal M, Goswami A, Sharma P. Evaluating ChatGPT-3.5 and Claude-2 in answering and explaining conceptual medical physiology multiple-choice questions. Cureus 2023 Sep;15(9):e46222. [doi: 10.7759/cureus.46222]
9. Welcome to Claude. Anthropic. URL: https://docs.anthropic.com/en/docs/welcome [accessed 2024-09-06]
10. GPT-4 turbo and GPT-4. OpenAI. URL: https://tinyurl.com/ycec959y [accessed 2024-09-06]
11. Gemini models. Google AI. URL: https://ai.google.dev/gemini-api/docs/models/gemini [accessed 2024-09-06]
12. Microsoft 365 Copilot. Microsoft. URL: https://learn.microsoft.com/en-us/office365/servicedescriptions/office-365-platform-service-description/microsoft-365-copilot [accessed 2024-09-06]
13. Mavrych V, Ganguly P, Bolgova O. Using large language models (ChatGPT, Copilot, PaLM, Bard, and Gemini) in gross anatomy course: comparative analysis. Clin Anat 2025 Mar;38(2):200-210. [doi: 10.1002/ca.24244] [Medline: 39573871]
14. Laupichler MC, Rother JF, Grunwald Kadow IC, Ahmadi S, Raupach T. Large language models in medical education: comparing ChatGPT- to human-generated exam questions. Acad Med 2024 May 1;99(5):508-512. [doi: 10.1097/ACM.0000000000005626] [Medline: 38166323]
15. Liu M, Okuhara T, Chang X, et al. Performance of ChatGPT across different versions in medical licensing examinations worldwide: systematic review and meta-analysis. J Med Internet Res 2024 Jul 25;26:e60807. [doi: 10.2196/60807] [Medline: 39052324]
16. Mavrych V, Bolgova O. Evaluating AI performance in answering questions related to thoracic anatomy. MOJ Anat Physiol 2023;10(1):55-59. [doi: 10.15406/mojap.2023.10.00339]
17. Brin D, Sorin V, Vaid A, et al. Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. Sci Rep 2023 Oct 1;13(1):16492. [doi: 10.1038/s41598-023-43436-9] [Medline: 37779171]
18. Bharatha A, Ojeh N, Fazle Rabbi AM, et al. Comparing the performance of ChatGPT-4 and medical students on MCQs at varied levels of Bloom's taxonomy. Adv Med Educ Pract 2024;15(393-400):393-400. [doi: 10.2147/AMEP.S457408] [Medline: 38751805]
19. Goyal M, Agarwal M, Goel A. Interactive learning: online audience response system and multiple choice questions improve student participation in lectures. Cureus 2023;15(7):e42527. [doi: 10.7759/cureus.42527]
20. Torres-Zegarra BC, Rios-Garcia W, Ñaña-Cordova AM, et al. Performance of ChatGPT, Bard, Claude, and Bing on the Peruvian National Licensing Medical Examination: a cross-sectional study. J Educ Eval Health Prof 2023;20:30. [doi: 10.3352/jeehp.2023.20.30] [Medline: 37981579]
21. Gilson A, Safranek CW, Huang T, et al. Correction: How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2024 Feb 27;10:e57594. [doi: 10.2196/57594] [Medline: 38412478]
22. Friederichs H, Friederichs WJ, März M. ChatGPT in medical school: how successful is AI in progress testing? Med Educ Online 2023 Dec;28(1):2220920. [doi: 10.1080/10872981.2023.2220920] [Medline: 37307503]
23. Lin Z. Why and how to embrace AI such as ChatGPT in your academic life. R Soc Open Sci 2023 Aug;10(8):230658. [doi: 10.1098/rsos.230658] [Medline: 37621662]
24. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023 Feb;2(2):e0000198. [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
25. Lai UH, Wu KS, Hsu TY, Kan JKC. Evaluating the performance of ChatGPT-4 on the United Kingdom Medical Licensing Assessment. Front Med (Lausanne) 2023;10:1240915. [doi: 10.3389/fmed.2023.1240915] [Medline: 37795422]
26. Kleinig O, Gao C, Bacchi S. This too shall pass: the performance of ChatGPT-3.5, ChatGPT-4 and New Bing in an Australian medical licensing examination. Med J Aust 2023 Sep 4;219(5):237. [doi: 10.5694/mja2.52061] [Medline: 37528548]
27. Abbas A, Rehman MS, Rehman SS. Comparing the performance of popular large language models on the National Board of Medical Examiners sample questions. Cureus 2024 Mar;16(3):e55991. [doi: 10.7759/cureus.55991] [Medline: 38606229]
28. Surapaneni KM, Rajajagadeesan A, Goudhaman L, et al. Evaluating ChatGPT as a self-learning tool in medical biochemistry: a performance assessment in undergraduate medical university examination. Biochem Mol Biol Educ 2024;52(2):237-248. [doi: 10.1002/bmb.21808] [Medline: 38112255]

XSL•FO
RenderX

29. Ghosh A, Bir A. Evaluating ChatGPT's ability to solve higher-order questions on the competency-based medical education curriculum in medical biochemistry. Cureus 2023;15(4):e37023. [doi: 10.7759/cureus.37023]

Abbreviations

AI: artificial intelligence
AMC: Australian Medical Council
CBME: Competency-Based Medical Education
LLMs: large language models
MCQ: multiple-choice question
NBME: National Board of Medical Examiners
RBC: red blood cell
UKMLA: United Kingdom Medical Licensing Assessment
USMLE: United States Medical Licensing Examination

Edited by B Lesselroth; submitted 06.10.24; peer-reviewed by M Agarwal, M Malhotra; revised version received 20.01.25; accepted 08.03.25; published 10.04.25.

Please cite as:
Bolgova O, Shypilova I, Mavrych V
Large Language Models in Biochemistry Education: Comparative Evaluation of Performance
JMIR Med Educ 2025;11:e67244
URL: https://mededu.jmir.org/2025/1/e67244
doi:10.2196/67244

© Olena Bolgova, Inna Shypilova, Volodymyr Mavrych. Originally published in JMIR Medical Education (https://mededu.jmir.org), 10.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Medical Students' Acceptance of Tailored e–Mental Health Apps to Foster Their Mental Health: Cross-Sectional Study

Catharina Grüneberg[1]; Alexander Bäuerle[1,2], PhD; Sophia Karunakaran[1], MSc; Dogus Darici[3], Dr med; Nora Dörrie[1,2], Dr med; Martin Teufel[1,2], Prof Dr med; Sven Benson[2,4], Prof Dr; Anita Robitzsch[1,2], Dr med

[1]Clinic for Psychosomatic Medicine and Psychotherapy, LVR-University Hospital, University of Duisburg-Essen, Virchowstraße 174, Essen, Germany

[2]Center for Translational Neuro- and Behavioral Sciences, University of Duisburg-Essen, Essen, Germany

[3]Institute of Anatomy and Neurobiology, University of Münster, Münster, Germany

[4]Institute for Medical Education, University Hospital Essen, University of Duisburg-Essen, Essen, Germany

Corresponding Author:
Anita Robitzsch, Dr med
Clinic for Psychosomatic Medicine and Psychotherapy, LVR-University Hospital, University of Duisburg-Essen, Virchowstraße 174, Essen, Germany

Abstract

Background: Despite the high prevalence of mental health problems among medical students and physicians, help-seeking remains low. Digital mental health approaches offer beneficial opportunities to increase well-being, for example, via mobile apps.

Objective: This study aimed to assess the acceptance, and its underlying predictors, of tailored e–mental health apps among medical students by focusing on stress management and the promotion of personal skills.

Methods: From November 2022 to July 2023, a cross-sectional study was conducted with 245 medical students at the University of Duisburg-Essen, Germany. Sociodemographic, mental health, and eHealth-related data were assessed. The Unified Theory of Acceptance and Use of Technology (UTAUT) was applied. Differences in acceptance were examined and a multiple hierarchical regression analysis was conducted.

Results: The general acceptance of tailored e–mental health apps among medical students was high (mean 3.72, SD 0.92). Students with a job besides medical school reported higher acceptance ($t_{107.3}$=–2.16; P=.03; P_{adj}=.027; Cohen d=4.13) as well as students with higher loads of anxiety symptoms ($t_{92.4}$=2.36; P=.02; P_{adj}=.03; Cohen d=0.35). The t values were estimated using a 2-tailed t test. Regression analysis revealed that acceptance was significantly predicted by anxiety symptoms (β=.11; P=.045), depressive symptoms (β=–.11; P=.05), internet anxiety (β=–.12; P=.01), digital overload (β=.1; P=.03), and the 3 UTAUT core predictors—performance expectancy (β=.24; P<.001), effort expectancy (β=.26; P<.001), and social influence (β=.43; P<.001).

Conclusions: The high acceptance of e–mental health apps among medical students and its predictors lay a valuable basis for the development and implementation of tailored e–mental health apps within medical education to foster their mental health. More research using validated measures is needed to replicate our findings and to further investigate medical students' specific needs and demands regarding the framework of tailored e–mental health apps.

(JMIR Med Educ 2025;11:e58183) doi:10.2196/58183

KEYWORDS

eHealth; medical education; medical students; tailored interventions; UTAUT; intention to use; e–mental health apps; app; foster; cross-sectional study; mental health problems; physician; well-being; mobile apps; acceptance; assessment; mental health apps

Introduction

Background

Medical students have a heightened incidence of mental health problems, namely anxiety [1,2] and depression [3-5], and are confronted with stressful situations throughout their careers [6,7]. Elevated levels of depression and anxiety among medical students and physicians exert considerable influence on personal well-being and patient safety [8], emphasizing the urgent need for targeted preventive and support programs [7,9-11]. The necessity for assessable and easily accessible interventions to foster mental health and well-being is of utmost importance in the medical student population [12,13].

In recent years, the surge in the significance of digitalization within health care and medical education has been noteworthy [14-16]. This trend has been particularly pronounced during the COVID-19 pandemic and persists afterward [17]. A report disseminated by a German health insurance entity in 2023

XSL•FO
RenderX

scrutinized students' health, with a specific focus on postpandemic developments and the pivotal role of digital education and instruction [18]. The report underscored the critical importance of stress prevention and mental health initiatives [18]. Digital mental health approaches present promising avenues for surmounting barriers and enhancing the use of mental health support, for example, through mobile apps [13,19,20].

Analyzing factors influencing the acceptance of a mobile app is essential, and further research on actual uptake, adoption, and adherence is needed [21-26]. Incorporating future users directly into the development process is crucial for optimizing the adherence of new technologies and should be focused within research [27,28].

Few studies have delved into e–mental health promotion and the prevention of psychological distress among medical students and have shown that uptake of mental health support remains low due to barriers such as mental health stigma or data safety [6,29-31]. To date and to the best of our knowledge, no study has examined the acceptance of tailored e–mental health apps among medical students using a validated model. For this reason, the Unified Theory of Acceptance and Use of Technology (UTAUT) was applied in this study to lay the foundation for the development of an application especially tailored to the students' needs and demands to foster mental health by focusing on stress management and promotion of personal skills at University Duisburg-Essen. The UTAUT evaluates the acceptance of technological systems consisting of 4 primary predictors—performance expectancy (PE), effort expectancy (EE), social influence (SI), and facilitating conditions (FC)—and has been adjusted to investigate the acceptance of eHealth interventions along with their underlying factors [23-25]. Numerous studies have used the UTAUT framework in the context of eHealth interventions among different samples [32-35].

Objectives

Due to the evident progression of digitalization and its concomitant potential to enhance mental health while simultaneously acknowledging the existing impediments to leveraging these opportunities, this study is specifically oriented toward investigating the acceptance of tailored e–mental health apps and their foundational predictors among medical students, using the validated UTAUT model as the analytical framework.

While prior research has underscored the significance of promoting mental health among medical students [7,9,12,31], limited attention has been given to evaluate e–mental health approaches focusing on the promotion and the prevention of psychological distress among medical students using validated measures, such as the UTAUT model, and tailored approaches [28,36,37].

This study will address the following research questions: (1) What is the extent of acceptance of e–mental health apps among medical students? (2) Are there differences in acceptance among medical students based on sociodemographic and mental health data? (3) What factors predict acceptance among medical students?

Methods

Study Design and Participants

A cross-sectional study was conducted to assess acceptance and to analyze drivers and barriers of tailored e–mental health apps among medical students. The study was presented to medical students in the 5th year at the Medical Faculty of University Duisburg-Essen, North-Rhine-Westphalia, Germany, during the course of psychosomatic medicine. Following the course, students were given the opportunity to participate voluntarily. The participants of the study were recruited from November 2022 to July 2023. Of the 305 students attending the course, 245 (80.3%) students gave their informed consent to participate in the study. Of the 245 participants, 16 (6.5%) participants were eliminated from the sample because of missing data. In total, 229 (93.5%) students were included in the final data analysis. We applied no inclusion or exclusion criteria. Medical students were invited to participate in the study through direct contact in the context of psychosomatic medicine courses. All participants were aged 18 years or above.

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki and has been approved by the ethics committee of the Medical Faculty of the University of Duisburg-Essen (21 - 10196-BO). Participation was anonymous, voluntary, and without any compensation. Prior to the start of the questionnaire, written informed consent was obtained and the students received background information on the purpose of the study.

Assessment Instruments

The survey consisted of a paper-pencil questionnaire with self-developed items. Additionally, validated scales were used. The measures encompassed sociodemographic, eHealth-related, and mental health data. The primary outcome was the acceptance of an e–mental health app by using the conceptual framework of the UTAUT model's theory.

Sociodemographic Data

Sociodemographic data contained age, gender, marital status, employment besides medical school (occupational status), and working hours per week (0 - 5, 5 - 10, 10 - 15, and >15 hours).

Mental Health Data

To obtain mental health data, the validated PHQ-4 (Patient Health Questionnaire-4) measure consisting of two 2-item measures—PHQ-2 (symptoms of depression, Patient Health Questionnaire-2) and GAD-2 (symptoms of general anxiety disorder, Generalized Anxiety Disorder-2)—were used [38,39]. Answers were given on a 4-point Likert scale (0="never" to 3="nearly every day"). A cutoff score of 3 or more is described to be an indicator of depression (PHQ-2) [38] or general anxiety (GAD-2) [40]. Internal consistencies measured by the Cronbach α were sufficient with α=0.82 (95% CI 0.76 - 0.87) for GAD-2 and α=0.81 (95% CI 0.73 - 0.86) for PHQ-2. Self-generated questions were used to assess life quality (0="very low" to 10="very good"), mental health (0="very low" to 10="very good"), physical health (0="very low" to 10="very good"), and

XSL•FO
RenderX

importance of promoting mental well-being (0="not important" to 10="very important") on numerical rating scales.

eHealth-Related Data

eHealth-related data were assessed by measuring digital overload, internet anxiety, and digital competence. Internet anxiety and digital overload were both measured on a 5-point Likert scale (1="strongly disagree" to 5="strongly agree"). Internal consistency measured by the Cronbach α was low to sufficient with α=0.68 (95% CI 0.6 - 0.75) for the digital overload scale and sufficient with α=0.81 (95% CI 0.72 - 0.87) for the internet anxiety scale. These scales were previously published and established [34,35,41,42]. Digital competence was measured with a numerical rating scale (0="low" to 10="high").

Acceptance and UTAUT Predictors

To assess medical students' acceptance of using tailored e–mental health apps, a modified UTAUT questionnaire [24] was applied. The adapted UTAUT model consisted of 14 items and measured items on a 5-point Likert scale (1="strongly disagree" to 5="strongly agree"). Acceptance, operationalized as behavioral intention (BI) to use technology, is forecasted by PE, EE, and SI [25]. PE reflects the individual's belief in the benefits they will derive from using the technology. EE signifies the perceived ease of use. SI gauges the extent to which an individual believes that their relatives or friends would endorse the use of the technology. Four items were used to assess BI and PE. Acceptance, operationalized as BI, represented the dependent variable. Two predictors of acceptance—EE and SI—were measured with 3 items each. Internal consistency (Cronbach α) was excellent for BI (α=0.91, 95% CI 0.89 - 0.93) and PE (α=0.92, 95% CI 0.89 - 0.94), sufficient for SI (α=0.83, 95% CI 0.77 - 0.87), and low to sufficient for EE (α=.67, 95% CI 0.57 - 0.75).

Statistical Analysis

For data and statistical analysis, SPSS Statistics version 26 (IBM Corp) and R through RStudio version 4.3.1 (The R Foundation for Statistical Computing; Posit Software) were used. The raw data were collected from the survey, extracted, and processed. Relevant assumptions and prerequisites were tested prior to any statistical test [43-46]. The level of significance was set at α=0.05 for all tests. To minimize α error inflation for multiple comparisons Bonferroni correction was used and *P* values were adjusted. Sum scores (PHQ-4 scale, PHQ-2 scale, and GAD-2 scale) and mean scores (internet anxiety and digital overload) were computed. Mean scores for the UTAUT model were computed: BI, PE, EE, and SI. Consistent with previous research, acceptance scores, operationalized as BI, were categorized as "low acceptance" from 1 to 2.34, "moderate acceptance" from 2.35 to 3.67, and "high acceptance" from 3.68 to 5 [33,41,47]. Descriptive statistics (percentage and absolute count, mean scores, distributions, and standard deviations) of scales, items, and acceptance categories were performed. Additionally, explorative data analysis was conducted. Internal consistencies such as the Cronbach α and item-total correlation were calculated for scales. The normal distribution of the dependent variable (acceptance) was tested graphically and by the Kolmogorov-Smirnov test. Although violations against normal distribution were detected, parametric tests could be used according to the central limit theorem (n>30) and the robustness of the *t* test and Welch-ANOVA against normal distribution violations [44]. Means of acceptance were compared between groups using the *t* test (occupational status, PHQ-2, and GAD-2) and Welch-ANOVA (gender and marital status). The predictive model of acceptance was tested using multiple hierarchical regression analyses. The following predictors were included stepwise: sociodemographic data, mental health data (PHQ-2 and GAD-2), eHealth-related data, and the UTAUT core predictors (EE, SI, and PE). Linearity could be assumed and was analyzed using a scatter plot of the residuals against fitted values. Multicollinearity was not detected because all values of the variance inflation factor were <5. The normality of residuals could be assumed due to the central limit theorem. Homoscedasticity was proven and analyzed using a scatter plot of the standardized residuals and adjusted predicted values. According to Cohen *d*, effect sizes were reported and interpreted as small (0.2), medium (0.5), and large (0.8) [48].

Results

Study Population

In this sample, participants' age ranged from 20 to 37 years (mean 25.05, SD 2.82 years). Medical students experienced low digital overload (mean 2.85, SD 0.92) and low internet anxiety (mean 1.72, SD 0.79). Digital competence was high among medical students (mean 6.97, SD 1.72; range 0 - 10). For detailed characteristics, see Table 1.

XSL•FO
RenderX

Table . Sociodemographic and mental health data of participants (n=229).

Variable		N (%)	Mean (SD)	Acceptance, n (%)		
				Low[a]	Moderate[b]	High[c]
Gender						
	Woman	157 (68.6)	—[d]	13 (8.3)	42 (26.8)	102 (65)
	Man	70 (30.6)	—	9 (12.9)	21 (30)	40 (57.1)
	Nonbinary	2 (0.9)	—	0 (0)	2 (100)	0 (0)
Marital status (n=228)						
	Single, divorced, or separated	139 (61)	—	14 (10.1)	45 (32.4)	80 (57.6)
	Married or in a relationship	89 (39)	—	8 (9)	20 (22.5)	61 (68.5)
Job						
	Yes	165 (72.1)	—	14 (8.5)	43 (26.1)	108 (65.5)
	No	64 (28)	—	8 (12.5)	22 (34.4)	34 (53.1)
Working hours per week (n=166)						
	0 - 5	37 (22.3)	—	2 (5.4)	9 (24.3)	26 (70.3)
	5 - 10	84 (50.6)	—	8 (3.8)	21 (25)	55 (65.5)
	10 - 15	25 (15.1)	—	2 (8)	7 (28)	16 (64)
	>15	20 (12.1)	—	2 (10)	7 (35)	11 (55)
Mental health[e]		—	6.82 (1.72)	7.4 (2.2)	6.9 (2.3)	6.7 (2.4)
Physical health[e]		—	8.00 (1.84)	8.5 (1.6)	8 (1.8)	7.9 (1.9)
Life quality[e]		—	7.92 (1.67)	8.4 (1.1)	7.9 (1.7)	7.9 (1.7)
Promotion of mental well-being[e]		—	8.68 (1.80)	7.9 (2.3)	7.9 (1.7)	9 (1.7)
PHQ-2[f] score (range 0 - 6)		—	1.26 (1.42)	—	—	—
	Low (≤2)	201 (87.8)	0.84 (0.84)	21 (10.5)	57 (28.4)	123 (61.2)
	High (≥3)	28 (12.2)	4.25 (1.11)	1 (1.7)	8 (28.6)	19 (67.9)
GAD-2[g] score (range 0 - 6)		—	1.85 (1.51)	—	—	—
	Low (≤2)	178 (77.7)	1.19 (0.76)	20 (11.2)	52 (29.2)	106 (59.6)
	High (≥3)	51 (22.3)	4.16 (1.17)	2 (3.9)	13 (25.5)	36 (70.6)

[a]Low acceptance, with scores ranging from 1 to 2.34.

[b]Moderate acceptance, with scores ranging from 2.35 to 3.67.

[c]High acceptance, with scores ranging from 3.68 to 5.

[d]Not applicable.

[e]Higher scores indicate higher levels of mental health, physical health, life quality, or importance of promoting mental well-being (range 0 - 10).

[f]PHQ-2: Patient Health Questionnaire-2.

[g]GAD-2: Generalized Anxiety Disorder-2.

Acceptance of Tailored e–Mental Health Apps

The general acceptance of tailored e–mental health apps among medical students was high (mean 3.72, SD 0.92). Dividing the acceptance categories from low to high, 62% (142/229) participants showed high acceptance (mean 4.31, SD 0.45), 28.4% (65/229) showed moderate acceptance (mean 3.11, SD 0.28), and 9.6% (22/229) showed low acceptance (mean 1.76, SD 0.42).

Between groups, significant differences in acceptance were identified between occupational status ($t_{107.3}$=–2.16; P=.03; P_{adj}=.03; Cohen d=4.13) and GAD-2 groups ($t_{92.4}$=2.36; P=.02; P_{adj}=.03; Cohen d=0.35) using a 2-tailed t test. Students with a job besides medical school reported higher acceptance of tailored e–mental health apps than students without a job. Medical students with high GAD-2 levels (high load of anxiety symptoms) showed higher acceptance than students with low GAD-2 levels (low load of anxiety symptoms). No significant

XSL•FO
RenderX

differences between acceptance were found regarding PHQ-2 groups (low and high), gender (female, male, and divers), and marital status via ANOVA and *t* test (P_{adj}>.5).

Hierarchical Linear Regression Analysis and Predictors of Acceptance

A hierarchical linear regression analysis was conducted to evaluate predictors of acceptance among medical students regarding tailored e–mental health apps.

Sociodemographic data were included in the first step, explaining 3.6% of the variance in acceptance (R^2=0.036; R^2_{adj}=0.022; $F_{3,222}$=2.72; P=.045). Occupational status emerged as a significant positive predictor (β=.31; P=.03).

In the second step, mental health data were added to the analysis, increasing the explained variance to 6.4% (R^2=0.064; R^2_{adj}=0.042; $F_{5,220}$=2.99; P=.01). GAD-2 was identified as a significant predictor (β=.12; P=.03) of acceptance.

In the third step, eHealth-related data were added to the model, which further explained 8.2% of the variance in acceptance (R^2=0.082; R^2_{adj}=0.048; $F_{8,217}$=2.14; P=.02).

In the fourth and final step, the UTAUT predictors (EE, PE, and SI) were added (overall model), resulting in a comprehensive model that explained 65.8% of the variance in acceptance (R^2=0.658; R^2_{adj}=0.647; $F_{11,214}$=37.47; P<.001). The following variables (UTAUT core predictors) showed a significant positive prediction: UTAUT PE (β=.22, P<.001), UTAUT EE (β=.32, P<.001), and UTAUT SI (β=.44; P<.001).

To sum up, within the overall model, the UTAUT predictors, PHQ-2 and GAD-2 sum scores, internet anxiety, and digital overload were associated with the acceptance of tailored e–mental health apps among medical students. For a detailed overview of the hierarchical regression model of acceptance, see Table 2.

XSL•FO
RenderX

Table . Hierarchical regression model of acceptance (the extended Unified Theory of Acceptance and Use of Technology model; n=226).

Predictors		β[a]	β[b]	*t* [c]	R^2 [d]	ΔR^2[e]	*P* value
Intercept		–.22	–.00	–0.46	—[f]	—	.64
Step 1: Sociodemographic data		—	—	—	0.036	0.036	—
	Gender	.04	.02	0.53	—	—	.59
	Age	.01	.03	0.80	—	—	.43
	Occupational status	.16	.08	1.78	—	—	.08
Step 2[g]: Mental health data		—	—	—	0.064	0.028	—
	PHQ-2[h], sum score	–.07	–.11	–1.98	—	—	.05
	GAD-2[i], sum score	.07	.11	2.02	—	—	.04
Step 3[g]: eHealth-related data		—	—	—	0.082	0.018	—
	Digital overload	.10	.10	2.18	—	—	.03
	Internet anxiety	–.14	–.12	–2.49	—	—	.01
	Digital competence	.01	.03	0.47	—	—	.64
Step 4[g]: UTAUT[j] core predictors		—	—	—	0.658	0.576	—
	Social influence	.44	.43	7.57	—	—	<.001
	Performance expectancy	.22	.24	4.31	—	—	<.001
	Effort expectancy	.32	.26	5.24	—	—	<.001

[a]Unstandardized coefficient beta.
[b]Standardized coefficient beta.
[c]Test statistics were estimated using a 2-tailed *t* test.
[d]Multiple R^2 reported, determination coefficient.
[e]Changes in R^2.
[f]Not applicable.
[g]In steps 2, 3, and 4, only the newly included variables are presented.
[h]PHQ-2: Patient Health Questionnaire-2.
[i]GAD-2: Generalized Anxiety Disorder-2.
[j]UTAUT: Unified Theory of Acceptance and Use of Technology.

Discussion

Principal Findings

This study focused on examining the acceptance of tailored e–mental health apps and the factors influencing their use to promote medical students' mental health.

The general acceptance was high. Students with a job besides medical school reported higher acceptance as well as students with higher loads of anxiety symptoms. Acceptance was significantly predicted by occupational status, anxiety symptoms, depressive symptoms, internet anxiety, digital overload, and the 3 UTAUT core predictors—PE, EE, and SI.

The participants in this sample reported higher overall acceptance compared to previous research involving different target groups [23,33,42]. A qualitative study conducted by Dederichs et al [12] corroborates our findings, elucidating universally positive perspectives among medical students regarding internet- and mobile-based interventions. Preceding investigations have posited that augmented levels of educational attainment are concomitant with elevated acceptance scores [32,49], concurrently accentuating the advantages of e–mental health methodologies, including their low-threshold nature, temporal flexibility, and provision of anonymous support [12].

Among our cohort, self-rated promotion of mental well-being was highly valued, indicating general interest in mental health promotion as an important prerequisite and determinant of increasing acceptance.

The UTAUT core predictors elucidated the majority of the variance in acceptance, substantiating the model's efficacy in appraising e–mental health acceptance among medical students

XSL•FO
RenderX

and aligning with antecedent research [25,33,42]. Despite prior investigations indicating age [25,32,42] and gender [25,49] as salient determinants influencing acceptance within heterogeneous populations, these variables did not achieve statistical significance in this study. This lack of significance may be attributed to the existence of comparable stress factors affecting all participants uniformly.

A notable proportion, 12.2% (28/229), displayed indicators suggestive of depressive symptoms (PHQ-2), while 22.3% (51/229) exhibited symptoms indicative of a general anxiety disorder (GAD-2). These findings are consistent with extant research documenting the psychological vulnerability of medical students, illustrating elevated levels of anxiety and depression [1,6,50]. This underscores the imperative for psychological support interventions [2,3,10]. Our analysis revealed that mental health data concerning anxiety symptoms positively predicted acceptance within our model, aligning with prior research [51]. In contrast to that, depressive symptoms were associated with lower acceptance within our model. The acceptability may be decreased among students with higher depressive symptoms due to fear of additional loads. Furthermore, barriers, such as mental health stigma or data safety, were described as known challenges within previous research focusing on help-seeking behavior [30,36]. Additional information and educative programs or interventions may have beneficial effects to increase help-seeking and decrease stigma [29,52-55], but their impact needs to be investigated further.

Students concurrently managing part-time employment and medical school responsibilities demonstrated higher acceptance scores. Research specifically focusing on the mental health of working medical students is scarce [9,56]. Based on the findings, we would suggest that the additional load due to a part-time job results in higher acceptance levels of mental health support programs but this needs to be investigated further.

A study by Joiner et al [57] found that individuals born after 1993 exhibited lower internet anxiety and higher internet identification, reinforcing our findings. In our sample, most of the participants were born in the 1990s and 2000s. Internet anxiety and digital overload were observed at low levels and significant predictors of acceptance in the overall regression model. Aligning with previous research [23], high levels of internet anxiety were associated with decreased acceptance.

Digital competence was high within our sample. High internet identification and regular use of digital media might have influenced digital competence within our sample. Information on digital skills [58], preventive strategies, and digitalization need to be integrated further within medical education [15].

While acceptance and potential usage constitute crucial prerequisites for the implementation of digital approaches [23,59], it is imperative to acknowledge additional factors, including barriers and risks associated with the promotion of such approaches. Notably, skepticism and a lack of knowledge regarding e–mental health apps among medical students underscore the necessity for augmented information dissemination and increased personal experience with digital health approaches [22,36]. Attention must be directed toward addressing stigma and concerns related to data security [30,36]. Comprehensive assessments of additional barriers influencing actual usage and dropout rates are warranted in the implementation of e–mental health approaches [19,60].

The outcomes of this study establish a foundational framework for subsequent research endeavors and the implementation of e–mental health apps within the realm of medical education. The imperative for further implementation and rigorous evaluation of digital interventions for medical students is underscored.

Limitations

This study has limitations that should be considered when interpreting the presented results. It should be noted that studies assessing medical students' acceptance with validated instruments are still scarce and comparability is limited. The cross-sectional design does not allow causal inferences. Overall, overrepresentation may diminish representativeness, generalizability, and external validity, which is a common bias in research. In the context of a tailored design approach, additional stakeholders should be integrated into future studies [61]. The intention-behavior gap should be considered, as our study assessed theoretical willingness rather than actual usage. Within this study, the Cronbach α, a conservative measure assessing reliability, was used, and it should be noted that the Cronbach α of the EE scale and digital overload scale were lower compared to those observed in previous studies [33,35,41,42]. One possible explanation may be inconsistent response patterns; therefore, the interpretation should be done with caution. According to previous studies [21-28], adherence, actual usage, and dropout rates of e–mental health approaches should be investigated further. While the 3 fundamental predictors of the UTAUT model—EE, PE, and SI—remain crucial, additional factors should be focused on to comprehensively grasp and optimize acceptance levels further.

Conclusions

In this investigation, the focus was on evaluating the acceptance of tailored e–mental health apps and its influencing factors in promoting medical students' mental health. The overall acceptance was found to be high, with students having part-time jobs alongside medical school and students with elevated anxiety levels reporting even higher levels of acceptance. Besides the 3 UTAUT core predictors (PE, EE, and SI), additional significant predictors influence acceptance among medical students including occupational status, anxiety symptoms, depressive symptoms, internet anxiety, and digital overload. As digitalization transforms the medical sector, integrating supportive digital tools into medical education requires a focus on promoting a healthy learning environment and well-being among future physicians. Preventive strategies, including addressing barriers like stigma, are crucial. This study contributes valuable insights in order to develop and implement a digital application to foster medical students' mental health focusing on stress management and promotion of personal skills at Medical University Duisburg-Essen, Germany.

XSL•FO
RenderX

Acknowledgments

We gratefully acknowledge the support of the Open Access Publication Fund at the University of Duisburg-Essen.

Data Availability

The datasets analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: AR, MT, AB
Data curation: AR, AB, CG
Formal analysis: CG, SK
Investigation: AR, MT, AB, SB
Methodology: AR, MT, AB, SB
Project administration: AR, AB, SB
Supervision: AR, MT, AB, SB
Writing – original draft: CG
Writing – review & editing: SB, DD, ND

Conflicts of Interest

None declared.

References

1. Quek TTC, Tam WWS, Tran BX, et al. The global prevalence of anxiety among medical students: a meta-analysis. Int J Environ Res Public Health 2019 Jul 31;16(15):2735. [doi: 10.3390/ijerph16152735] [Medline: 31370266]
2. Thapa B, Sapkota S, Khanal A, Aryal BK, Hu Y. Predictors of depression and anxiety among medical students. J Nepal Health Res Counc 2023 Sep 8;21(1):63-70. [doi: 10.33314/jnhrc.v21i1.4514] [Medline: 37742151]
3. Rotenstein LS, Ramos MA, Torre M, et al. Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: a systematic review and meta-analysis. JAMA 2016 Dec 6;316(21):2214-2236. [doi: 10.1001/jama.2016.17324] [Medline: 27923088]
4. Santabárbara J, Olaya B, Bueno-Notivol J, et al. Prevalence of depression among medical students during the COVID-19 pandemic: a systematic review and meta-analysis. Rev Med Chil 2021 Nov;149(11):1579-1588. [doi: 10.4067/S0034-98872021001101579] [Medline: 35735320]
5. Cohen AM, Braun K, Hübner N, Scherner PV, Jurkat HB. Influencing factors on stress management in medical students—with special consideration of depression [In German]. Nervenarzt 2022 May;93(5):468-475. [doi: 10.1007/s00115-021-01183-0] [Medline: 34487197]
6. de Sá e Camargo ML, Torres RV, Cotta KCG, da Silva Ezequiel O, Lucchetti G, Lucchetti ALG. Mental health throughout the medical career: a comparison of depression, anxiety, and stress levels among medical students, residents, and physicians. Int J Soc Psychiatry 2023 Aug;69(5):1260-1267. [doi: 10.1177/00207640231157258]
7. Voltmer E, Köslich-Strumann S, Voltmer JB, Kötter T. Stress and behavior patterns throughout medical education—a six year longitudinal study. BMC Med Educ 2021 Aug 28;21(1):454. [doi: 10.1186/s12909-021-02862-x] [Medline: 34454487]
8. Melnyk BM, Kelly SA, Stephens J, et al. Interventions to improve mental health, well-being, physical health, and lifestyle behaviors in physicians and nurses: a systematic review. Am J Health Promot 2020 Nov;34(8):929-941. [doi: 10.1177/0890117120920451] [Medline: 32338522]
9. Wege N, Muth T, Li J, Angerer P. Mental health among currently enrolled medical students in Germany. Pub Health (Fairfax) 2016 Mar;132:92-100. [doi: 10.1016/j.puhe.2015.12.014] [Medline: 26880490]
10. Pelzer A, Sapalidis A, Rabkow N, Pukas L, Günther N, Watzke S. Does medical school cause depression or do medical students already begin their studies depressed? A longitudinal study over the first semester about depression and influencing factors. GMS J Med Educ 2022;39(5):Doc58. [doi: 10.3205/zma001579] [Medline: 36540560]
11. Krümmel A, Laiker I, Wrona KJ, Aschentrup L, Dockweiler C. Acceptance and use of digital interventions for distress prevention among students: results from a qualitative interview study using the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) [In German]. Präv Gesundhförder 2023;18(4):508-516. [doi: 10.1007/s11553-022-00985-7]
12. Dederichs M, Weber J, Pischke CR, Angerer P, Apolinário-Hagen J. Exploring medical students' views on digital mental health interventions: a qualitative study. Internet Interv 2021 Sep;25:100398. [doi: 10.1016/j.invent.2021.100398] [Medline: 34026567]
13. D'Adamo L, Paraboschi L, Grammer AC, et al. Reach and uptake of digital mental health interventions based on cognitive-behavioral therapy for college students: a systematic review. J Behav Cogn Ther 2023 Jun;33(2):97-117. [doi: 10.1016/j.jbct.2023.05.002] [Medline: 37724304]

XSL•FO
RenderX

14. Apolinário-Hagen J. Current perspectives on e-mental-health self-help treatments: exploring the “black box” of public views, perceptions, and attitudes toward the digitalization of mental health care. In: Menvielle L, Audrain-Pontevia AF, Menvielle W, editors. The Digitization of Healthcare: New Challenges and Opportunities: Palgrave Macmillan; 2017:205-223. [doi: 10.1057/978-1-349-95173-4_12]
15. Kuhn S, Frankenhauser S, Tolks D. Digital learning and teaching in medical education: already there or still at the beginning? Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2018 Feb;61(2):201-209. [doi: 10.1007/s00103-017-2673-z] [Medline: 29234823]
16. Aulenkamp J, Mikuteit M, Löffler T, Schmidt J. Overview of digital health teaching courses in medical education in Germany in 2020. GMS J Med Educ 2021;38(4):Doc80. [doi: 10.3205/zma001476] [Medline: 34056069]
17. Smith KA, Blease C, Faurholt-Jepsen M, et al. Digital mental health: challenges and next steps. BMJ Ment Health 2023 Feb;26(1):e300670. [doi: 10.1136/bmjment-2023-300670]
18. Gesundheitsreport 2023 – wie geht’s Deutschlands Studierenden? Techniker Krankenkasse. 2023. URL: https://www.tk.de/resource/blob/2149886/e5bb2564c786aedb3979588fe64a8f39/2023-tk-gesundheitsreport-data.pdf [accessed 2025-01-08]
19. Paganin G, Apolinário-Hagen J, Simbula S. Introducing mobile apps to promote the well-being of German and Italian university students: a cross-national application of the technology acceptance model. Curr Psychol 2022 Oct 27:1-12. [doi: 10.1007/s12144-022-03856-8] [Medline: 36320558]
20. Harrer M, Apolinário-Hagen J, Fritsche L, et al. Effect of an internet- and app-based stress intervention compared to online psychoeducation in university students with depressive symptoms: results of a randomized controlled trial. Internet Interv 2021 Apr;24:100374. [doi: 10.1016/j.invent.2021.100374] [Medline: 33718001]
21. Apolinário-Hagen J, Hennemann S, Fritsche L, Drüge M, Breil B. Determinant factors of public acceptance of stress management apps: survey study. JMIR Ment Health 2019 Nov 7;6(11):e15373. [doi: 10.2196/15373] [Medline: 31697243]
22. Apolinário-Hagen J, Hennemann S, Kück C, et al. Exploring user-related drivers of the early acceptance of certified digital stress prevention programs in Germany. Health Serv Insights 2020;13:1178632920911061. [doi: 10.1177/1178632920911061] [Medline: 32206013]
23. Philippi P, Baumeister H, Apolinário-Hagen J, et al. Acceptance towards digital health interventions: model validation and further development of the Unified Theory of Acceptance and Use of Technology. Internet Interv 2021 Dec;26:100459. [doi: 10.1016/j.invent.2021.100459] [Medline: 34603973]
24. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
25. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
26. Dwivedi YK, Rana NP, Tamilmani K, Raman R. A meta-analysis based modified Unified Theory of Acceptance and Use of Technology (meta-UTAUT): a review of emerging literature. Curr Opin Psychol 2020 Dec;36:13-18. [doi: 10.1016/j.copsyc.2020.03.008] [Medline: 32339928]
27. Oti O, Pitt I. Online mental health interventions designed for students in higher education: a user-centered perspective. Internet Interv 2021 Dec;26:100468. [doi: 10.1016/j.invent.2021.100468] [Medline: 34703772]
28. Dederichs M, Nitsch FJ, Apolinário-Hagen J. Piloting an innovative concept of e-mental health and mHealth workshops with medical students using a participatory co-design approach and app prototyping: case study. JMIR Med Educ 2022 Jan 10;8(1):e32017. [doi: 10.2196/32017] [Medline: 35006085]
29. Chew-Graham CA, Rogers A, Yassin N. “I wouldn’t want it on my CV or their records”: medical students’ experiences of help-seeking for mental health problems. Med Educ 2003 Oct;37(10):873-880. [doi: 10.1046/j.1365-2923.2003.01627.x] [Medline: 12974841]
30. Berliant M, Rahman N, Mattice C, Bhatt C, Haykal KA. Barriers faced by medical students in seeking mental healthcare: a scoping review. MedEdPublish 2016;12:70. [doi: 10.12688/mep.19115.1]
31. Michaeli D, Keough G, Perez-Dominguez F, et al. Medical education and mental health during COVID-19: a survey across 9 countries. Int J Med Educ 2022 Feb 26;13:35-46. [doi: 10.5116/ijme.6209.10d6] [Medline: 35226614]
32. Hennemann S, Beutel ME, Zwerenz R. Drivers and barriers to acceptance of web-based aftercare of patients in inpatient routine care: a cross-sectional survey. J Med Internet Res 2016 Dec 23;18(12):e337. [doi: 10.2196/jmir.6003] [Medline: 28011445]
33. Damerau M, Teufel M, Musche V, et al. Determining acceptance of e-mental health interventions in digital psychodiabetology using a quantitative web-based survey: cross-sectional study. JMIR Form Res 2021 Jul 30;5(7):e27436. [doi: 10.2196/27436] [Medline: 34328429]
34. Bäuerle A, Frewer AL, Rentrop V, et al. Determinants of acceptance of weight management applications in overweight and obese individuals: using an extended Unified Theory of Acceptance and Use of Technology model. Nutrients 2022 May 8;14(9):1968. [doi: 10.3390/nu14091968] [Medline: 35565935]
35. Bäuerle A, Mallien C, Rassaf T, et al. Determining the acceptance of digital cardiac rehabilitation and its influencing factors among patients affected by cardiac diseases. J Cardiovasc Dev Dis 2023 Apr 17;10(4):174. [doi: 10.3390/jcdd10040174] [Medline: 37103053]

XSL•FO
RenderX

36. Mayer G, Gronewold N, Alvarez S, Bruns B, Hilbel T, Schultz JH. Acceptance and expectations of medical experts, students, and patients toward electronic mental health apps: cross-sectional quantitative and qualitative survey study. JMIR Ment Health 2019 Nov 25;6(11):e14018. [doi: 10.2196/14018] [Medline: 31763990]
37. Ungar P, Schindler AK, Polujanski S, Rotthoff T. Online programs to strengthen the mental health of medical students: a systematic review of the literature. Med Educ Online 2022 Dec;27(1):2082909. [doi: 10.1080/10872981.2022.2082909] [Medline: 35642839]
38. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
39. Kroenke K, Spitzer RL, Williams JBW, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 6;146(5):317-325. [doi: 10.7326/0003-4819-146-5-200703060-00004] [Medline: 17339617]
40. Plummer F, Manea L, Trepel D, McMillan D. Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. Gen Hosp Psychiatry 2016;39:24-31. [doi: 10.1016/j.genhosppsych.2015.11.005] [Medline: 26719105]
41. Stoppok P, Teufel M, Jahre L, et al. Determining the influencing factors on acceptance of eHealth pain management interventions among patients with chronic pain using the Unified Theory of Acceptance and Use of Technology: cross-sectional study. JMIR Form Res 2022 Aug 17;6(8):e37682. [doi: 10.2196/37682] [Medline: 35976199]
42. Rentrop V, Damerau M, Schweda A, et al. Predicting acceptance of e–mental health interventions in patients with obesity by using an extended unified theory of acceptance model: cross-sectional study. JMIR Form Res 2022 Mar 17;6(3):e31229. [doi: 10.2196/31229] [Medline: 35297769]
43. Pfeifer A. Datenanalyse Mit SPSS Für Windows [Book in German]: De Gruyter; 1996.
44. Gollwitzer M, Eid M, Schmitt M. Statistik und Forschungsmethoden [Book in German]: Beltz Verlagsgruppe; 2017.
45. Sedlmeier P, Burkhardt M. Datenanalyse mit R: Beschreiben, Explorieren, Schätzen und Testen [Book in German]: Pearson Deutschland; 2021.
46. Wollschläger D. Grundlagen der Datenanalyse mit R : Eine anwendungsorientierte Einführung. In: Ruhmann I, editor. Eine Anwendungsorientierte Einführung: Springer Spektrum; 2020. [doi: 10.1007/978-3-662-53670-4]
47. Esber A, Teufel M, Jahre L, In der Schmitten J, Skoda EM, Bäuerle A. Predictors of patients' acceptance of video consultation in general practice during the coronavirus disease 2019 pandemic applying the Unified Theory of Acceptance and Use of Technology model. D Health 2023;9:20552076221149317. [doi: 10.1177/20552076221149317] [Medline: 36815005]
48. Cohen J. Statistical Power Analysis for the Behavioral Sciences: Routledge; 1977. [doi: 10.4324/9780203771587]
49. Crisp DA, Griffiths KM. Participating in online mental health interventions: who is most likely to sign up and why? Depress Res Treat 2014;2014:790457. [doi: 10.1155/2014/790457] [Medline: 24804089]
50. Nikolic A, Bukurov B, Kocic I, et al. Smartphone addiction, sleep quality, depression, anxiety, and stress among medical students. Front Public Health 2023;11:1252371. [doi: 10.3389/fpubh.2023.1252371] [Medline: 37744504]
51. Lin J, Faust B, Ebert DD, Krämer L, Baumeister H. A web-based acceptance-facilitating intervention for identifying patients' acceptance, uptake, and adherence of internet- and mobile-based pain interventions: randomized controlled trial. J Med Internet Res 2018 Aug 21;20(8):e244. [doi: 10.2196/jmir.9925] [Medline: 30131313]
52. Wimsatt LA, Schwenk TL, Sen A. Predictors of depression stigma in medical students: potential targets for prevention and education. Am J Prev Med 2015 Nov;49(5):703-714. [doi: 10.1016/j.amepre.2015.03.021] [Medline: 26141915]
53. Apolinário-Hagen J, Fritsche L, Bierhals C, Salewski C. Improving attitudes toward e-mental health services in the general population via psychoeducational information material: a randomized controlled trial. Internet Interv 2018 Jun;12:141-149. [doi: 10.1016/j.invent.2017.12.002] [Medline: 30135778]
54. Kirschner B, Goetzl M, Curtin L. Mental health stigma among college students: test of an interactive online intervention. J Am Coll Health 2022;70(6):1831-1838. [doi: 10.1080/07448481.2020.1826492] [Medline: 33048656]
55. Nazari A, Garmaroudi G, Foroushani AR, Hosseinnia M. The effect of web-based educational interventions on mental health literacy, stigma and help-seeking intentions/attitudes in young people: systematic review and meta-analysis. BMC Psychiatry 2023 Sep 4;23(1):647. [doi: 10.1186/s12888-023-05143-7] [Medline: 37667229]
56. Shao R, He P, Ling B, et al. Prevalence of depression and anxiety and correlations between depression, anxiety, family functioning, social support and coping styles among Chinese medical students. BMC Psychol 2020 Apr 22;8(1):38. [doi: 10.1186/s40359-020-00402-8] [Medline: 32321593]
57. Joiner R, Gavin J, Brosnan M, et al. Comparing first and second generation digital natives' internet use, internet anxiety, and internet identification. Cyberpsychol Behav Soc Netw 2013 Jul;16(7):549-552. [doi: 10.1089/cyber.2012.0526] [Medline: 23675995]
58. Burzyńska J, Bartosiewicz A, Januszewicz P. Dr. Google: Physicians—the web—patients triangle: Digital skills and attitudes towards e-health solutions among physicians in south eastern Poland—a cross-sectional study in a pre-COVID-19 era. Int J Environ Res Public Health 2023 Jan 5;20(2):978. [doi: 10.3390/ijerph20020978] [Medline: 36673740]
59. Bautista J, Schueller SM. Understanding the adoption and use of digital mental health apps among college students: secondary analysis of a national survey. JMIR Ment Health 2023 Mar 22;10:e43942. [doi: 10.2196/43942] [Medline: 36947115]

XSL•FO
RenderX

60. Apolinario-Hagen J, Harrer M, Salewski C, Lehr D, Ebert DD. Acceptance and use of e-mental health services among university students: secondary analysis of an experiment. Prav Und Gesundhford 2023;18(2):196-203. [doi: 10.1007/s11553-022-00945-1]
61. Irish M, Kuso S, Simek M, et al. Online prevention programmes for university students: stakeholder perspectives from six European countries. Eur J Public Health 2021 Jul 7;31(31 Suppl 1):i64-i70. [doi: 10.1093/eurpub/ckab040] [Medline: 34240152]

Abbreviations

BI: behavioral intention
EE: effort expectancy
FC: facilitating conditions
GAD-2: Generalized Anxiety Disorder-2
PE: performance expectancy
PHQ-2: Patient Health Questionnaire-2
PHQ-4: Patient Health Questionnaire-4
SI: social influence
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by B Lesselroth; submitted 08.03.24; peer-reviewed by C Papan, K Koelkebeck, M Marendic; revised version received 13.09.24; accepted 24.09.24; published 24.01.25.

Please cite as:
Grüneberg C, Bäuerle A, Karunakaran S, Darici D, Dörrie N, Teufel M, Benson S, Robitzsch A
Medical Students' Acceptance of Tailored e–Mental Health Apps to Foster Their Mental Health: Cross-Sectional Study
JMIR Med Educ 2025;11:e58183
URL: https://mededu.jmir.org/2025/1/e58183
doi:10.2196/58183

© Catharina Grüneberg, Alexander Bäuerle, Sophia Karunakaran, Dogus Darici, Nora Dörrie, Martin Teufel, Sven Benson, Anita Robitzsch. Originally published in JMIR Medical Education (https://mededu.jmir.org), 24.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Integration of an Audiovisual Learning Resource in a Podiatric Medical Infectious Disease Course: Multiple Cohort Pilot Study

Garrik Hoyt[1*], BTech; Chandra Shekhar Bakshi[2], PhD; Paramita Basu[1,3*], PhD

[1]Touro University, New York, NY, United States

[2]New York Medical College, New York, NY, United States

[3]New York College of Podiatric Medicine, 53 E 124th St, New York, NY, United States

[*]these authors contributed equally

Corresponding Author:
Paramita Basu, PhD
Touro University, New York, NY, United States

Abstract

Background: Improved long-term learning retention leads to higher exam scores and overall course grades, which is crucial for success in preclinical coursework in any podiatric medicine curriculum. Audiovisual mnemonics, in conjunction with text-based materials and an interactive user interface, have been shown to increase memory retention and higher order thinking.

Objective: This pilot study aims to evaluate the effectiveness of integrating web-based multimedia learning resources for improving student engagement and increasing learning retention.

Methods: A quasi-experimental study was conducted with 2 cohorts totaling 158 second-year podiatric medical students. The treatment group had access to Picmonic's audiovisual resources, while the control group followed traditional instruction methods. Exam scores, final course grades, and user interactions with Picmonic were analyzed. Logistic regression and correlation analyses were conducted to examine the relationships between Picmonic access, performance outcomes, and student engagement.

Results: The treatment group (n=91) had significantly higher average exam scores ($P<.001$) and final course grades ($P<.001$) than the control group (n=67). Effect size for the average final grades ($d=0.96$) indicated the practical significance of these differences. Logistic regression analysis revealed a positive association between Picmonic access with an odds ratio of 2.72 with a 95% confidence interval, indicating that it is positively associated with the likelihood of achieving high final grades. Correlation analysis revealed a positive relationship ($r=0.25$, $P=.02$) between the number of in-video questions answered and students' final grades. Survey responses reflected increased student engagement, comprehension, and higher user satisfaction (3.71 out of 5 average rating) with the multimedia-based resources compared to traditional instructional resources.

Conclusions: This pilot study underscores the positive impact of animation-supported web-based instruction on preclinical medical education. The treatment group, equipped with Picmonic, exhibited improved learning outcomes, enhanced engagement, and high satisfaction. These results contribute to the discourse on innovative educational methods and highlight the potential of multimedia-based learning resources to enrich medical curricula. Despite certain limitations, this research suggests that animation-supported audiovisual instruction offers a valuable avenue for enhancing student learning experiences in medical education.

(JMIR Med Educ 2025;11:e55206) doi:10.2196/55206

KEYWORDS

learning retention; preclinical education; podiatric medical education; audiovisual learning resources; multimedia-based learning resource; animation-supported learning tools; mnemonics; spaced repetition

Introduction

Long-term retention is crucial for higher exam scores and overall course grades in preclinical coursework. A recent examination of popular board preparatory resources has provided insight into the different trends that students have experienced in their self-directed learning [1]. Incorporating digital resources appeared to be as effective, if not more so, than regular text-based learning [2]. With greater interest shown in digital learning, curiosity has arisen regarding students' sentimental value of animation-styled instruction. One study has shown students' fondness for learning increased with animation instruction as an exciting new way to learn, further increasing permanent learning [3].

Various methods of increasing learning efficiency in medical education have been explored, such as digital recordings, visual mnemonics, and flashcard systems. Many students find the aforementioned methods to be an excellent supplement to the

XSL•FO
RenderX

usual textbook-based learning, resulting in higher test scores, particularly within the medical field [4-7]. Enhanced use of digital teaching tools is effective in providing students with basic science information and has shown to be useful in improving their preparation for clerkship [8].

Mnemonics are a commonly used memory technique in medical school. A mnemonic links to well-known knowledge, sometimes invoking humor or emotions [9]. Web-based learning positively impacts information retention and learning efficiency [10,11]. Audiovisual (AV) mnemonics, in conjunction with text-based materials and an interactive user interface, have been shown to increase memory retention and higher order thinking [12,13].

Picmonic [14], a web-based AV learning resource, uses immersive videos, clinical case questions, flashcards, and high-yield notes, as well as picture mnemonics to cover various aspects of the preclinical and clinical practice curriculum [15-17]. A study by Yang et al [12] observed improved student performance in free-recall and paired-matching tests when using Picmonic. Another study by Abdalla et al [13] underscores how important memory and knowledge retention are to a medical student's grades. Adding AV modalities increases a student's ability to remember information over an extended period of time. Students who had undergone AV sessions had higher marks on response answer questions, shorter time spent answering questions, and a higher memory consolidation after specific time benchmarks. Further studies of using mnemonics, particularly food eponyms in pathology-related education, have shown relevance in learning and retaining pathology knowledge in addition to being useful for United States Medical Licensing Examination boards preparation, clinical clerkship preparation, and future practice [18].

Other studies have examined the usefulness of incorporating AV instructional tools in various levels of education [19-21], including medical education [1,12,22,23], and found them helpful for improving student engagement and learning experiences [12,22,23]. However, there seems to be a dearth of studies exploring the usage of multimedia web-based learning resources in podiatric medical education. Our goal is to evaluate how integrating a multimedia web-based learning resource affects student engagement and learning retention in a preclinical course. Though there are many online learning resources available for medical students to bolster their learning, we selected the commercially available web-based platform Picmonic due to the shorter length of the videos. Since this resource was integrated into the course in the form of low-stakes assignments with the purpose of serving as a supplementary resource, in addition to the textbook and the instructor-provided materials, it was important to ensure that the videos did not take up too much time.

This pilot study aims to offer insight into the use of tools like Picmonic that uses AV media and mnemonics to supplement traditional learning resources in podiatric medical education. To achieve this goal, we have tried to determine in second-year students in a podiatric medicine program (P) if students who have access to Picmonic, an interactive video-based learning system with mnemonics as an additional supplementary resource (I), show higher course performance and experience better learning retention and engagement with the learning material (O), compared to those with access to textbooks and other instructor-provided course materials only (C), when enrolled in a preclinical infectious disease course in their third semester (T). Course performance and knowledge retention were determined by comparing average final exam scores between a treatment group that had access to Picmonic in addition to textbooks and other instructor-provided material and a control group that relied only on the same textbooks and instructor provided materials. Students' perception of the usefulness of this platform as a learning resource and engagement with course materials were assessed using a survey instrument and analysis of correlation between the number of in-video questions answered, the number of times the video was watched, and the accuracy of video-embedded quiz attempts.

Methods

Study Design

A sample of 158 second-year podiatric medical students enrolled in the Infectious Disease course in the third semester at New York College of Podiatric Medicine (NYCPM) were observed in 2 consecutive cohorts. The cohorts consisted of a control group of 67 students taking the course in 2021 and a treatment group of 91 students taking the course in 2022. Participants in the treatment group used the multimedia web-based learning tool Picmonic as a learning resource, while participants in the control group did not. All students were given the same didactic instruction, textbooks, and other traditional learning resources. The study was conducted as a posttest-only, nonequivalent group, quasi-experimental design [24,25]. Although sample selection was nonrandom, it is assumed that the 2 sample groups are similar in their baseline characteristics as they were both in the same curricular level within the program at the time of taking the course. In addition, the initial knowledge and skill level of the students in the 2 cohorts were determined to be equivalent based on their average cumulative grade point average (GPA) data from the earlier semesters in the program and the average incoming Medical College Admission Test (MCAT) scores and undergraduate GPA. Both cohorts started the course and third semester with similar average standardized test scores, similar mean incoming cumulative GPAs, and were given similar course content and assessments. Other confounders like educator quality and digitalization were addressed by using the same instructors and learning management system for the delivery of course content to both cohorts. There were also no changes made in course instruction, course content, syllabus, grading, or objectives between the 2 cohorts. It was also ensured that contextual confounders such as new academic initiatives or changes in course leadership, program objectives, and fallout from the pandemic did not occur during the period of the study.

In the treatment group, students were given 5% participation credit, which was awarded on the completion of the video-based assignments. A customized playlist of assigned videos (aligned to the lecture topics) curated from the Picmonic video database was created by the instructors to be watched by the students on their own time and answer the embedded quiz questions shown in Multimedia Appendix 1. Each set of assigned videos and

XSL•FO
RenderX

quizzes had to be completed before the scheduled in-class lecture on that topic to get credit. Data on the number and frequency of videos watched, multiple attempts at answering video-embedded questions, and quiz accuracy was recorded and monitored using the instructor's dashboard provided to faculty in the Picmonic platform.

Similarly, in the control group, a 5% participation grade was awarded for active participation in the live discussions held during class time based on prior review of the posted instructional materials and assigned readings from the course textbook to be completed before lecture sessions. The contribution of all other course assessments was weighed identically in both control and treatment cohorts. The instructors, textbooks, lectures, instructor-provided materials, and exams used were kept the same between the 2 cohorts.

Exam scores for the treatment group were collected as posttest observations over the course of the semester. The control group underwent comparable nontreated observations [15,16]. Feedback about user experience was gathered from students enrolled in the treatment group at the end of the course through an electronic web-based semi-structured survey questionnaire modified from Haleem et al [17] consisting of 7 required questions included in the survey instrument as shown in Multimedia Appendix 2. NYCPM's Institutional Review Board (IRB) granted ethical approval for this study. The students enrolled in the treatment group were sent the survey link, which the student participants voluntarily filled out. The responses to the 7 questions listed under 4 items in the survey instrument were collected, then analyzed and reported as detailed in the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) adapted from [26] and included as Multimedia Appendix 3. Standard strategies appropriate for this type of quasi-experimental study design, consisting of nonrandomized sampling and posttest-only nonequivalent groups, were used for analysis of the data collected [24,25,27].

Data Analysis

Statistical analysis, data processing, and model fitting were performed in Jupyter Notebook, MATLAB, and Excel. Descriptive statistics, including mean, standard deviation, minimum, maximum, and quartile ranges, were calculated for both the treatment and control group's exam scores and final grade. Boxplots and histograms display score distributions, central tendency, spread, and outliers. Statistical analysis encompassed Levene tests, Welch *t* tests, and the calculation of Cohen *d* as the effect size [28].

Logistic regression analysis enables us to explore the association between access to Picmonic and receiving a score of 90% or higher through a logistic regression model built using Python's Statsmodels library for statistical analysis [29]. To fit the model, we used access to Picmonic as a binary predictor (1=access, 0=no access) to predict whether a student achieved a final grade of 90% or higher in the course (1=final grade ≥0.9; 0=final grade <0.9). We obtained the odds ratio from the model by exponentiating the coefficient for the independent variable with the base of the natural logarithm [30]. The log-likelihood ratio was used to evaluate if access to the resource is a relevant predictor of high final grades.

Correlation analysis was performed to determine the strength of the relationship between usage metrics—number of questions answered, videos played, and quiz accuracy —and students' final grades [30]. We calculated Pearson *r* using a dataset of user interactions with the assigned videos and embedded quizzes on the platform [31].

Survey analysis was conducted using user experience data gathered from students enrolled in the treatment group at the end of the course through an electronic web-based questionnaire sent out by email. Students first answered 4 questions about Picmonic, focusing on information retention, concept understanding, higher test scores, and its usefulness as a learning supplement. Next, students were asked to answer 3 questions regarding their level of satisfaction, frequency of use, and favorite features of the platform—as shown in the Student Experience Survey Instrument in Multimedia Appendix 1. Researchers manually categorized answers to questions regarding their favorite features in Excel. Accordingly, summary statistics were calculated and compared using the data collected from student survey responses.

Ethical Considerations

This study was approved by NYCPM's IRB (23575) in May 2022. Informed consent was waived off by the IRB since students agree to the use of unidentifiable education data for research purposes at registration.

As per institutional policy, the IRB approval for this study is a blanket approval provided for all curriculum-related studies, which are undertaken at the college using deidentified aggregate course data rather than individual scores. The original consent or blanket IRB approval covers secondary analysis without additional consent since all incoming new students are required to sign a consent form agreeing to the use of unidentifiable course and education-related data for research purposes at the beginning and is applicable throughout their enrollment in the program.

All students enrolled in the courses that were included in this study were informed about the research and were made aware that the deidentified aggregate course performance data and their feedback would be used to gather data for this pilot study. This information was also reiterated when they were given the survey instruments to record their feedback which was optional for them to fill out.

All data used in this study are course-level aggregate data calculated from score-related data that are anonymized or deidentified.

No compensation was provided for participation in the research study as the courses used are required as part of the podiatric medical curriculum. The students were made aware their feedback would be collected in the form of responses to a survey questionnaire which was optional to complete. Transparency and fairness were ensured by clarifying that the survey instrument was not mandatory and without any consequences for participants who opted out of responding to the questions included in the survey.

XSL•FO
RenderX

Results

Statistical Analysis

The difference in distribution of exam scores and final grades among the treatment and control groups was visualized using bar graphs (Figure 1A), and box plots (Figure 1B). The summary statistics (Table 1) show the central tendencies using mean exam scores and mean final grade, the spread of the scores using standard deviation, and the shape of the score distributions within each group (Figure 1B), which were used to identify potential differences between the groups. The treatment group had significantly higher average exam scores for most of the course exams and had higher final grades compared to the control group (Figure 1A). The difference in the average scores of the first 2 exams (P<.001), the third exam (P=.04), and the final course grades (grand total) (P<.001) between the 2 groups with and without access to Picmonic was significant. There were also significant differences in variance for exam 1, exam 2, exam 3, and the final grade between the treatment and control groups (Table 2).

Figure 1. Comparison of student performance in course assessment between treatment group (T) and control group (C) based on (**a**) scaled mean test scores and final course grades from treatment group (having access to Picmonic or AV instruction) and control group (without access to Picmonic or audiovisual instruction) and (**b**) score distributions within each group.

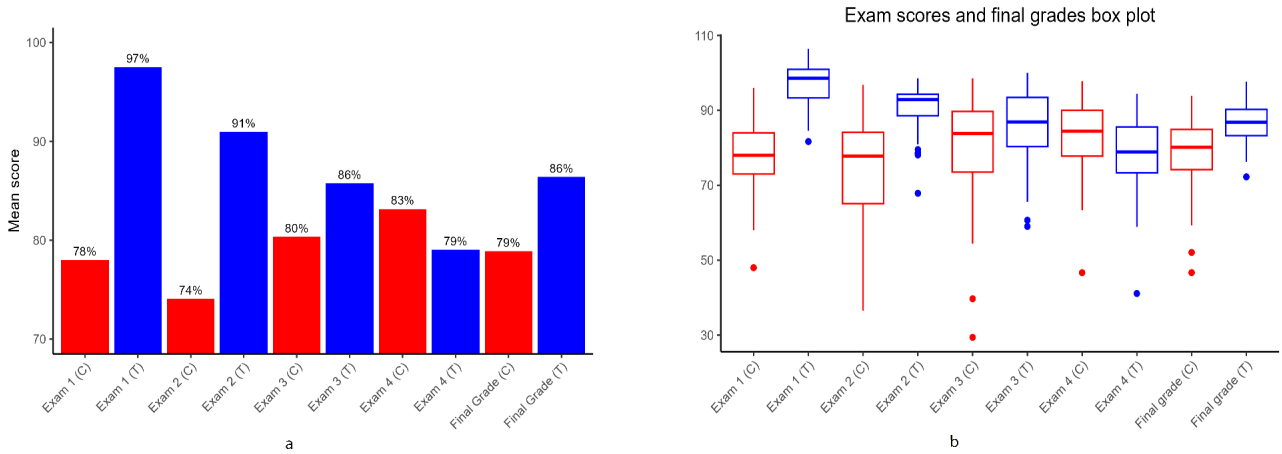


Table . Statistical analysis of student performance in course assessments for treatment and control groups.

Exam	Control, mean (SD)	Treatment, mean (SD)	Mean difference	Welch t test P value	Pooled SD	Cohen d
Exam 1	0.780 (0.098)	0.975 (0.052)	0.195	<.001	0.075	2.398
Exam 2	0.741 (0.143)	0.910 (0.052)	0.169	<.001	0.101	1.486
Exam 3	0.803 (0.133)	0.858 (0.089)	0.054	.005	0.110	0.463
Exam 4	0.832 (0.091)	0.790 (0.092)	−0.041	.006	0.092	−0.449
Final grade	0.789 (0.094)	0.864 (0.049)	0.075	<.001	0.072	0.957

Table . Analysis of variance of exam scores and final grades between treatment and control groups.

Assessment	Levene test statistic	P value	
Exam 1	19.442	<.001[a]	Significant compared to control
Exam 2	52.632	<.001[b]	Significant
Exam 3	4.354	.04	Significant
Exam 4	0.127	.72	Not significant
Final grade	14.652	<.001[c]	Significant

[a](1.92×10^{-5})
[b](1.77×10^{-11})
[c](1.87×10^{-4})

Levene test statistic values for potential differences in variance of exam scores and final grades between the 2 groups revealed significant differences in variance for exam 1, exam 2, exam 3, and the final grade (Table 2). The results of Welch t test used to compare the average final exam scores and final grades indicated statistically significant differences between the 2 groups' first 3 exams and final grades for the course (P<.01) (Table 1). Cohen d values calculated to quantify the observed

XSL•FO
RenderX

differences between the treatment and control groups across all exams and the final grade revealed a large effect size for exam 1 (d=2.397), exam 2 (d=1.486), and the final grade (d=0.957) (Table 1).

Logistic Regression

A logistic regression analysis between access to Picmonic and the likelihood of achieving a high final course grade of 90 out of 100 (90%) or above resulted in an odds ratio which indicates that, assuming all other factors are constant, students in our study with access to Picmonic were 2.72 times more likely to have received a final grade of 90% or higher and a letter grade of A in the course. The log-likelihood ratio P value is .02 (P<.05); therefore, we reject the null hypothesis that the base model with only the intercept is better than the model with access to Picmonic used as the predictor (Table 3).

Table . Regression analysis of the association between access to Picmonic and receiving a high final grade.

Predictor	Coefficient	SE	z value	P value[a]	Lower CI	Upper CI
Intercept	−1.998	0.377	−5.303	<.001	−2.737	−1.260
Picmonic access	0.971	0.446	2.180	.03	0.098	1.845

[a]Model log-likelihood ratio P value =.02.

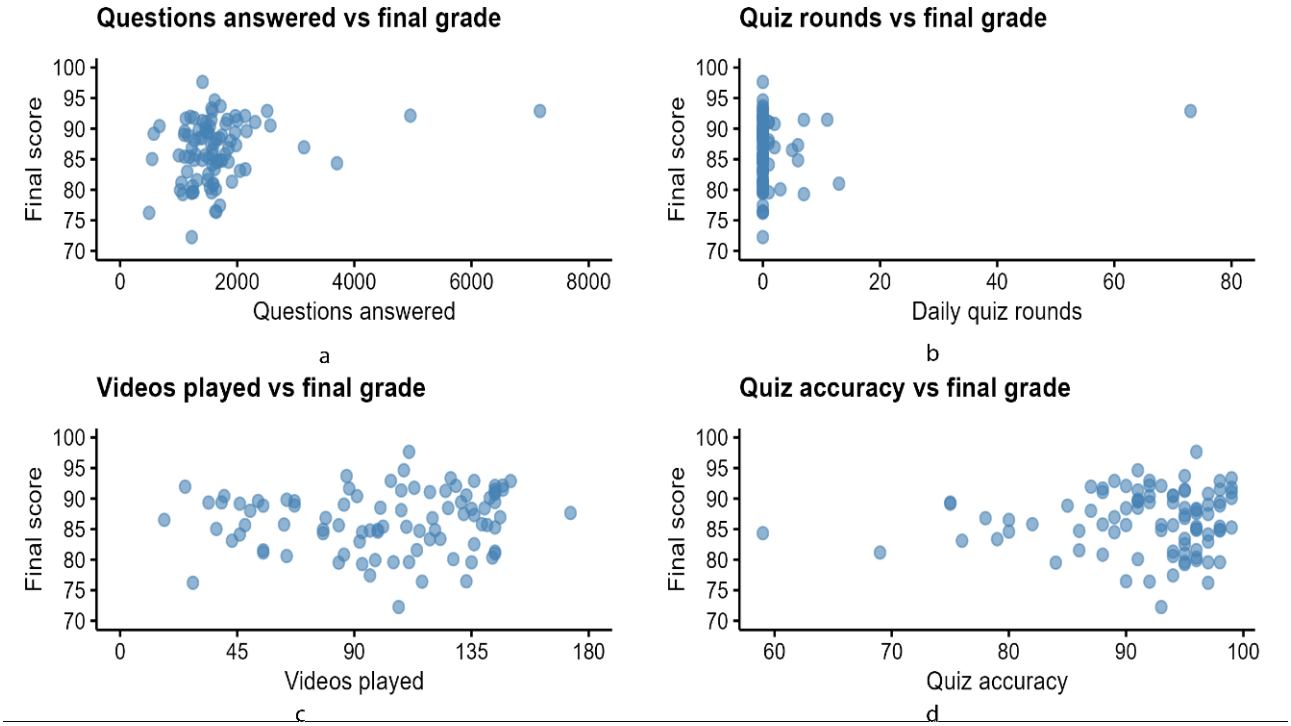
Correlation Analysis

Pearson correlation coefficient calculated to explore the relationship between final course grade and various usage statistics (number of videos played, questions answered, and quiz accuracy) show that students in the treatment group who answered more questions on the platform tended to get a higher final score (r=0.25, P=.02) (Table 4 and Figure 2).

Table . Analysis of correlation between final grades and platform usage metrics.

Final grade with:	Pearson r	P value
Questions answered	0.247965674	.02
Daily quiz rounds	0.124515676	.24
Videos played	0.09980258	.35
Quiz accuracy	0.074970912	.48

Figure 2. Correlation analysis between final course grades and (**a**) the number of in-video questions answered, (**b**) number of daily quiz round attempts, (**c**) number of times videos watched, and (**d**) overall accuracy of embedded quizzes.



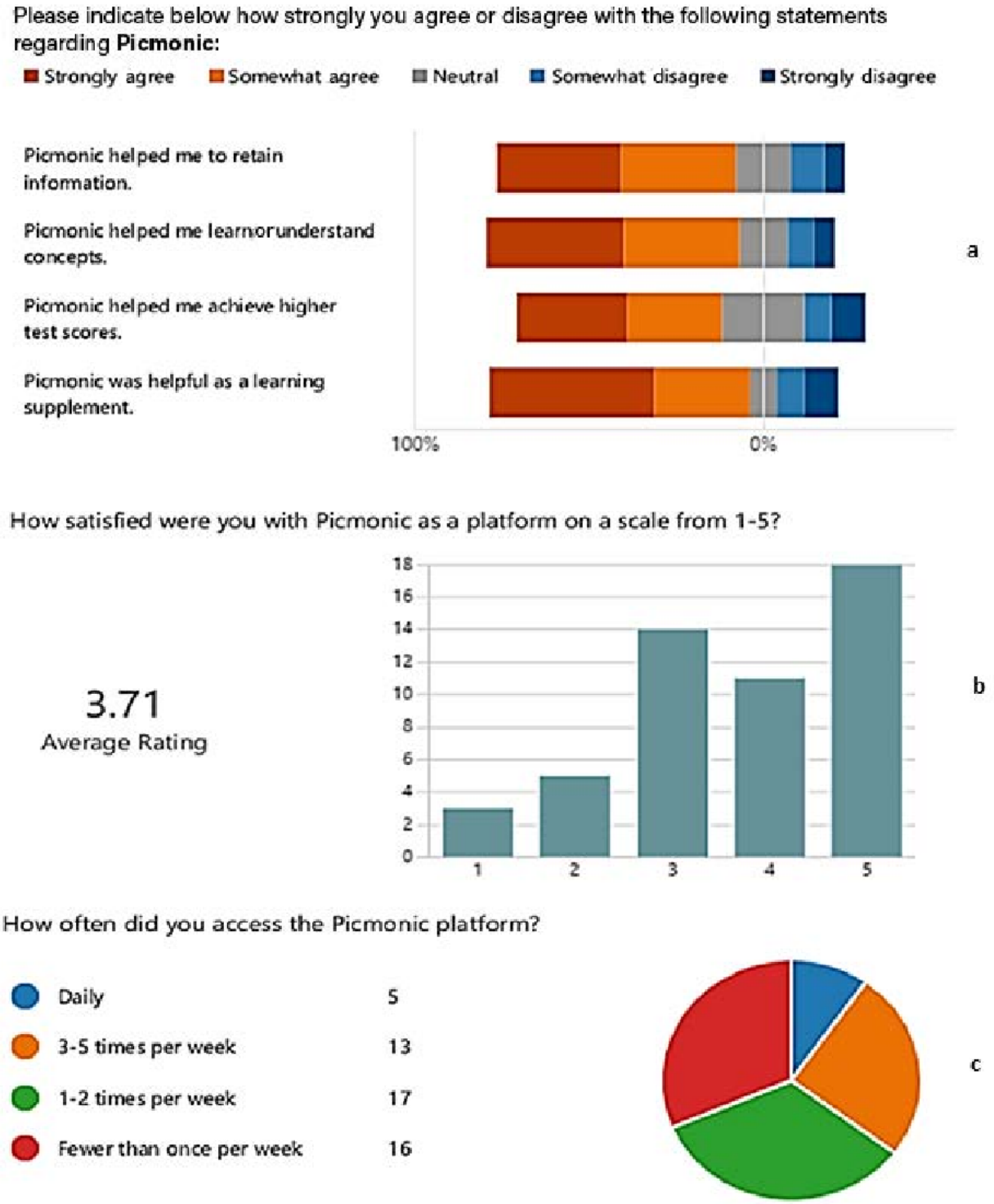
XSL•FO
RenderX

Survey Analysis

The response rate for item 1 of the survey instrument, which consisted of 4 questions about the students' perceived usefulness of Picmonic, was 73% (66 out of 91). In this item, students indicated strong agreement regarding Picmonic's positive impact on information retention, concept understanding, higher test scores, and usefulness as a supplementary learning tool (Figure 3A). In item number 4, students also reported an average satisfaction rating of 3.71 out of 5 (Figure 3B). Out of 53 students who responded to item number 2 in the survey questionnaire, 36 accessed Picmonic at least once a week—predominantly 1‑2 times per week (Figure 3C). In item number 3, 50 out of 91 students responded to the open-ended questions regarding user experience or preferences about their favorite feature of Picmonic with some choosing more than 1 feature. The number of times each feature was reported as preferred is listed and compared in Table 5. Noteworthy features of Picmonic highlighted by students included videos, quizzes and questions, mnemonics, and content (Table 5).

XSL•FO
RenderX

Figure 3. Analysis of student experience feedback data. Student experience survey insights and summary of qualitative open-ended student feedback data showing (**a**) perceived effectiveness of Picmonic on various learning outcomes and retention, (**b**) user satisfaction level, and (**c**) frequency of use.



XSL•FO
RenderX

Table . Comparative analysis of student feedback performed on open-ended questions regarding user experience or preferences.

	Feedback (N=50), n (%)
Videos	15 (30)
Quizzes and questions	9 (18)
Mnemonics	9 (18)
Content	9 (18)
Images	6 (12)
User interface	4 (8)
General	2 (4)

Discussion

Principal Findings

In our study, we observed that students with access to the multimedia and mnemonic-based AV learning resource scored higher on most exams, had higher final grades, and were more likely to receive a final grade of 90% or higher. The resulting effect sizes for the treatment group are large enough to be meaningful in the real world; however, other factors may have contributed to the observed effect size. Students in the treatment group who answered more questions on the platform tended to get a higher final score. Most students used Picmonic at least once a week to learn the course content. The lower response rate on this item asking about the frequency of accessing Picmonic in a week could be due to the inability of participants to remember their frequency of access at the time of completing the survey several weeks after the completion of the course and semester. Students had positive opinions regarding Picmonic's platform user experience and its effectiveness in helping them retain information, learn and understand concepts, and achieve higher test scores. In this course, the students reported the videos to be their most preferred feature, followed by mnemonics and self-assessment quiz questions associated with the videos, which allowed them to test their knowledge both during and after watching the assigned videos on their own time with unlimited attempts. In this context, it is important to note that though students generally prefer to watch videos on their own rather than attending class, the 2 groups were treated equally since both the reading assignments (for the control group) and the Picmonic-based video assignments (for the treatment group) were assigned to be completed outside of class on their own time. Additionally, the students in the treatment group were assigned to watch the videos and complete the video-embedded quizzes within a specific time frame to mimic the reading assignments given to the control group. To receive the 5% participation grade, the treatment group had to complete the Picmonic video assignments within the instructor-provided deadline corresponding to the weekly topical schedule, which is similar to that of the control group, rather than watching the assigned videos at their own pace throughout the semester.

Comparison With Previous Research

Our experimental results indicate that students with access to the multimedia-based AV mnemonic learning resource scored higher on most exams, had higher final grades, and were more likely to receive a final grade of 90% or higher. The difference between the grades achieved by the 2 groups is large enough to be meaningful in the real world. These findings are consistent with previous studies that have demonstrated the effectiveness of AV mnemonics and web-based learning tools in enhancing memory retention and learning outcomes in medical education [12,16].

Yang et al [12] observed improved student performance in free-recall and paired-matching tests when using an earlier version of the multimedia-based learning platform that we have used here that was released almost 10 years ago, while the current version that we have used has newer, redesigned, more impactful, and shorter videos, though still based on the same type of picture mnemonics and principles of spaced repetition and visual learning. The currently available version that was used in this study also has improved dashboard features and assessment capabilities compared to the older version. Abdalla et al [13] also found that students who had undergone AV sessions had higher marks on response answer questions, shorter time spent answering questions, and higher memory consolidation after specific time benchmarks. These studies underscore the importance of memory and knowledge retention in medical students' academic performance.

Results examining user interaction with the resource showed that the more questions a student answered on a multimedia-based AV learning platform (like Picmonic) using spaced repetition and mnemonics, the higher their grade tended to be. This finding aligns with research highlighting the benefits of interactive user interfaces and spaced repetition in increasing memory retention and higher order thinking [12].

Survey responses indicated that students found the resource useful for learning concepts, retaining information, and achieving higher test scores. This is consistent with previous research demonstrating increased student engagement, comprehension, and satisfaction with multimedia-based resources compared to traditional instructional methods [15,18-20]. Studies by Tackett et al [23] examined student engagement with commercially produced medical education videos incorporated into a preclinical course and also found the videos to be helpful for student learning and improved students' experiences.

Overall, our findings contribute to the growing body of evidence supporting the integration of multimedia-based learning resources and AV mnemonics in medical education curricula

XSL•FO
RenderX

to enhance student learning experiences and outcomes [7,8,12,16,17].

Limitations

While this investigation showed promising initial findings, we recognize that the study design limits its ability to make unequivocal causal inferences about the impact of the multimedia tool alone on the outcomes. The sample size is relatively small, and the lack of randomization in the study design may limit the generalizability of the findings. The quasi-experimental design with nonrandom assignment to treatment and control groups restricts the establishment of a cause-and-effect relationship [24], which could potentially affect the internal validity of the study and its ability to accurately infer whether the change in outcomes was caused by the intervention. Without randomized selection of the 2 groups, we cannot rule out potential unmeasured differences due to systematic differences in sample selection [23]. The study only examined the effect of the implementation of the intervention in 1 preclinical infectious disease course, limiting assessment of the tool's effectiveness. Additionally, the 5% participation credit component was implemented differently for the control (class participation) versus treatment (video watching) groups, which could impact effort levels.

The 2 most likely influential unmeasured confounding variables in our study are potential differences in the overall academic aptitude of the students in each sample [13], as well as potential differences in the amount of time spent with study materials between the 2 groups due to the spaced repetition provided in the Picmonic platform, which could impact the results. The potential difference in student aptitude between the 2 groups should be somewhat mitigated by the fact that both groups were 2nd-year medical students at the same college when being tested. Additionally, the 2 groups had similar average standardized test scores and mean incoming GPAs at the beginning of the semester. The absence of pretest observations comparing the treatment group to the control group makes it difficult to know whether any differences between the 2 groups could potentially be attributed to pre-existing factors rather than the multimedia learning tool alone [14].

Since this study was done on a very limited number of students and involved only 1 course focused on the topic of preclinical medical microbiology and infectious diseases, this effect may not be generalized for all topics or types of courses. This course is heavily dependent on memorization and recall due to the nature of the topics covered, which may also contribute to the impact of the integration of visual learning with mnemonics and spaced repetition and therefore may not be equally applicable in another course that does not require extensive memorization.

Conclusions

Our study shows strong agreement amongst students that Picmonic helped them achieve key learning outcomes. Usage data revealed a positive relationship between final grades and students' usage of platform features; the number of in-video questions answered had a stronger correlation than the number of times videos were watched or the accuracy of topic-associated quiz attempts. Students that were given access to Picmonic did better on exams; however, it must be noted that due to the nonrandomized sampling process, posttest-only study design, limited implementation, and small sample size, it is difficult to conclude whether the difference was due solely to the integration of the new tool. The improved course grades and test scores in the treatment group may have been due to the inherent confounding factors like comprehension and retention skills or increased contact time with the course topics provided by the platform features. Our pilot study focused on the integration of Picmonic, a multimedia-based learning resource, in only 1 course, but implementing it across more courses over multiple semesters would strengthen the assessment of the tool's effectiveness. Despite its limitations, this study provides insight into the potential benefits of integrating multimedia learning resources in podiatric medical education. However, a larger study that implements this type of learning resource on a larger scale, and in more preclinical and clinical courses throughout the curriculum, is needed to further analyze its effectiveness in the podiatric medical curriculum.

Acknowledgments

This work originated at New York College of Podiatric Medicine; the authors express their deep gratitude for the support received. The authors of this paper would like to thank Rhonda Altonen, MLS, MS (Director of Library Services, Sheldon L. Sirota Memorial Library, Touro College of Osteopathic Medicine and College of Pharmacy, Touro University) for her insights about the specific features of Picmonic, the multimedia-based learning resource used in this study, as well Adrian Rice, Associate Registrar, New York College of Osteopathic Medicine, for providing summarized cumulative grade point average data of the control and treatment cohorts. The authors of this paper would also like to thank Picmonic for providing additional data on usage of platform features for further analysis which provided detailed insights.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example of the playlist of Picmonic videos and embedded quiz questions on Gram positive bacilli assigned to the treatment group students.

XSL•FO
RenderX

[PNG File, 417 KB - mededu_v11i1e55206_app1.png]

Multimedia Appendix 2
Student user experience survey instrument.
[PNG File, 186 KB - mededu_v11i1e55206_app2.png]

Multimedia Appendix 3
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[PDF File, 104 KB - mededu_v11i1e55206_app3.pdf]

References

1. O'Hanlon R, Laynor G. Responding to a new generation of proprietary study resources in medical education. J Med Libr Assoc 2019 Apr;107(2):251-257. [doi: 10.5195/jmla.2019.619] [Medline: 31019395]
2. Kyaw BM, Posadzki P, Paddock S, Car J, Campbell J, Tudor Car L. Effectiveness of digital education on communication skills among medical students: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Aug 27;21(8):e12967. [doi: 10.2196/12967] [Medline: 31456579]
3. Nur Yasak Cevi˙zci˙ R. An investigation into the student opinions and observations towards animation-supported instruction in social sciences lesson. Int J Turkish Lit Cult Educ 2023 Jan 1;12(1):374-391. [doi: 10.7884/teke.5619]
4. Harris DM, Chiang M. An analysis of Anki usage and strategy of first-year medical students in a structure and function course. Cureus 2022 Mar;14(3):e23530. [doi: 10.7759/cureus.23530] [Medline: 35494926]
5. Wechsler B. Improvement of medical terminology education with interactive learning module. S D Med 2022 Oct;75(10):459. [Medline: 36889270]
6. Brahler CJ, Walker D. Learning scientific and medical terminology with a mnemonic strategy using an illogical association technique. Adv Physiol Educ 2008 Sep;32(3):219-224. [doi: 10.1152/advan.00083.2007]
7. Maag M. The effectiveness of an interactive multimedia learning tool on nursing students' math knowledge and self-efficacy. Comput Inform Nurs 2004;22(1):26-33. [doi: 10.1097/00024665-200401000-00007] [Medline: 15069846]
8. Myers SR. The changing face of medical education. J Med Educ Curric Dev 2014 Jan;1:S17270. [doi: 10.4137/JMECD.S17270]
9. West N. Mnemonics are useful memory tools in modern medicine. Ugeskr Laeger 2014 Dec 8;176(50):V66204. [Medline: 25498182]
10. Noor U, Younas M, Saleh Aldayel H, Menhas R, Qingyu X. Learning behavior, digital platforms for learning and its impact on university student's motivations and knowledge development. Front Psychol 2022;13:933974. [doi: 10.3389/fpsyg.2022.933974] [Medline: 36506979]
11. Abarghouie MHG, Omid A, Ghadami A. Effects of virtual and lecture-based instruction on learning, content retention, and satisfaction from these instruction methods among surgical technology students: a comparative study. J Educ Health Promot 2020;9:296. [doi: 10.4103/jehp.jehp_634_19] [Medline: 33426100]
12. Yang A, Goel H, Bryan M, et al. The Picmonic(®) Learning System: enhancing memory retention of medical sciences, using an audiovisual mnemonic Web-based learning platform. Adv Med Educ Pract 2014;5:125-132. [doi: 10.2147/AMEP.S61875] [Medline: 24868180]
13. Abdalla MMI, Azzani M, Rajendren R, et al. Effect of story-based audiovisual mnemonics in comparison with text-reading method on memory consolidation among medical students: a randomized controlled trial. Am J Med Sci 2021 Dec;362(6):612-618. [doi: 10.1016/j.amjms.2021.07.015] [Medline: 34606752]
14. Picmonic. URL: https://www.picmonic.com/ [accessed 2025-01-24]
15. Bhangu S, Provost F, Caduff C. Introduction to qualitative research methods - part I. Perspect Clin Res 2023;14(1):39-42. [doi: 10.4103/picr.picr_253_22] [Medline: 36909216]
16. Krishnan P. A review of the non-equivalent control group post-test-only design. Nurse Res 2019 Sep 21;26(2):37-40. [doi: 10.7748/nr.2018.e1582] [Medline: 30226337]
17. Haleem A, Javaid M, Qadri MA, Suman R. Understanding the role of digital technologies in education: a review. Sust Oper Comput 2022;3:275-285. [doi: 10.1016/j.susoc.2022.05.004]
18. Seto C, Zayat V. A spoonful of eponyms helps the pathology go down: using food eponyms and visual mnemonics in preclinical pathology education. Med Sci Educ 2022 Feb;32(1):131-140. [doi: 10.1007/s40670-021-01474-w] [Medline: 35154897]
19. Harris DA, Krousgrill C. Distance education: new technologies and new directions. Proc IEEE 2008 May 20;96(6):917-930. [doi: 10.1109/JPROC.2008.921612]
20. Giovannella C. Effect induced by the Covid-19 pandemic on students' perception about technologies and distance learning. Ludic, co-design and tools supporting smart learning ecosystems and smart education. Smart Innov Syst Technol 2020;197:105-116. [doi: 10.1007/978-981-15-7383-5_9]

XSL•FO
RenderX

21. Sosa Neira EA, Salinas J, De Benito B. Emerging technologies (ETs) in education: a systematic review of the literature published between 2006 and 2016. Int J Emerg Technol Learn 2017;12(5):128. [doi: 10.3991/ijet.v12i05.6939]
22. Grainger R, Liu Q, Geertshuis S. Learning technologies: a medium for the transformation of medical education? Med Educ (Chicago Ill) 2021 Jan;55(1):23-29. [doi: 10.1111/medu.14261]
23. Tackett S, Green D, Dyal M, et al. Use of commercially produced medical education videos in a cardiovascular curriculum: multiple cohort study. JMIR Med Educ 2021 Oct 7;7(4):e27441. [doi: 10.2196/27441] [Medline: 34617911]
24. Harris AD, McGregor JC, Perencevich EN, et al. The use and interpretation of quasi-experimental studies in medical informatics. J Am Med Inform Assoc 2006;13(1):16-23. [doi: 10.1197/jamia.M1749] [Medline: 16221933]
25. Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-Experimental Designs for Generalized Causal Inference: Houghton Mifflin; 2002.
26. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004;6(3):e34. [doi: 10.2196/jmir.6.3.e34]
27. Maxwell SE, Delaney HD. Designing Experiments and Analyzing Data: A Model Comparison Approach, 2nd edition: Lawrence Erlbaum Associates; 2004. [doi: 10.4324/9781410609243]
28. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Routledge; 1988.
29. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. Presented at: Python in Science Conference; Jun 28 to Jul 3, 2010; Austin, Texas. [doi: 10.25080/Majora-92bf1922-011]
30. Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression, 3rd edition: Wiley; 2013. [doi: 10.1002/9781118548387]
31. Witte R, Witte J. Statistics, 9th edition: Wiley; 2010.

Abbreviations

AV: audiovisual
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
GPA: grade point average
IRB: Institutional Review Board
MCAT: Medical College Admission Test
NYCPM: New York College of Podiatric Medicine

Edited by B Lesselroth; submitted 05.12.23; peer-reviewed by K Malale, S Mukhida; revised version received 15.10.24; accepted 03.12.24; published 11.02.25.

Please cite as:
Hoyt G, Bakshi CS, Basu P
Integration of an Audiovisual Learning Resource in a Podiatric Medical Infectious Disease Course: Multiple Cohort Pilot Study
JMIR Med Educ 2025;11:e55206
URL: https://mededu.jmir.org/2025/1/e55206
doi:10.2196/55206

© Garrik Hoyt, Chandra Shekhar Bakshi, Paramita Basu. Originally published in JMIR Medical Education (https://mededu.jmir.org), 11.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Performance of ChatGPT-3.5 and ChatGPT-4 in the Taiwan National Pharmacist Licensing Examination: Comparative Evaluation Study

Ying-Mei Wang[1,2,3,4], MBA; Hung-Wei Shen[1,2,4], MBA; Tzeng-Ji Chen[5,6,7], Dr Med; Shu-Chiung Chiang[1,8], PhD; Ting-Guan Lin[2,4], BS

[1]Department of Medical Education and Research, Taipei Veterans General Hospital Hsinchu Branch, 81, Section 1, Zhongfeng Road, Zhudong, Hsinchu, Taiwan

[2]Department of Pharmacy, Taipei Veterans General Hospital Hsinchu Branch, Hsinchu, Taiwan

[3]School of Medicine, National Tsing Hua University, Hsinchu, Taiwan

[4]Hsinchu County Pharmacists Association, Hsinchu, Taiwan

[5]Department of Family Medicine, Taipei Veterans General Hospital Hsinchu Branch, Hsinchu, Taiwan

[6]Department of Family Medicine, Taipei Veterans General Hospital, Taipei, Taiwan

[7]Department of Post-Baccalaureate Medicine, National Chung Hsing University, Taichung, Taiwan

[8]Institute of Hospital and Health Care Administration, School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

Corresponding Author:
Ying-Mei Wang, MBA
Department of Medical Education and Research, Taipei Veterans General Hospital Hsinchu Branch, 81, Section 1, Zhongfeng Road, Zhudong, Hsinchu, Taiwan

Abstract

Background: OpenAI released versions ChatGPT-3.5 and GPT-4 between 2022 and 2023. GPT-3.5 has demonstrated proficiency in various examinations, particularly the United States Medical Licensing Examination. However, GPT-4 has more advanced capabilities.

Objective: This study aims to examine the efficacy of GPT-3.5 and GPT-4 within the Taiwan National Pharmacist Licensing Examination and to ascertain their utility and potential application in clinical pharmacy and education.

Methods: The pharmacist examination in Taiwan consists of 2 stages: basic subjects and clinical subjects. In this study, exam questions were manually fed into the GPT-3.5 and GPT-4 models, and their responses were recorded; graphic-based questions were excluded. This study encompassed three steps: (1) determining the answering accuracy of GPT-3.5 and GPT-4, (2) categorizing question types and observing differences in model performance across these categories, and (3) comparing model performance on calculation and situational questions. Microsoft Excel and R software were used for statistical analyses.

Results: GPT-4 achieved an accuracy rate of 72.9%, overshadowing GPT-3.5, which achieved 59.1% ($P<.001$). In the basic subjects category, GPT-4 significantly outperformed GPT-3.5 (73.4% vs 53.2%; $P<.001$). However, in clinical subjects, only minor differences in accuracy were observed. Specifically, GPT-4 outperformed GPT-3.5 in the calculation and situational questions.

Conclusions: This study demonstrates that GPT-4 outperforms GPT-3.5 in the Taiwan National Pharmacist Licensing Examination, particularly in basic subjects. While GPT-4 shows potential for use in clinical practice and pharmacy education, its limitations warrant caution. Future research should focus on refining prompts, improving model stability, integrating medical databases, and designing questions that better assess student competence and minimize guessing.

(JMIR Med Educ 2025;11:e56850) doi:10.2196/56850

KEYWORDS

artificial intelligence; ChatGPT; chat generative pre-trained transformer; GPT-4; medical education; educational measurement; pharmacy licensure; Taiwan; Taiwan national pharmacist licensing examination; learning model; AI; Chatbot; pharmacist; evaluation and comparison study; pharmacy; statistical analyses; medical databases; medical decision-making; generative AI; machine learning

XSL•FO
RenderX

Introduction

Background

With the advent of the artificial intelligence (AI) era, applications of AI in the medical field have increased with ChatGPT (OpenAI) being the most notable examples. ChatGPT is a large language model based on a generative pretrained transformer developed by OpenAI. ChatGPT-3.5 (GPT-3.5) was the first publicly accessible version, while ChatGPT-4 (GPT-4) was the subscription version. GPT-4 surpasses GPT-3.5 in advanced reasoning, almost nearing human-level performance in professional and academic examinations [1,2]. For instance, GPT-4 ranked in the top 10% of scores on a law examination, whereas GPT-3.5 ranked in the bottom 10% [3]. Additionally, GPT-3.5 resolved 90% of false-belief tasks, achieving the level of a 7-year-old child, whereas GPT-4 resolved 95% of these tasks [4]. Following its launch, ChatGPT has been extensively studied and discussed in both the medical and educational fields [5]. The most widely recognized performance of GPT-3.5 has been on the United States Medical Licensing Examination (USMLE) [6,7]; however, GPT-3.5's performance did not meet expectations in other examinations [8-11]. Gradually, Nori et al [12]observed that the accuracy of GPT-4 was higher than that of the GPT-3.5 on the USMLE, and further studies confirmed that GPT-4 outperforms GPT-3.5 [13-16]. However, there has been limited research on its performance in pharmacy examinations.

In the field of pharmacy, GPT-3.5 has exhibited commendable performance in clinical toxicology and pharmacology [17,18], although it has not passed the National Pharmacist Licensing Examination (NPLE) in Taiwan [19]. However, GPT-4 has outperformed GPT-3.5 in drug information [20] and China's Pharmacist Licensing Examination [21]. Generative AI models, a large language model, has been applied in drug development and novel drug design [22-24], pharmacovigilance [25,26], pharmacokinetic model development [27], pharmacy education, and research writing [28,29].

Goal of the Study

According to previous studies, GPT-3.5 failed to pass the NPLE, indicating its limitations in pharmacy education. Based on these findings, we hypothesized that GPT-4 would outperform GPT-3.5 in this context, demonstrating greater proficiency. To test this hypothesis, this study compared the performance of GPT-3.5 and GPT-4 on Taiwan's NPLE. Additionally, we conducted a comprehensive assessment of their performance across various question types, with a focus on pharmacy-related tasks such as pharmacokinetic calculation and clinical decision-making scenarios. This analysis aims to determine the practical applications of GPT-4 in pharmacy education and establish guidelines for its optimal use in this field.

Methods

Background

The NPLE in Taiwan is divided into 2 stages. The first stage focuses on 3 basic subjects: pharmacology and pharmaceutical chemistry, pharmaceutical analysis and pharmacognosy (including traditional Chinese medicine), and pharmaceutics and biopharmaceutics. The second stage focuses on 3 clinical subjects: dispensing and clinical pharmacy, pharmacotherapy, and pharmacy administration and pharmacy law. The first and second stages of the examination have 240 and 210 multiple-choice questions, respectively. Pharmacy students typically complete the first-stage exam after completing their third year of university coursework. They become eligible for the second-stage exam only after passing the first examination, completing their internships and obtaining their graduation certificates. After passing the second-stage examination, candidates receive their pharmacist certificate, allowing them to practice as a pharmacist legally.

Data Source

This study used the 2-stage NPLE questions released by the Ministry of Examination in February 2023, with each subject exam lasting for 1 hour. The version of NPLE used in this study was the most recent available at the time of research. We used both GPT-3.5 (free version) and GPT-4 (licensed version). No temperature settings were applied. Examination questions were manually fed into GPT-4 and GPT-3.5 sequentially. To simulate student responses, complete questions were entered into the models without tailored prompts. One question was input at a time, and the responses were recorded for analysis. Since GPT-3.5 cannot process images and image functionality of GPT-4 was unavailable during the analysis, only text-based questions were used. Questions containing graphics, such as chemical structures, tables, symbols, and formulas were excluded. Both models were presented with the same set of questions under identical conditions. Due to the limitations on the number of times the model could be used and required cooling time between queries, all questions were answered sequentially and not timed to avoid any potential bias introduced by time constraints.

Study Design

The study was divided into 3 parts; the first part compared the accuracy of GPT-4 and GPT-3.5, as well as in different subjects. The second part compared the accuracy of GPT-4 and GPT-3.5 across different question types. These questions were categorized into 4 types: memory-based questions (1 correct word answer out of 4 options, low-level thinking; Figure 1), judgment questions (1 correct statement out of 4, medium-level thinking; Figure 2), reverse questions (1 incorrect statement out of 4, medium to high-level thinking; Figure 3), and comprehension questions (multiple-choice or matching types, high-level thinking; Figure 4). One pharmacist classified the questions according to these established categories and the second pharmacist reviewed the classifications. In the event of disagreement, a third pharmacist was consulted for the final decision. All pharmacists had over 10 years of experience in medical center hospitals or community teaching hospitals. The third part compared the accuracy of GPT-4 and GPT-3.5 for calculation-based and case scenario questions (Figure 5). Model testing for this study was conducted from May 10 to July 20, 2023.

XSL•FO
RenderX

Figure 1. Template of a memory-based question (choose 1 correct word from 4 options, requiring low-level thinking).

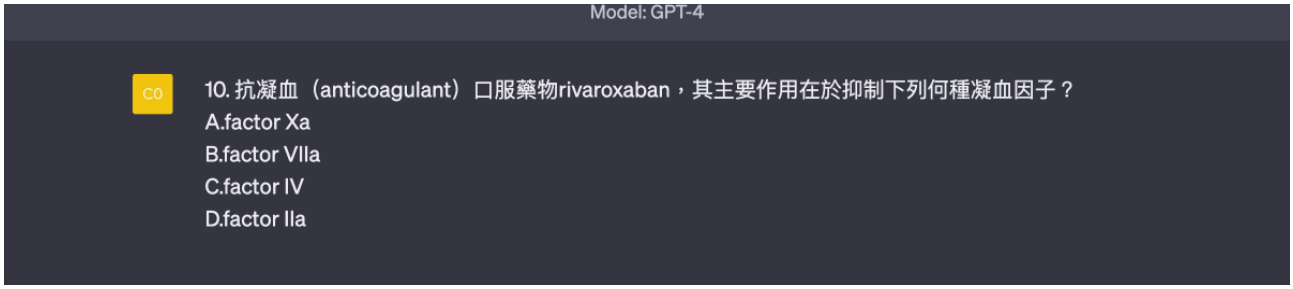


Figure 2. Template of a judgment question (choose 1 correct statement from 4 options, requiring medium-level thinking).

Figure 3. Template of a reverse question (choose 1 incorrect statement from 4 options, requiring medium- to high- level thinking).

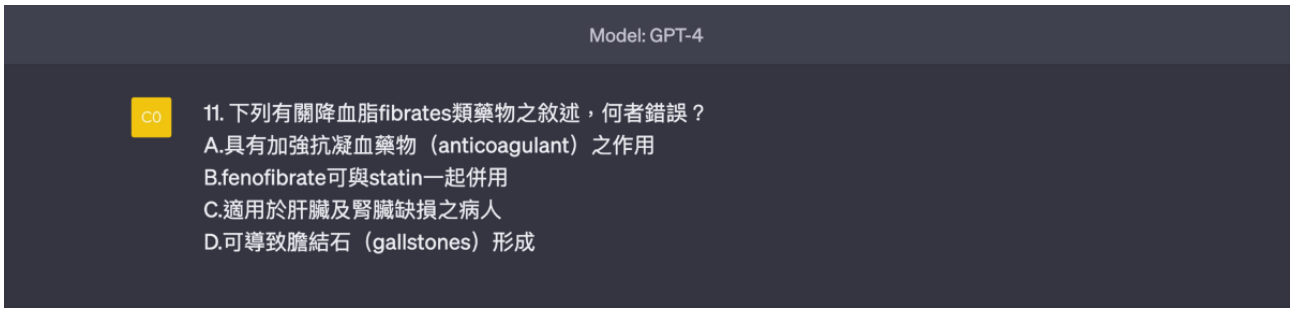
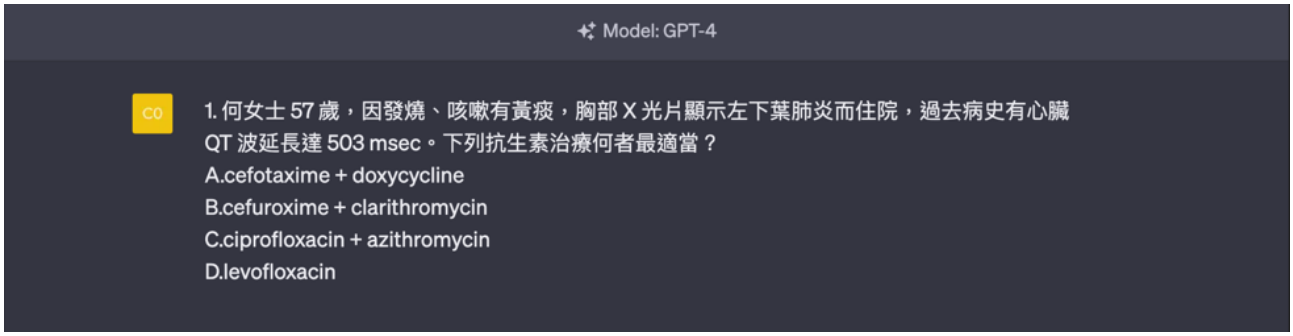


Figure 4. Template of a comprehension questions (multiple-choice or matching types, requiring high- level thinking).

XSL•FO
RenderX

Figure 5. Template of a case scenario question.



Statistical Analysis

Microsoft Excel 2019 was used to compare the accuracy rates of the 2 models. χ^2 tests were used to compare the overall accuracy rates of answers obtained using GPT-3.5 and GPT-4. McNemar tests were used to compare the consistency in answers between GPT-3.5 and GPT-4, and for the calculation-based and situational question types using R software (version 4.2.2; R Foundation for Statistical Computing).

Ethical Considerations

This study involved comparing the performance of ChatGPT-4 and ChatGPT-3.5 in the pharmacist licensing examination. It did not involve human participants. As per the guidelines of the 'Human Research Cases Exempted from Ethics Review Board' issued by the Ministry of Health and Welfare, Taiwan, this study was exempted from Ethics Review Board analysis.

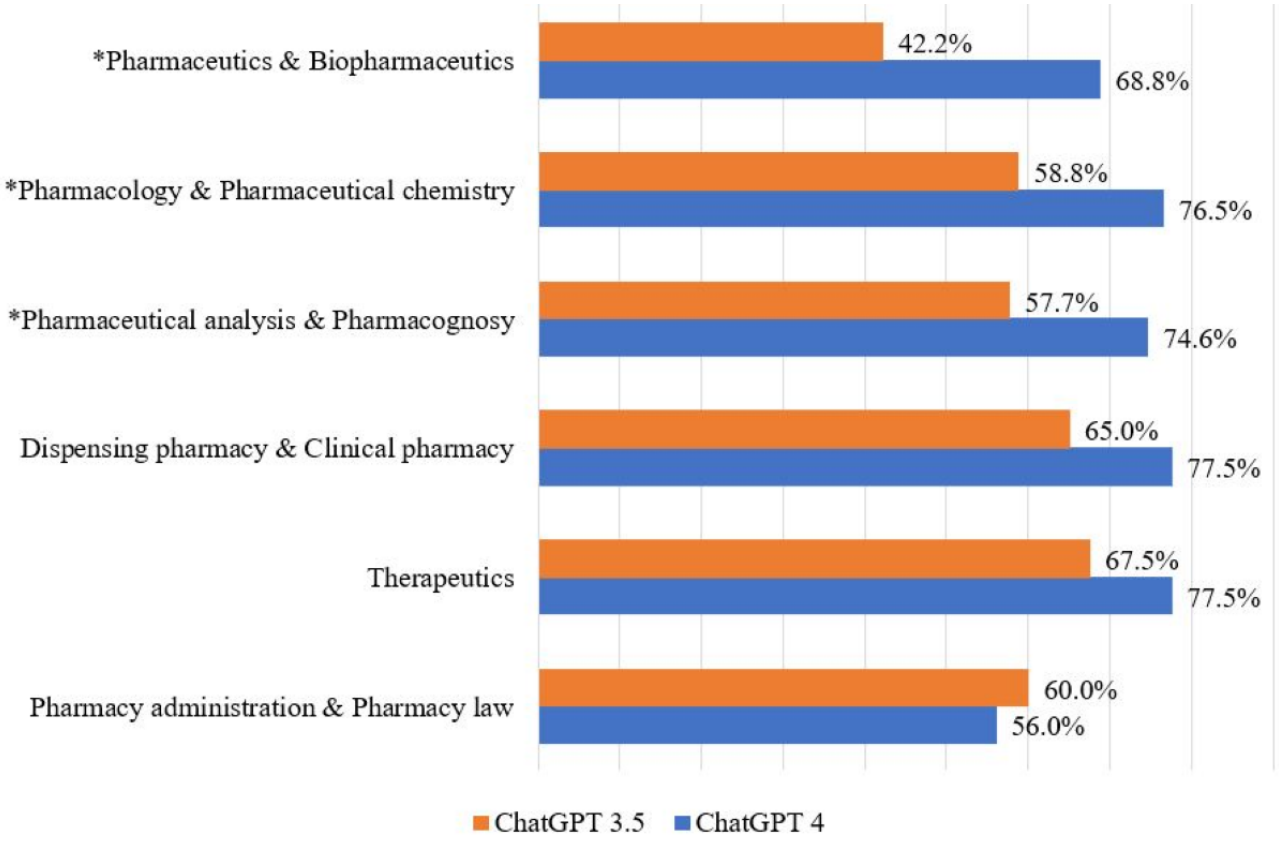
Results

Accuracy in Different Subjects

In total, 203 and 210 questions were included for analysis from the first- and second-stage examinations, respectively, after excluding 37 questions containing graphical elements (N=413) (Figure 6). GPT-4 had an overall accuracy of 72.9% (301/413), easily passing the test (60% threshold) and outperforming GPT-3.5 which achieved an accuracy of 59.1% (244/413; $P<.001$). In terms of accuracy by stage, GPT-4's overall accuracy was significantly higher than that of GPT-3.5 (73.4% vs 53.2% or 149/203 vs 108/203; $P<.001$) in basic subjects of the first stage. GPT-4 also significantly outperformed GPT-3.5 in each of the 3 basic subjects. In the clinical subjects of the second stage, GPT-4's accuracy was higher but not statistically significant than that of GPT-3.5 (72.4% vs 64.8% or 152/210 vs 136/210; $P=.096$). In pharmacy administration and pharmacy law, GPT-4's accuracy was lower than that of GPT-3.5 (56% vs 60% or 28/50 vs 30/50; $P=.96$). Among individual subjects, significant differences were observed in pharmacology and pharmaceutical chemistry ($P=.02$), pharmaceutical analysis and pharmacognosy ($P=.02$), and pharmaceutics and biopharmaceutics ($P=.002$). No significant differences were noted in dispensing pharmacy and clinical pharmacy ($P=.07$), pharmacotherapeutics ($P=.10$), and pharmacy administration and pharmacy law ($P=.48$).

XSL•FO
RenderX

Figure 6. Accuracy comparison of ChatGPT-3.5 and ChatGPT-4 across different subjects. *P<.05.



The overall consistency among answers significantly differed between the 2 models (68%, P<.001), with GPT-4 showing consistent correct answers in 49.4% (n=204) of cases and consistent incorrect answers in 18.6% (n=77) of cases (Table 1).

Table . Performance comparison of consistency between ChatGPT-3.5 and ChatGPT-4.

ChatGPT-3.5 responses	GPT-4	
	Correct answers, n (%)	Incorrect answers, n (%)
Correct answer	204 (49.4)	38 (9.2)
Incorrect answer	94 (22.8)	77 (18.6)

Accuracy in Different Question Types

Among the 413 examination questions analyzed, memory-based questions were the most common (n=254, 61.5%), followed by judgment questions (n=82, 19.9%), reverse questions (n=46, 11.1%), and comprehension questions (n=31, 7.5%). GPT-4 and GPT-3.5 did not differ significantly in terms of accuracy of answers between question types (P=.461 vs P=.18; Table 2). GPT-4 is significantly better than GPT-3.5 in memory-based questions (P<.001) and comprehension-based questions(P=.03).

Table . Accuracy comparison of ChatGPT-3.5 and ChatGPT-4 by question type.

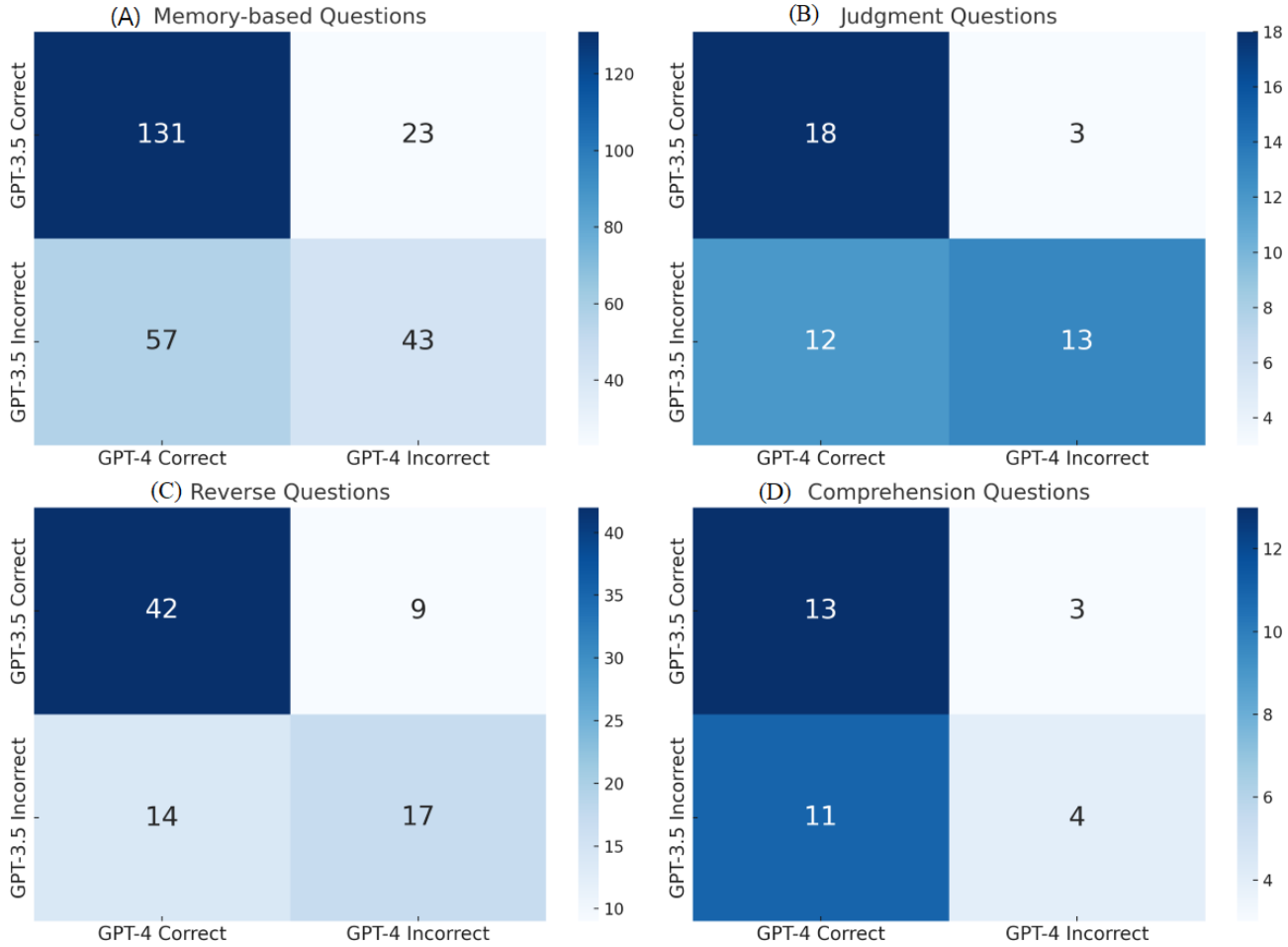
Question type	GPT-3.5 Correct answers, n (%)	GPT-4 Correct answers, n (%)	Total, n (%)	P value
Memory-based questions	155 (61)	188 (74)	254 (61.5)	<.001[a]
Judgment questions	21 (45.7)	30 (65.2)	46 (11.1)	.06
Reverse questions	51 (62.6)	56 (68.3)	82 (19.9)	.41
Comprehension questions	16 (51.6)	24 (77.4)	31 (7.5)	.03[a]

[a]P<.05.

Figure 7 shows the performance comparison of GPT-3.5 and GPT-4 across question types. The data provided insights into the relative strengths and weaknesses of each model.

XSL•FO
RenderX

Figure 7. Performance comparison of GPT-3.5 and GPT-4 across question types (A) memory-based, (B) judgement, (C) reverse , and (D) comprehension. The heatmaps display the number of answers, with darker shades indicating higher counts of correct responses and highlighting model performance.



Further analysis of the discrepancies between the models revealed no significant difference in questions answered incorrectly by GPT-3.5 but correctly by GPT-4 (n=94) and vice versa (n=38) across the 4 question types (*P*=.27 vs *P*=.95).

For calculation-based questions, GPT-4 showed higher accuracy than that of GPT-3.5 (80% vs 40%, *P*=.03), with the most pronounced difference in pharmaceutics and biopharmaceutics subjects. In scenario-based questions, GPT-4 also outperformed GPT-3.5 in terms of accuracy (63% vs 44.4%, *P*=.41), though the difference was nonsignificant.

Discussion

Principal Findings

This study demonstrates that GPT-4 significantly outperformed GPT-3.5 in the Taiwan NPLE, surpassing the passing threshold, especially in basic pharmacy subjects. These subjects, which have only a 13.82% passing rate among human students, are particularly challenging. GPT-4 excelled in areas such as pharmacology, pharmaceutical chemistry, pharmaceutical analysis, and pharmaceutics, consistently providing correct answers and comprehensive explanations. Although GPT-4 also performed better than GPT-3.5 in clinical subjects such as dispensing pharmacy and therapeutics, the performance gap was narrower in these areas.

In specific subjects like pharmacodynamics, pharmacokinetics, and drug-related topics in the autonomic nervous system, GPT-4 consistently provided accurate responses, where GPT-3.5 often faltered. Additionally, GPT-4 exhibited superior accuracy in bioavailability, dosing, and pharmacokinetic calculations. However, GPT-4's accuracy dropped in topics like herbal medicines and pharmacy law, emphasizing the need for further model refinement in these areas [30].

Comparison with Literature

Previous studies have established that GPT-4 consistently outperforms GPT-3.5 in various medical exams, including the Australian Medical Licensing Examination [31], Canadian Radiology Examination [15], Turkish Medical Examination [32], and Japanese Medical Licensing Examination [33]. In many of these examinations, GPT-4 consistently achieved scores above 70% [34-36]. This study aligns with those findings, showing GPT-4's superior performance in the Taiwan NPLE. Unlike prior research that focused on real-world clinical applications [37-43], this study comprehensively assessed the models across various pharmacy domains.

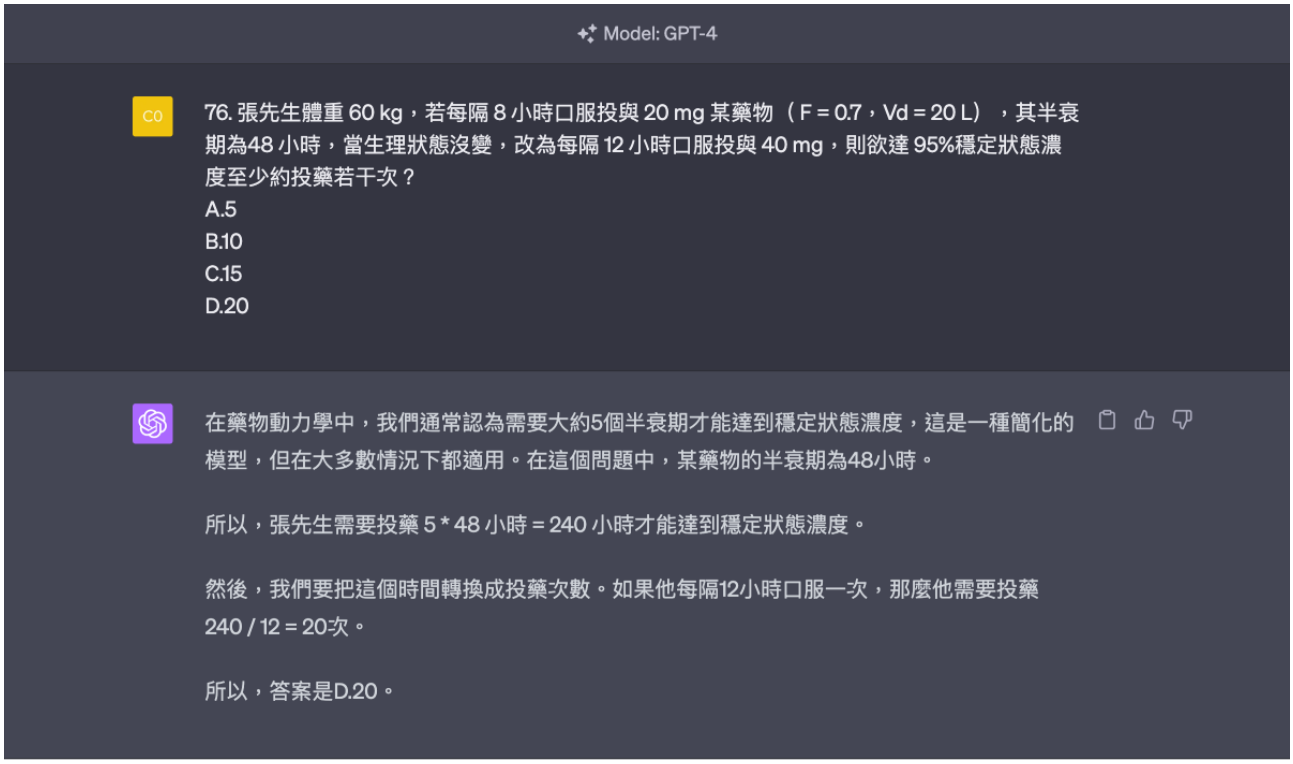
A study by Choi [44] reported that GPT-3.5 performed well on memory-based questions but struggled with problem-solving, whereas GPT-4 demonstrated better performance in comprehension and judgment tasks. Similarly, a radiology study suggested that GPT-4 outperformed GPT-3.5 on higher-order thinking questions but not on lower-order questions [15]. These

XSL•FO
RenderX

findings slightly differ from the results of our study, where GPT-3.5 exhibited higher accuracy in both memory-based (low-level thinking) and reverse (mid-level thinking) questions. However, GPT-4 surpassed GPT-3.5 across all question types, particularly in comprehension (high-level thinking) and memory-based (low-level thinking) questions. In judgment, reverse, and comprehension questions—tasks that demand more advanced reasoning—GPT-4 demonstrated superior accuracy with fewer errors compared to GPT-3.5. Additionally, GPT-4's ability to correct errors made by GPT-3.5 reinforces its potential as a more reliable model for pharmacy-related assessments.

Further, GPT-4 significantly outperformed GPT-3.5 in calculation questions. While GPT-3.5 provided step-by-step explanations but often guessed the final answer—a phenomenon known as hallucination' due to insufficient training—GPT-4 exhibited stronger logical reasoning (Figure 8) with over 80% accuracy. However, it still made errors in 20% of cases, indicating the need for needed during its use [21,45]. In clinical applications, modifying prompts has been shown to improve GPT's accuracy [46]. For integrated analysis questions, GPT-4's performance was slightly better than GPT-3.5, consistent with findings from a nursing licensure examination in Japan [14].

Figure 8. Template of the questions that GPT-4 exhibited stronger logical reasoning.



Implications for Education

The study highlights GPT-4's potential as an educational tool, particularly in pharmacy education. GPT-4 can offer extensive practice opportunities for pharmacy students across both basic and clinical subjects, providing both correct answers and detailed explanations [18,47] to enhance understanding. Given the lower passing rates among pharmacy students in basic subjects among that were challenging, GPT-4 could assist in individualized learning. Its strength in comprehension and integrated analysis questions makes it a valuable resource for fostering critical thinking skills.

Despite its advancements over GPT-3.5, GPT-4's occasional inconsistencies suggest that model stability is not yet perfect. Questions correctly answered by GPT-3.5 were not always consistently answered by GPT-4. Nevertheless, GPT-4's accuracy, approaching 80% suggests that it can serve as an effective learning supplement, provided educators guide students in minimizing potential errors. For instance, specifying clearer prompts, such as "Please do not add your own opinions", may help mitigate hallucinations and enhance its use in educational settings.

In addition, educators should consider adjusting the format of examinations by replacing memory-based questions with comprehension questions, which can reduce the chances of guessing and better assess students' true intelligence.

Limitations

The primary limitation of this study is the time frame during which the models were tested (ie, from May 10 to July 20, 2023), which may affect the reproducibility of the results if retested in the future. Additionally, both GPT-3.5 and GPT-4 struggled with recognizing structural diagrams, limiting their performance in areas such as pharmaceutical chemistry and pharmacognosy. These limitations, consistent with previous research, highlight the need for cautious application of GPT models in fields that require visual recognition [11,48,49]. Additionally, the models showed poorer performance in subjects with less available training data and specific medical knowledge such as pharmacy law and traditional medicine, indicating potential biases in the models' training. We suggest that future

XSL•FO
RenderX

efforts in model development should focus on incorporating more diverse and comprehensive data to reduce such biases.

Conclusions

This study demonstrates that GPT-4 outperforms GPT-3.5 in the Taiwan NPLE, particularly in pharmacy expertise, calculation ability, and situational case studies, with a notable advantage in basic subjects. It is recommended that GPT-4 be applied in clinical pharmacy practice (ie, patient education, drug consultation) and pharmacy education, particularly to support self-directed learning. However, given its limitations, caution is advised when integrating GPT-4 into clinical settings and educational programs. Future research should focus on refining prompts, improving model stability, integrating medical databases, and enhancing comprehensive questions to evaluate student competence more effectively while minimizing the chance of guessing correct answers.

Acknowledgments

This work was supported by Taipei Veterans General Hospital Hsinchu Branch (2024-VHCT-P-0008) and the authors would like to thank Wallace Academic Editing (https://www.editing.tw/) for English language editing.

Conflicts of Interest

None declared.

References

1. ChatGPT: optimizing language models for dialogue. OpenAI. URL: https://chatgpt.r4wand.eu.org/ [accessed 2023-03-03]
2. Research index. OpenAI. URL: https://openai.com/research/gpt-4 [accessed 2023-08-03]
3. OpenAI. GPT-4 technical report. arXiv. Preprint posted online on Mar 4, 2024. [doi: 10.48550/arXiv.2303.08774]
4. Kosinski M. Evaluating large language models in theory of mind tasks. Proc Natl Acad Sci U S A 2024 Nov 5;121(45):e2405460121. [doi: 10.1073/pnas.2405460121] [Medline: 39471222]
5. Wang YM, Chen TJ. ChatGPT surges ahead: GPT-4 has arrived in the arena of medical research. J Chin Med Assoc 2023 Sep 1;86(9):784-785. [doi: 10.1097/JCMA.0000000000000955] [Medline: 37406215]
6. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 8;9:e45312. [doi: 10.2196/45312] [Medline: 36753318]
7. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Dig Health 2023 Feb;2(2):e0000198. [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
8. Huh S. Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination? A descriptive study. J Educ Eval Health Prof 2023;20:1. [doi: 10.3352/jeehp.2023.20.1] [Medline: 36627845]
9. Fijačko N, Gosak L, Štiglic G, Picard CT, John Douma M. Can ChatGPT pass the life support exams without entering the American Heart Association course? Resuscitation 2023 Apr;185:109732. [doi: 10.1016/j.resuscitation.2023.109732] [Medline: 36775020]
10. Weng TL, Wang YM, Chang S, Chen TJ, Hwang SJ. ChatGPT failed Taiwan's Family Medicine Board Exam. J Chin Med Assoc 2023 Aug 1;86(8):762-766. [doi: 10.1097/JCMA.0000000000000946] [Medline: 37294147]
11. Antaki F, Touma S, Milad D, El-Khoury J, Duval R. Evaluating the performance of ChatGPT in ophthalmology: an analysis of its successes and shortcomings. Ophthalmol Sci 2023 Dec;3(4):100324. [doi: 10.1016/j.xops.2023.100324] [Medline: 37334036]
12. Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on medical challenge problems. arXiv. Preprint posted online on Apr 12, 2023. [doi: 10.48550/arXiv.2303.13375]
13. Rosoł M, Gąsior JS, Łaba J, Korzeniewski K, Młyńczak M. Evaluation of the performance of GPT-3.5 and GPT-4 on the Polish Medical Final Examination. Sci Rep 2023 Nov 22;13(1):20512. [doi: 10.1038/s41598-023-46995-z] [Medline: 37993519]
14. Kaneda Y, Takahashi R, Kaneda U, et al. Assessing the performance of GPT-3.5 and GPT-4 on the 2023 Japanese Nursing Examination. Cureus 2023 Aug;15(8):e42924. [doi: 10.7759/cureus.42924] [Medline: 37667724]
15. Bhayana R, Bleakney RR, Krishna S. GPT-4 in radiology: improvements in advanced reasoning. Radiology 2023 Jun;307(5):e230987. [doi: 10.1148/radiol.230987] [Medline: 37191491]
16. Oh N, Choi GS, Lee WY. ChatGPT goes to the operating room: evaluating GPT-4 performance and its potential in surgical education and training in the era of large language models. Ann Surg Treat Res 2023 May;104(5):269-273. [doi: 10.4174/astr.2023.104.5.269] [Medline: 37179699]
17. Sabry Abdel-Messih M, Kamel Boulos MN. ChatGPT in clinical toxicology. JMIR Med Educ 2023 Mar 8;9:e46876. [doi: 10.2196/46876] [Medline: 36867743]

XSL•FO
RenderX

18. Nisar S, Aslam MS. Is ChatGPT a good tool for T&CM students in studying pharmacology? SSRN. Preprint posted online on Jan 17, 2023. [doi: 10.2139/ssrn.4324310]
19. Wang YM, Shen HW, Chen TJ. Performance of ChatGPT on the Pharmacist Licensing Examination in Taiwan. J Chin Med Assoc 2023 Jul 1;86(7):653-658. [doi: 10.1097/JCMA.0000000000000942] [Medline: 37227901]
20. He N, Yan Y, Wu Z, et al. Chat GPT-4 significantly surpasses GPT-3.5 in drug information queries. J Telemed Telecare 2023 Jun 22:1357633X231181922. [doi: 10.1177/1357633X231181922] [Medline: 37350055]
21. Li D, Yu J, Hu B, Xu Z, Zhang M. ExplainCPE: A free-text explanation benchmark of Chinese Pharmacist Examination. arXiv. Preprint posted online on Oct 26, 2023. [doi: 10.48550/arXiv.2305.12945]
22. Vert JP. How will generative AI disrupt data science in drug discovery? Nat Biotechnol 2023 Jun;41(6):750-751. [doi: 10.1038/s41587-023-01789-6] [Medline: 37156917]
23. Blanco-González A, Cabezón A, Seco-González A, et al. The role of AI in drug discovery: challenges, opportunities, and strategies. Pharmaceuticals (Basel) 2023 Jun 18;16(6):891. [doi: 10.3390/ph16060891] [Medline: 37375838]
24. Savage N. Drug discovery companies are customizing ChatGPT: here's how. Nat Biotechnol 2023 May;41(5):585-586. [doi: 10.1038/s41587-023-01788-7] [Medline: 37095351]
25. Wang H, Ding YJ, Luo Y. Future of ChatGPT in pharmacovigilance. Drug Saf 2023 Aug;46(8):711-713. [doi: 10.1007/s40264-023-01315-2] [Medline: 37306853]
26. Carpenter KA, Altman RB. Using GPT-3 to build a lexicon of drugs of abuse synonyms for social media pharmacovigilance. Biomolecules 2023 Feb 18;13(2):387. [doi: 10.3390/biom13020387] [Medline: 36830756]
27. Cloesmeijer ME, Janssen A, Koopman SF, Cnossen MH, Mathôt RAA, consortium S. ChatGPT in pharmacometrics? Potential opportunities and limitations. Br J Clin Pharmacol 2024 Jan;90(1):360-365. [doi: 10.1111/bcp.15895] [Medline: 37621112]
28. Sallam M, Salim N, Barakat M, Al-Tammemi A. ChatGPT applications in medical, dental, pharmacy, and public health education: a descriptive study highlighting the advantages and limitations. Narra J 2023 Mar 29;3(1):e103. [doi: 10.52225/narra.v3i1.103]
29. Zhu Y, Han D, Chen S, Zeng F, Wang C. How can ChatGPT benefit pharmacy: a case report on review writing. Preprints. Preprint posted online on Feb 20, 2023. [doi: 10.20944/preprints202302.0324.v1]
30. Hsu HY, Hsu KC, Hou SY, Wu CL, Hsieh YW, Cheng YD. Examining real-world medication consultations and drug-herb interactions: ChatGPT performance evaluation. JMIR Med Educ 2023 Aug 21;9:e48433. [doi: 10.2196/48433] [Medline: 37561097]
31. Kleinig O, Gao C, Bacchi S. This too shall pass: the performance of ChatGPT-3.5, ChatGPT-4 and New Bing in an Australian Medical Licensing Examination. Med J Aust 2023 Sep 4;219(5):237. [doi: 10.5694/mja2.52061] [Medline: 37528548]
32. Kılıç ME. AI in medical education: A comparative analysis of GPT-4 and GPT-3.5 on turkish medical specialization exam performance. medRxiv. Preprint posted online on Jul 12, 2023. [doi: 10.1101/2023.07.12.23292564]
33. Takagi S, Watari T, Erabi A, Sakaguchi K. Performance of GPT-3.5 and GPT-4 on the Japanese Medical Licensing Examination: comparison study. JMIR Med Educ 2023 Jun 29;9:e48002. [doi: 10.2196/48002] [Medline: 37384388]
34. Guerra GA, Hofmann H, Sobhani S, et al. GPT-4 artificial intelligence model outperforms ChatGPT, medical students, and neurosurgery residents on neurosurgery written board-like questions. World Neurosurg 2023 Nov;179:e160-e165. [doi: 10.1016/j.wneu.2023.08.042] [Medline: 37597659]
35. Lewandowski M, Łukowicz P, Świetlik D, Barańska-Rybak W. ChatGPT-3.5 and ChatGPT-4 dermatological knowledge level based on the Specialty Certificate Examination in dermatology. Clin Exp Dermatol 2024 Jun 25;49(7):686-691. [doi: 10.1093/ced/llad255] [Medline: 37540015]
36. Khorshidi H, Mohammadi A, Yousem DM, et al. Application of ChatGPT in multilingual medical education: How does ChatGPT fare in 2023's Iranian Residency Entrance Examination. Inform Med Unlocked 2023;41:101314. [doi: 10.1016/j.imu.2023.101314]
37. Huang X, Estau D, Liu X, Yu Y, Qin J, Li Z. Evaluating the performance of ChatGPT in clinical pharmacy: a comparative study of ChatGPT and clinical pharmacists. Br J Clin Pharmacol 2024 Jan;90(1):232-238. [doi: 10.1111/bcp.15896] [Medline: 37626010]
38. Jairoun AA, Al-Hemyari SS, Shahwan M, Humaid Alnuaimi GR, Zyoud SH, Jairoun M. ChatGPT: threat or boon to the future of pharmacy practice? Res Social Adm Pharm 2023 Jul;19(7):975-976. [doi: 10.1016/j.sapharm.2023.03.012] [Medline: 37061346]
39. Juhi A, Pipil N, Santra S, Mondal S, Behera JK, Mondal H. The capability of ChatGPT in predicting and explaining common drug-drug interactions. Cureus 2023 Mar;15(3):e36272. [doi: 10.7759/cureus.36272] [Medline: 37073184]
40. Davies NM. Adapting artificial intelligence into the evolution of pharmaceutical sciences and publishing: technological darwinism. J Pharm Pharm Sci 2023;26:11349. [doi: 10.3389/jpps.2023.11349] [Medline: 37034476]
41. Kleebayoon A, Wiwanitkit V. Performance and risks of ChatGPT used in drug information: comment. Eur J Hosp Pharm 2023 Dec 27;31(1):85-86. [doi: 10.1136/ejhpharm-2023-003864] [Medline: 37339863]
42. Mohammed M, Kumar N, Zawiah M, et al. Psychometric properties and assessment of knowledge, attitude, and practice towards ChatGPT in pharmacy practice and education: a study protocol. J Racial Ethn Health Disparities 2024 Aug;11(4):2284-2293. [doi: 10.1007/s40615-023-01696-1] [Medline: 37428357]

XSL•FO
RenderX

43. Abu-Farha R, Fino L, Al-Ashwal FY, et al. Evaluation of community pharmacists' perceptions and willingness to integrate ChatGPT into their pharmacy practice: a study from Jordan. J Am Pharm Assoc 2023 Nov;63(6):1761-1767. [doi: 10.1016/j.japh.2023.08.020]
44. Choi W. Assessment of the capacity of chatgpt as a self-learning tool in medical pharmacology: a study using mcqs. BMC Med Educ 2023 Nov 13;23(1):864. [doi: 10.1186/s12909-023-04832-x] [Medline: 37957666]
45. Snoswell CL, Falconer N, Snoswell AJ. Pharmacist vs machine: pharmacy services in the age of large language models. Res Social Adm Pharm 2023 Jun;19(6):843-844. [doi: 10.1016/j.sapharm.2023.03.006] [Medline: 36907776]
46. Meskó B. Prompt engineering as an important emerging skill for medical professionals: tutorial. J Med Internet Res 2023 Oct 4;25:e50638. [doi: 10.2196/50638] [Medline: 37792434]
47. Krumborg JR, Mikkelsen N, Damkier P, et al. ChatGPT: first glance from a perspective of clinical pharmacology. Basic Clin Pharma Tox 2023 Jul;133(1):3-5. [doi: 10.1111/bcpt.13879]
48. Fergus S, Botha M, Ostovar M. Evaluating academic answers generated using ChatGPT. J Chem Educ 2023 Apr 11;100(4):1672-1675. [doi: 10.1021/acs.jchemed.3c00087]
49. Massey PA, Montgomery C, Zhang AS. Comparison of ChatGPT-3.5, ChatGPT-4, and orthopaedic resident performance on orthopaedic assessment examinations. J Am Acad Orthop Surg 2023 Dec 1;31(23):1173-1179. [doi: 10.5435/JAAOS-D-23-00396] [Medline: 37671415]

Abbreviations

GPT-3.5: ChatGPT-3.5
GPT-4: ChatGPT-4
NPLE: National Pharmacist Licensing Examination
USMLE: United States Medical Licensing Examination

Edited by B Lesselroth; submitted 28.01.24; peer-reviewed by G Farid, S Zhai, WY Jen; revised version received 26.09.24; accepted 17.12.24; published 17.01.25.

Please cite as:
Wang YM, Shen HW, Chen TJ, Chiang SC, Lin TG
Performance of ChatGPT-3.5 and ChatGPT-4 in the Taiwan National Pharmacist Licensing Examination: Comparative Evaluation Study
JMIR Med Educ 2025;11:e56850
URL: https://mededu.jmir.org/2025/1/e56850
doi:10.2196/56850

© Ying-Mei Wang, Hung-Wei Shen, Tzeng-Ji Chen, Shu-Chiung Chiang, Ting-Guan Lin. Originally published in JMIR Medical Education (https://mededu.jmir.org), 17.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Guidelines for Patient-Centered Documentation in the Era of Open Notes: Qualitative Study

Anita Vanka[1,2], MD; Katherine T Johnston[2,3], MD; Tom Delbanco[1,2], MD; Catherine M DesRoches[1], DrPH; Annalays Garcia[1], MD; Liz Salmi[1], AS; Charlotte Blease[4], PhD

[1]Division of General Medicine, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States
[2]Harvard Medical School, Boston, MA, United States
[3]Department of Medicine, Massachusetts General Hospital, Boston, MA, United States
[4]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

Corresponding Author:
Anita Vanka, MD
Division of General Medicine
Department of Medicine
Beth Israel Deaconess Medical Center
330 Brookline Avenue
Deaconess 301
Boston, MA, 02215-5400
United States
Phone: 1 617 632 8350
Fax: 1 617 632 8261
Email: avanka@bidmc.harvard.edu

Abstract

Background: Patients in the United States have recently gained federally mandated, free, and ready electronic access to clinicians' computerized notes in their medical records ("open notes"). This change from longstanding practice can benefit patients in clinically important ways, but studies show some patients feel judged or stigmatized by words or phrases embedded in their records. Therefore, it is imperative that clinicians adopt documentation techniques that help both to empower patients and minimize potential harms.

Objective: At a time when open and transparent communication among patients, families, and clinicians can spread more easily throughout medical practice, this inquiry aims to develop informed guidelines for documentation in medical records.

Methods: Through a series of focus groups, preliminary guidelines for documentation language in medical records were developed by health professionals and patients. Using a structured focus group decision guide, we conducted 4 group meetings with different sets of 27 participants: physicians experienced with writing open notes (n=5), patients accustomed to reviewing their notes (n=8), medical student educators (n=7), and resident physicians (n=7). To generate themes, we used an iterative coding process. First-order codes were grouped into second-order themes based on the commonality of meanings.

Results: The participants identified 10 important guidelines as a preliminary framework for developing notes sensitive to patients' needs.

Conclusions: The process identified 10 discrete themes that can help clinicians use and spread patient-centered documentation.

(JMIR Med Educ 2025;11:e59301) doi:10.2196/59301

KEYWORDS

open notes; patient-centered documentation skills; medical student education; 21st Century Cures Act

Introduction

Reflecting long-standing tradition, medical record notes documenting clinical encounters have primarily served the doctors or other health professionals preparing them. Their diverse functions include accurate documentation of a patient's unique circumstance, refreshing clinicians' memories, documenting diagnostic reasoning, communicating cogently with colleagues, justifying charges for encounters, and serving as material for assaying quality of care.

XSL•FO
RenderX

Until about 15 years ago, clinicians prepared such notes rarely with patients or their families envisioned as potential recipients [1]. Since the turn of this century, however, the movement toward more open and transparent communication with patients has grown, and since April 2021, federal rules in the United States now mandate that all patients (with very limited permitted exceptions) are offered online and rapid access to their clinical records, including the notes written by clinicians ("open notes") [2,3].

Whether one note can serve diverse recipients remains an open question, however, studies suggest room for optimism. A large majority of clinicians experienced with open notes favor their continuation, and few report "dumbing down" what they write [4]. Moreover, extensive survey and qualitative research demonstrate that patients who review their records and read their notes feel more involved in and knowledgeable about their care, report being better prepared for visits, and indicate they are more likely to follow their clinicians' advice [5-11]. However, words matter, and studies also show that patients can feel judged, stigmatized, or offended by their notes, with potentially adverse effects on the clinician-patient relationship [12-15]. For example, a recent study at 3 diverse health systems found that 1 in 10 patients reported feeling judged or offended by an outpatient note, reflecting the perception that the note contained errors, surprises, inappropriate labeling, or evidence of disrespect [16].

To date, little empirically informed counsel about best practices with respect to patient-centered documentation has been published [17]. Although studies have demonstrated that certain language in notes influences clinician attitudes toward patients and that specific words used can negatively impact the clinician-patient relationship, what language providers should use or avoid has not been clearly described [11,15]. Within the current medical literature, some recommendations have been offered on how clinicians might better prepare for the era of open notes [18]. However, there are no evidence-informed guidelines describing concrete approaches to patient-centered language in clinical notes, and such guidance may be important for developing mindful practices.

In preparation for an educational intervention with medical students and practitioners supervising their work [19], we aimed to address this gap in knowledge and practice by drawing on the perspectives of patients, physicians facile with open notes, medical educators responsible for teaching clinical skills to early learners, and medical residents, who often teach students directly. Following dedicated discussions with these 4 groups and subsequent thematic analysis of their perspectives, we developed a set of guidelines as a preliminary framework for future initiatives aimed at teaching patient-centered documentation skills to medical students, their preceptors, and a broad range of clinicians.

Methods

We conducted 4 focus groups with discrete groups of individuals to gain an understanding of their perspectives and experiences with written medical documentation. Our goal was to develop guiding principles for best practices in patient-centered documentation skills.

Design

Our focus groups addressed experiences with written medical documentation (patient notes), with particular attention to the language used in notes. We sought a range of perspectives and structured the groups based on the type of participant. Within their respective groups, the aim was to create opportunities for interaction and comparison of responses among participants [20,21]. This methodology can generate a large volume of responses, and we anticipated a robust discussion of experiences among participants. Although there is no consensus about the ideal number of participants in focus groups, our goal was to recruit between 6 and 10 participants per group [22]. All study procedures received ethical approval in March 2022 and met exempt status both from the Beth Israel Deaconess Medical Center institutional review board (reference number 2022P000079) and the Mass General Brigham institutional review board (reference number 2022P000635).

Recruitment and Participants

We convened four groups of participants: (1) patients familiar with open notes, (2) practicing physicians facile with open notes, (3) medical student educators involved with teaching clinical skills, and (4) resident physicians working closely with medical students. Between March and April 2022, we used various forms of outreach to recruit participants: For the patient focus group, LS identified, recruited, and contacted patients through email with a flyer attachment. For the group of practicing physicians, TD identified and AG recruited individuals having active experiences with open notes at academic institutions across the country. For the group of medical student teachers, AV and KJ recruited physician educators through the national medical educator list, which included Clerkship Directors in Internal Medicine and Directors of Clinical Skills. For the focus group of residents, AV and KJ identified and contacted through email residents from Beth Israel Deaconess Medical Center and Massachusetts General Hospital who had expressed interest in medical student education. For their time, patient participants were offered a US $100 honorarium, and physicians a US $50 honorarium.

Between April and May 2022, we convened four 90-minute focus groups, each composed of specific types of participants as described above. Focus groups were conducted through Zoom (Zoom Video Communications) and recorded for transcription.

Format of Focus Groups

Each group followed the same script, with the meetings facilitated by 2 of the study leaders (AV and KJ). Drawing on a variety of stakeholder perspectives, the research team collectively devised the structured interview script. The team included medical educators, practicing physicians, a health services researcher, a patient advocacy researcher, a health services researcher, and a medical ethicist. The script was developed based on the group's experiential, ethical, and practical experience. To assess face validity, the questions were further honed, refined, and pretested with 4 doctors and 3 patients, leading to further refinements in the wording of

XSL•FO
RenderX

questions and prompts (Textbox 1). Groups opened with a standard anonymity disclosure, along with a description of the reason behind the inquiries. We offered participants the opportunity to opt out at any time during the process. Participants were encouraged to set their screen name, however, they felt comfortable and to keep their video off if so desired. To preserve anonymity, we asked participants not to mention each other's names during the session.

Textbox 1. Focus group questions for patient and physician groups.

(10 minutes per question)

Question 1: What do you recall from any previous learning experiences that focused on writing notes that patients will read, or reading notes from the perspective of a patient?

- For the patient group: "What do you recall from your experience of reading your clinical notes?"
- Follow-up question: What are some concrete examples of this?

Question 2: What should early medical students know about writing notes that will be useful to patients and encourage partnering or engagement in their care?

- Follow-up question: What are some concrete examples of this?

Question 3: What should early medical students know about writing notes with words or phrases that could be harmful to patients or their relationship with their physician?

- Follow-up question: What are some concrete examples of this?

Question 4: What should early medical students know regarding how medical vernacular or acronyms may be perceived by patients?

- For patient group: "What has been your experience when reading notes containing medical vernacular or acronyms?"
- Follow-up questions:
- What are examples of vernacular that should be avoided due to possible patient harm or creating unwanted bias?
- Are there common acronyms that should be avoided?

Question 5: What should early medical students know about including the patient's voice in the notes?

- For patient group: Describe some ways in which a physician could write your words, or ensure that the writing of your medical concerns represents your lived experience.
- Follow-up question:
- What documentation approaches could a student follow to help the patient feel authentically seen and heard when reading their clinical notes?

Question 6: What should early medical students know about how words or phrases in clinical notes may convey bias?

- For patient group: "What has been your experience in reading notes with words or phrases that you feel convey bias?"
- Follow-up question: What are some concrete examples of this?

Question 7: What key content areas are more sensitive for patient readers and should require students to receive specific guidance on documentation?

- Follow-up question:
- What topics or themes should receive special attention and teaching?
- Are there any key areas that should be omitted or avoided in notes unless discussed and reviewed with faculty?
- Prompt examples: race, obesity, firearm ownership, gender identity and health, sexual identity and health, substance use, and so on.

(5 minutes for wrap-up and debrief and closing thoughts)

Analysis

We designed and conducted an inductive thematic content analysis of the transcribed focus groups [23]. This approach was used because it is particularly appropriate for analyzing textual data by identifying patterns, themes, or categories that emerge from the data itself, rather than imposing predetermined categories or codes. This approach facilitates the identification of nuanced themes arising directly from the data, making it particularly useful when exploring new or less understood phenomena [24]. Responses were analyzed by 2 members of the research team (AV and KJ). Both are medical educators and general internists in the United States with experience in sharing online access to patients' health records (AV is an inpatient physician and KJ is a primary care physician). The research team was diverse in age and background: the lead author and one of the coders (AV) identified as of Indian background, another author (AG) identified as having Cuban descent, and

XSL•FO
RenderX

the patient-researcher author (LS) identified as having both physical and cognitive disabilities.

First, AV and KJ read the transcripts to familiarize themselves with the responses. Second, AV and KJ independently created codes through the selection of excerpts from each of the 4 focus groups. The codes were then reviewed jointly by AV and KJ to come to a consensus. Using the common list, new excerpts from each of the 4 transcripts were coded by AV and KJ, and the codes were further refined until a consensus was reached. Subsequently, first-order codes were grouped into second-order themes based on commonality of meaning. Representative comments for each theme were identified by authors AV and AG.

Ethical Considerations

All study procedures received ethical approval in March 2022 and met exempt status both from the Beth Israel Deaconess Medical Center institutional review board (reference number 2022P000079) and the Mass General Brigham institutional review board (reference number 2022P000635). This study did not meet criteria for human subjects research at either institution and thereby was deemed exempt.

Results

Overview

A total of 27 individuals participated in the focus groups: 8 patients (5 women, 2 men, and 1 individual identifying as gender nonbinary), 5 physicians with experience preparing open notes (2 women and 3 men), 7 medical student educators (6 women and 1 man), and 7 resident physicians (6 women and 1 man). The participants from the patient, physician, and medical educator groups were from diverse geographical regions of the United States. Based at large academic centers, physician participants represented different clinical specialties including internal medicine, pediatrics, and behavioral health. Educator participants were all involved with either leadership or teaching foundational clinical skills at various medical schools in the country. The resident physician participants were recruited from 2 large academic health centers in the greater Boston area.

Using the iterative coding process [25], we identified 10 major themes that could serve as guidelines for patient-centered documentation (Textbox 2). We discuss these in greater detail below, with representative comments illustrative for each theme in Table 1.

Textbox 2. Checklist of guidelines for patient-centered documentation: major themes.

Themes

- Use person-first language.
- Refer to your patients as how they want to be identified.
- Avoid abbreviations and acronyms, especially if not officially approved by the practice.
- Say what you write and write what you say.
- Verify past history information before recording it in the note.
- Avoid words that may convey bias or judgment.
- Keep descriptions of physical examinations objective.
- Empower your patients with encouraging words and clear next steps.
- Pay close attention to sensitive topics, including but not limited to sexual history, trauma history, substance use history, mental health, or illness.
- Write from your perspective.

XSL•FO
RenderX

Table 1. Examples and suggestions for each identified best practice for patient-centered documentation.

Theme	Examples and suggestions
Use person-first language	• I don't see necessarily a boundary between the language you use to describe a patient out loud and what you write in your note. It's all part of a single way that you frame patients. It's not a diabetic patient. It's a patient with diabetes. I really don't think about the way that I document my notes differently than the way that I just think about my patients, which is to think about them in a person-centered way. [#2, Medical Educator] • Even though I've lost weight, the word morbidly obese is in every note, every note I see. That doesn't bring me joy or comfort me or get me to want to interact in a positive manner. My issues have nothing to do with my weight and never did. I developed a MRSA staph infection; it was a healthcare-acquired infection.There are certain things where weight doesn't always factor in, and you don't always need weight. [#1, Patient]
Refer to your patients as how they want to be identified	• I have found the doctors that I have currently are really good about using he/him pronouns for me and referring to me as using masculine identifying language. I have had providers who are not. Never knowing what I'm going to run into in those notes is very anxiety-inducing on the gender part alone, right? Let alone, does this person think that my gender is influenced by or influencing any of my other health issues? [#5, Patient] • I tend to write and call the patient whatever they like to go by, so I'll say like, Bill is a 75-year-old man, instead of Mr. Smith, for instance. For nonbinary or transgender patients, I try to ask them what gender they would like me to document rather than just assuming it's transgender male. Then in the social history and medical history, perhaps elaborating on it more, but trying to give them voice in what I'm documenting. I also think that mentioning race without having any clear connection to anything should be discouraged completely. [#1, Resident]
Avoid abbreviations and acronyms, especially if not officially approved by the practice	• I was struck by the idea that it almost felt, as a medical student, as if you were being inducted in a secret society. It was this secret language that you now all understood, and that's what unified everyone who was part of that. It really is just so unnecessary at this point. I think what's already come out here is that there clearly are regional differences. I've never heard of "MOP" ["mother of patient"], but "ISO"["in the setting of"] is rampant around here, so clearly, there are differences in terms of some of the abbreviations that are used in specific regions and areas, and we can't then even understand each other's language. [#2, Physician] • Just the other day, we were talking about PSA and how it's prostate-specific antigen, but also pseudomonas, and also pseudoaneurysm. If a patient reads PSA and looks that up for themselves on Google, they're going to find a million things that could be. If you're going to use an acronym, maybe don't use ones that have multiple different meanings even in the medical world, and the most important parts of their diagnoses should not be abbreviated. [#1, Resident] • Medical jargon and all this evolved as a way of having more succinct communication, being able to communicate among clinical teams. I think there is some value to that, but I think what's important for a lot of educators to keep in mind is, this is an area where the students are actually—while they're learning, they also should, in some ways, be teachers for all of us who have been doing this for some time, right, and how they can take their lack of being indoctrinated by the system and bring that from the bottom up. [#3, Physician] • Then some acronyms. Point out we say SOB for shortness of breath, and, obviously, has other meanings. [#2, Resident]
Say what you write and write what you say	• There probably shouldn't be anything in the note that hasn't been discussed. If you're telling someone a diagnosis, I don't think you should go back in the note and say, "And this diagnosis is terminal" if you haven't discussed it with the patient. Knowing that the patient is likely going to go back and read the note, I don't think that it should be put in the note. Or, if the physician feels like it's important that it's put in the note, make sure it's discussed with the patient as well. There's no point in writing a good note if none of it was even said to the patient. [#7, Patient] • When we have a discussion and he puts it right in my note right then and there, those are kinds of things that help me. Today I had a call from the pharmacist about a med. They would have given me the wrong med if I wouldn't have stood up for myself and wouldn't have known what I was supposed to have, and wouldn't have had my notes there, because basically I read it right from my note to him. I think the notes become more and more valuable. [#1, Patient]

XSL•FO
RenderX

Theme	Examples and suggestions
Verify past history information before recording it in the note	• We've evolved as people from year 1 to year 5 over the course of a relationship of knowing a doctor, and our social histories often don't reflect that change. What could have been pertinent 5 years ago for a patient might not be anymore, and depending on what they were experiencing at that point in time, could be biasing. [#1, Resident] • So now it's on my to-do list when I go to appointments to make sure that's changed. Really, that shouldn't be a priority of mine when we go into medical appointments to say, "Hey, I read the note, and this needs to be changed, 'cause this doesn't represent who I am.' I think everyone has assumed the person before them has done it, and it's accurate. It takes a lot of time and a lot of effort and a lot of energy for us to be able to correct mistakes that are in our patient portals, our notes. I think it's 1 of those things where I just want to remind medical students, you can do 2 things to check with patients. Like, "Hey, I have all this down. Is that right?" Two, "If you see anything in here that's wrong, send me a MyChart message or EHR or whatever message, so I can fix it. [#5, Patient] • Some things that I notice that get copy-forwarded a lot in notes that create bias in the reader, 1 would be substance use disorders, and not really specifying or clarifying it, right? It will just say, "Polysubstance use disorder." Things like that, I think definitely have the ability to bias. [#1, Resident]
Avoid words that may convey bias or judgment	• You don't need to use the word complain. It's a pretty negative connotation. Patient "reports", or just "is having". I think avoiding the word complaint is pretty important." [#2, Resident] • I think, if you're ever thinking about using a term like that, it should be more about, what are the barriers this patient has to achieving care, and talking more about that and not just using 1 word. Using more language is actually helpful, and med students often have more time to actually ask patients about these things and have a conversation. I think that could be a really empowering point for medical students to learn more about the determinants of health that are affecting the patient more so than being like, "This patient doesn't take their medications. [#5, Resident] • I also think that mentioning race without having any clear connection to anything is—should be discouraged completely. It's such a part of vernacular, especially more physicians who are trained in a different era, so we've been actively encouraging students not to include that unless it has a very clear reason. [#2, Medical Educator] • I could imagine, sometimes, quotation marks are used to just directly capture what the patient said. I think it's about striking that balance of bringing the patient language in and having some of it as verbatim as directly as possible so that the patient can see what they said was actually recorded, but then a key part of what we're doing in notes is really clinical translation. It's about going from verbatim conversation to clinical processing and then recording our clinically processed thoughts. [#3, Physician]
Keep physical examination descriptions objective	• General appearance is very tricky, and I think we all have to ask ourselves, when is that relevant and why, and then really distill it down to what's clinically meaningful in the least judgmental way possible. Is it important that the patient is disheveled or older than stated age or has poor hygiene? It might be, but then you have to figure out, how do you relay that in the most respectful manner possible? [#5, Medical Educator]
Empower your patients with encouraging words and clear next steps	• There's a book called "The 15 Minute Hour" that's geared toward primary care providers, and 1 of my favorite takeaways from it that has always stuck with me over the years is the idea of using the word "yet" at the end of a sentence or a phrase. It's supposed to emphasize that you're not at a dead end. You can still do this. The patient has not been able to quit smoking yet. The patient has not lost weight yet. [#5, Physician] • I somehow think that I would like medical notes to be imbued with a different perspective, that we're trying to teach patients to be friends with their disease…[…]…disease is not the enemy. [#4, Patient] • I have heard from some patients that we had engaged in providing feedback on our student notes that they would really like, in reading the assessment and plan, to have some understanding of what they can do next and what the evaluation, diagnostic evaluation plan means for them. Maybe some additional content around health coaching and steps that they can take in their lives to support their health too. It could be a great way to extend that partnership beyond the visit. [#5, Medical Educator]
Pay close attention to sensitive topics, including but not limited to sexual history, trauma history, substance history, mental health or illness	

XSL•FO
RenderX

Theme	Examples and suggestions
	• Some advice I've gotten from the psychiatry consulting team at my primary care practice is […] to just state the type of trauma it was, the years that it happened, the relation to the abuser, and what the patient went through….[…]…There's no need to go into extreme detail and quotations about what the patient confides with you in clinic about, that you can still have a therapeutic bond…[…]…instead aim to strike a balance of what's useful for other providers and not needed for the patient to be reading about themselves. [#7, Resident] • There are certainly very specific topics around mental health, around sexually transmitted illnesses, reproductive health, substance use, et cetera, that definitely should be addressed in a particular manner that respects the patient's privacy… […]…early learners could certainly benefit from understanding what the implications of access to that information could be. [#2, Physician] • […] documenting your full differential diagnosis, where you might have a patient who is coming in for what seems like a respiratory illness or pneumonia, but maybe under differential, you also have a rare lung disorder or lung cancer. It's trying to strike that balance…[…]…trainees, as part of their training, are taught to document the full differential …[…]…this is an area where we're all still trying to figure out the best practices because you do want to indicate to some degree that you are thinking about these other diagnoses, but in the real world […]you might not start with an initial visit by sharing that this may be cancer, but it may be something you bring up on a follow-up visit when your initial working diagnosis doesn't seem to be correct and it seems more serious than that. I know the official recommendation has been to just document what you discuss…[…]…I think it's more complicated than that. [#3, Physician]
Write from your perspective	• Something I've been experimenting with a lot is using the first person in the assessment and plan, which I don't know if I was ever explicitly told not to in medical school, but I feel like I thought I wasn't supposed to. Over the past year, I've started to come around to this idea of, the objective is the objective, but the assessment and plan is my medical opinion. A lot of times, I'll say, "I'm worried about", or, "I think this might be at play. I think it's unlikely, but maybe cancer. I think anxiety might be driving some of this." I feel like that really couches it as, this is my opinion of what's going on. It doesn't negate what the patient thinks. [#5, Physician]

Use Person-First Language

Across all 4 focus groups, participants agreed that patients should be referred to as individuals with certain clinical conditions, rather than identifying them primarily by their medical condition (Table 1). Some patients identified a preference for the use of discrete numbers when addressing a person's weight, rather than describing the individual as "obese" or even "with obesity," and to be mindful of whether weight needs to be in the note if not relevant to issues being addressed (Table 1). All participants from the patient group described the importance of being mindful of phrases and words in notes that can trigger trauma for patients. Finally, participants in all groups identified descriptions of words and phrases in the medical record considered potentially depersonalizing or dehumanizing.

Refer to Your Patients as How They Want to Be Identified

The participants recommended that patients be identified according to their own expressed preferences. Participants from the resident physicians' group suggested that when first introducing oneself to a patient, one should request and document how the patient wishes to be identified in the record. Patient participants noted that honorifics and gender identity should never be presumed (Table 1). In addition, patient participants explained that patients may want to be identified by their life roles, such as their profession, or background, in addition to honorifics and names. Medical educators recommended that introductory "History of Present Illness" sentences should mention factors important to understanding and planning the care of that person at that point in time, with the understanding that these factors are dynamic and change over time. Physicians familiar with open notes suggested it is important to be thoughtful and deliberate about whether to include demographic factors, epidemiologic factors, and medical and social history in the opening sentences of a note, given the effect such information may have on framing how a patient, family member, or other clinician reads the note.

Avoid Abbreviations and Acronyms, Especially if not Officially Approved by the Practice

Participants in all groups felt that medical shorthand can cause confusion and misinterpretation by patients. Physician, resident, and educator participants noted that while some abbreviations are widely understood within health care, others are interpreted differently within and across a given specialty. They suggested that this, in turn, could lead to clinician confusion, thereby further amplifying the risk of patient confusion (Table 1). Physician participants noted that learning to use medical shorthand in notes was akin to being inducted into a "secret society" (Table 1). Illustrative examples suggested by the 4 focus groups included: "SOB" (shortness of breath), "F/U" (follow-up), "ISO" (in the setting of), "NAEON" (no acute events overnight), and "MOP" (mother of the patient). In addition, physician participants pointed out that shorthand for certain medical diagnoses was problematic, given the frequent lack of clarity. They cited "HFrEF" (heart failure with reduced ejection fraction); "HFpEF" (heart failure with preserved ejection fraction); and "AKI" (acute kidney injury). Several residents discussed the risk that abbreviations and acronyms may perpetuate discriminatory biases within notes. Medical educator participants noted that this was an area in which practicing clinicians can learn from students, given that many have not yet been inducted into contemporary systems of

XSL•FO
RenderX

medical vernacular. Finally, the medical educators noted that even if abbreviations were minimized, notes still may not be completely understandable, underscoring the need for open communication with patients.

Say What You Write and Write What You Say

Patients universally stressed the importance of notes accurately reflecting what was done and discussed during a visit. Residents and physicians emphasized that clinicians should be mindful about not documenting issues, such as possible diagnostics considerations, that were not discussed directly with the patient, and that this principle should be used consistently for all documentation (Table 1). Medical educators identified the importance of guiding learners carefully in interactions between documenting explicit clinical reasoning with robust documentation of differential diagnoses and setting expectations for patients. They suggested that clinicians point out to their patients that some notes are intended to be comprehensive and may include diagnostic possibilities that are unlikely, but nevertheless important to record. In contrast, patients suggested there are situations in which clinicians should not write too much about a potentially sensitive topic that may trigger a harmful response in patients. Instead, patients proposed clinicians might record that a discussion of a "challenging topic" took place during the visit; although details might be excluded, such reference would remind the patient of the conversation. On the other hand, several patient participants noted that accurate and detailed representations of a visit may help empower patients with managing their care, strengthen their agency, and enhance their understanding of medical recommendations.

Verify Past History Information Before Recording it in the Note

Another theme was the importance of verifying patients' past information in the note to avoid "note bloat" (ie, copy and paste from previous notes). Participants from all groups identified "cutting and pasting" as increasing the risk of mistakes. Notably, all patients agreed that the "note bloat" phenomenon can have a negative impact on how patients view their notes, and possibly on their relationship with clinicians, resulting in their feeling the need to advocate for themselves to ensure accurate representation of their stories (Table 1). Several patient participants suggested that copying over all aspects of the social history when not relevant to a specific visit might trigger trauma since this section of the note can replicate and reiterate sensitive information. Two residents noted the issue of "copy-forward" (copying previous information from the record into new notes), thereby propagating bias and stigma, and potentially risking mistrust of the clinician. Medical educators noted that untruthful documentation, at times propagated by copy-and-paste templates, can lead to mistrust in the patient-physician relationship and adversely impact care.

Avoid Words That May Convey Bias or Judgment

With striking overlap across all 4 groups, participants offered many examples of words and phrases that may reflect bias or unfair judgment of patients. Suggestions about common phrases and words to avoid included clinical vernacular such as "complains," "denies," "claims," "refuses," and "endorses." Participants recommended neutral and judgment-free words such as "says," "reports," "does not report," "concerns," and "did not tolerate." Several medical educators suggested renaming "chief complaint" as "chief concern." Participants also discussed problems with labels, such as "poor historian," "noncompliant," "difficult patient," "nonadherent," "uncooperative," "having poor health literacy," and "left AMA." Patient participants agreed on the importance of describing explanations for observed behaviors, rather than using judgmental language, such as, "Mr. X is unable to take his insulin most days of the week due to inability to refrigerate the medication at his place of work." Consideration about how to document race was another common concern. Given that race is without clear relevance to most medical concerns, many believed it should not be included in the written documentation. Finally, the pros and cons of using quotation marks and the patients' own language were discussed. Participants noted that doing so would accurately capture what was said. However, in certain contexts, this might be interpreted as sarcastic, questioning, or making light of what the patient communicated, and participants recommended that quotation marks should be used carefully.

Keep Descriptions of Physical Examinations Objective

Another emergent theme was ensuring that descriptions of physical examinations avoided phrases that could be perceived as offensive. Examples included: "disheveled," "older than stated age," "obese"; and judgmental descriptors, such as "argumentative." Participants recommended omitting descriptors such as "pleasant" or "delightful," in part because the absence of such terminology might inadvertently suggest to the reader that the patient is "unpleasant." Several participants urged considering whether language pertaining to appearance is helpful. All recommended avoiding the words "obese" or "morbidly obese" and to first consider whether the information is relevant to an issue at hand. If so, they recommended quantitative descriptors, for example, BMI or the actual weight.

Empower Patients With Encouraging Words and Clear Next Steps

Participants suggested patients often benefit from reading encouraging words and the next steps regarding their conditions. Physicians noted that empowering language, particularly in the "Assessment and Plan" section of a note, could stimulate patients to engage in their care more actively. One physician suggested adding the word "yet" at the end of the sentence or phrase, signifying progress with a specific goal (Table 1). Patients agreed that empowering notes could facilitate partnership with the medical team. Medical educators suggested adapting notes to ensure they capture expert or emerging clinical reasoning, while still engaging patients in the next steps of a plan. Relatedly, several participants suggested using notes to embed reminders to patients, their care partners, and other members of their care team by stating explicitly what was recommended during the visit, especially with respect to specific next steps and goals.

XSL•FO
RenderX

Pay Close Attention to Sensitive Topics, Including but not Limited to Sexual History, Trauma History, Substance Use History, Mental Health, or Illness

Another emergent theme was mindfulness in documenting topics such as substance use history, mental health and illness, gender identity, sexual identity and health, history of trauma, disagreements between patients and clinicians, significant illnesses, and comprehensive diagnostic reasoning. Medical educators noted the challenges of teaching early learners how best to document such information in the social history section. Physicians noted the importance of documenting any disagreements with patients in an objective and respectful manner.

Write From Your Perspective

Medical notes, in particular the "Assessment and Plan" portion of the note, reflect the perspective of the health care professional at a given point in time. Participants uniformly noted that the patient's perspective should also be included, and that descriptions should be factual and based on direct observations. Physician and resident participants alike suggested the use of the first-person pronoun "I" when writing the "Assessment and Plan," as well as using the phrase "at this point in time." Physicians advised that this wording signals that diagnostic considerations may change as a situation evolves.

Discussion

Principal Results

At a time of increasingly open and transparent communication between patients and clinicians, a growing body of research demonstrates the importance of words and phrases used in clinical notes, and the risk of bias and offense to patients. Patients who read their notes feel more involved in their care, better prepared for visits, and are more likely to follow their clinician's recommendations [4-10]. Just as importantly, patients can experience negative effects from their notes due to language perceived as judgmental, offensive, or stigmatizing [11-14]. Language in notes can also negatively influence other clinicians reading the note, leading to bias and impact on the patient's care [11]. To our knowledge, this is the first study attempting to define concrete guiding principles on best practices in patient-centered documentation. Informed by a qualitative analysis of active and interactive discussion among patients and clinicians, this inquiry, as described in detail above and in Textbox 2 and Table 1, identified 10 discrete guidelines for patient-centered documentation.

This inquiry was stimulated by our interest in facilitating open and transparent communication among health professionals and patients from the very beginning of a clinician's career. Medical school curricula and residency programs are just beginning to introduce the concept of patient-centered documentation, with few providing specific guidance [26-28]. Concrete advice is especially important for early medical learners. Moreover, many faculty and residents who teach medical students are also relatively new to the practice of open notes. Developing and describing guidelines, accompanied by clear and detailed examples, can help faculty both learn what we hope will evolve as best practices and teach students these skills more easily, creating an opportunity for standardized assessment of notes based on a checklist reflecting the guidelines. In addition to providing direct teaching and feedback, the presence of concrete guiding principles will allow faculty to role model these skills in patient interactions for their learners. Furthermore, given the importance and impact of language in documentation, we believe the guidelines identified in this research will serve all clinicians well.

The checklist this study generated has already been implemented within the first-year foundational clinical skills course students undertake at our medical school [19]. Students are introduced to the background and value of open notes, followed by the checklist of guidelines. Over the timeframe of this course and beyond into their core clinical training, students are expected to write patient notes using these guides and are subsequently assessed on the quality of their notes through a rubric based on the checklist. Faculty preceptors in this first-year course are also encouraged to use this checklist, with the aim of reviewing student notes and providing feedback based on these 10 themes.

Limitations

Our study has several limitations. First, owing to time constraints on the duration of the study caused in part by the COVID-19 pandemic, we did not conduct focus groups until data saturation was reached. Second, due to the need to develop this preliminary framework before the delivery of a patient-centered documentation curriculum for students, the data analysis was restricted to 2 reviewers only, limiting the validation of the data. Third, the small number of participants likely influenced the results. Fourth, and relatedly, a fuller spectrum of diversity including geography, clinical background, and health background, could produce more varied views and depth of experience with open notes. Fifth, studies have demonstrated that fewer patients with low income, limited education, or poor English proficiency are active users of the electronic health record, limiting our participants' understanding of the note-reading experience of these populations [29-31]. Our preliminary guidelines may therefore be limited when it comes to generalizing to wider patient populations, and future guidelines should seek to address this concern. Finally, the development of new charting practices and auditing may come with considerable resource demands. Typically, guidelines that assess resource-intensive interventions are more likely to include an assessment of implementation costs, real and intangible. It would also be useful to discuss the potential risks of open notes and any context-specific adjustments for special populations or clinics. This is, however, beyond the scope of this study.

Conclusions

The guidelines identified are preliminary. As our understanding grows, we expect clinicians will learn how to write notes in ways that patients find increasingly useful. In addition, there are nuanced areas to consider in various specialties and different patient populations that may require further adaptation of these guidelines. To this end, clinicians and patient advocates partnering to co-design medical education will be important, as it was in this study. Ideally, in the future, we will learn to teach practitioners effective documentation through a common

XSL•FO
RenderX

language and shared set of standard expectations. Although we expect norms and practice techniques to evolve, our study presents an initial attempt to develop practical and respectful guidelines for patient-centered documentation.

Acknowledgments

We thank the patients, physicians, and medical educators who took the time to share their experiences and perspectives, which helped our work. The authors also thank Alex Duncan who coordinated the logistics of the focus groups. This study is part of a project developing a patient-centered documentation curriculum for medical students and their prime clinical faculty. It is funded by an educational grant from the Josiah Macy Jr Foundation and a generous gift from Kate and Arnold Schmeidler. These funds supported honoraria for the focus group participants, of which there were a total of 27.

Authors' Contributions

AV, CB, KJ, TD, and CD performed conceptualization and visualization. CB and TD handled supervision. AV and KJ conducted data analysis. AV, KJ, and CB performed investigations. AV, KJ, AG, and LS handled project administration. AV and CB conducted writing–original draft preparation. AV, CB, KJ, CD, AG, TD, and LS performed writing–review and editing.

Conflicts of Interest

None declared.

References

1. Delbanco T, Walker J, Darer JD, Elmore JG, Feldman HJ, Leveille SG, et al. Open notes: doctors and patients signing on. Ann Intern Med 2010;153(2):121-125 [FREE Full text] [doi: 10.7326/0003-4819-153-2-201007200-00008] [Medline: 20643992]
2. Hudson KL, Collins FS. The 21st Century Cures Act - a view from the NIH. N Engl J Med 2017;376(2):111-113 [FREE Full text] [doi: 10.1056/NEJMp1615745] [Medline: 27959585]
3. Salmi L, Blease C, Hägglund M, Walker J, DesRoches CM. US policy requires immediate release of records to patients. BMJ 2021;372:n426. [doi: 10.1136/bmj.n426] [Medline: 33602667]
4. DesRoches CM, Leveille S, Bell SK, Dong ZJ, Elmore JG, Fernandez L, et al. The views and experiences of clinicians sharing medical record notes with patients. JAMA Netw Open 2020;3(3):e201753. [doi: 10.1001/jamanetworkopen.2020.1753] [Medline: 32219406]
5. Gerard M, Fossa A, Folcarelli PH, Walker J, Bell SK. What patients value about reading visit notes: a qualitative inquiry of patient experiences with their health information. J Med Internet Res 2017;19(7):e237 [FREE Full text] [doi: 10.2196/jmir.7212] [Medline: 28710055]
6. Gerard M, Chimowitz H, Fossa A, Bourgeois F, Fernandez L, Bell SK. The importance of visit notes on patient portals for engaging less educated or nonwhite patients: survey study. J Med Internet Res 2018;20(5):e191 [FREE Full text] [doi: 10.2196/jmir.9196] [Medline: 29793900]
7. Bell SK, Folcarelli P, Fossa A, Gerard M, Harper M, Leveille S, et al. Tackling ambulatory safety risks through patient engagement: what 10,000 patients and families say about safety-related knowledge, behaviors, and attitudes after reading visit notes. J Patient Saf 2021;17(8):e791-e799. [doi: 10.1097/PTS.0000000000000494] [Medline: 29781979]
8. Walker J, Leveille S, Bell S, Chimowitz H, Dong Z, Elmore JG, et al. OpenNotes after 7 years: patient experiences with ongoing access to their clinicians' outpatient visit notes. J Med Internet Res 2019;21(5):e13876 [FREE Full text] [doi: 10.2196/13876] [Medline: 31066717]
9. Mishra VK, Hoyt RE, Wolver SE, Yoshihashi A, Banas C. Qualitative and quantitative analysis of patients' perceptions of the patient portal experience with openNotes. Appl Clin Inform 2019;10(1):10-18 [FREE Full text] [doi: 10.1055/s-0038-1676588] [Medline: 30602196]
10. Nazi KM, Turvey CL, Klein DM, Hogan TP, Woods SS. VA openNotes: exploring the experiences of early patient adopters with access to clinical notes. J Am Med Inform Assoc 2015;22(2):380-389. [doi: 10.1136/amiajnl-2014-003144] [Medline: 25352570]
11. Wolff JL, Darer JD, Berger A, Clarke D, Green JA, Stametz RA, et al. Inviting patients and care partners to read doctors' notes: OpenNotes and shared access to electronic medical records. J Am Med Inform Assoc 2017;24(e1):e166-e172 [FREE Full text] [doi: 10.1093/jamia/ocw108] [Medline: 27497795]
12. P Goddu A, O'Conor KJ, Lanzkron S, Saheed MO, Saha S, Peek ME, et al. Do words matter? Stigmatizing language and the transmission of bias in the medical record. J Gen Intern Med 2018;33(5):685-691 [FREE Full text] [doi: 10.1007/s11606-017-4289-2] [Medline: 29374357]
13. Gudzune KA, Bennett WL, Cooper LA, Bleich SN. Patients who feel judged about their weight have lower trust in their primary care providers. Patient Educ Couns 2014;97(1):128-131 [FREE Full text] [doi: 10.1016/j.pec.2014.06.019] [Medline: 25049164]

XSL•FO
RenderX

14. Stanford FC, Kyle TK. Respectful language and care in childhood obesity. JAMA Pediatr 2018;172(11):1001-1002 [FREE Full text] [doi: 10.1001/jamapediatrics.2018.1912] [Medline: 30193352]
15. Bell SK, Mejilla R, Anselmo M, Darer JD, Elmore JG, Leveille S, et al. When doctors share visit notes with patients: a study of patient and doctor perceptions of documentation errors, safety opportunities and the patient-doctor relationship. BMJ Qual Saf 2017;26(4):262-270 [FREE Full text] [doi: 10.1136/bmjqs-2015-004697] [Medline: 27193032]
16. Fernández L, Fossa A, Dong Z, Delbanco T, Elmore J, Fitzgerald P, et al. Words matter: what do patients find judgmental or offensive in outpatient notes? J Gen Intern Med 2021;36(9):2571-2578 [FREE Full text] [doi: 10.1007/s11606-020-06432-7] [Medline: 33528782]
17. Beauchamp T, Childress J. Principles of biomedical ethics: marking its fortieth anniversary. Am J Bioeth 2019;19(11):9-12. [doi: 10.1080/15265161.2019.1665402] [Medline: 31647760]
18. Blease C, McMillan B, Salmi L, Davidge G, Delbanco T. Adapting to transparent medical records: international experience with "open notes". BMJ 2022;379:e069861. [doi: 10.1136/bmj-2021-069861] [Medline: 36410770]
19. Eng K, Johnston K, Cerda I, Kadakia K, Mosier-Mills A, Vanka A. A patient-centered documentation skills curriculum for preclerkship medical students in an open notes era. MedEdPORTAL 2024;20:11392 [FREE Full text] [doi: 10.15766/mep_2374-8265.11392] [Medline: 38533390]
20. McLafferty I. Focus group interviews as a data collecting strategy. J Adv Nurs 2004;48(2):187-194. [doi: 10.1111/j.1365-2648.2004.03186.x] [Medline: 15369499]
21. Morgan DL, Core SRM. Focus Groups as Qualitative Research. Thousand Oaks, CA: Sage Publications; 1997.
22. Carlsen B, Glenton C. What about N? A methodological study of sample-size reporting in focus group studies. BMC Med Res Methodol 2011;11:26 [FREE Full text] [doi: 10.1186/1471-2288-11-26] [Medline: 21396104]
23. Mayring P. Qualitative Content Analysis: Theoretical Foundation, Basic Procedures and Software Solution. Klagenfurth, Austria: Beltz; 2014.
24. Marks D, Yardley L. Research Methods for Clinical and Health Psychology. Thousand Oaks, CA: Sage; 2004.
25. Lester JN, Cho Y, Lochmiller CR. Learning to do qualitative data analysis: a starting point. Hum Resour Dev Rev 2020;19(1):94-106 [FREE Full text] [doi: 10.1177/1534484320903890]
26. Crotty BH, Anselmo M, Clarke D, Elmore JG, Famiglio LM, Fossa A, et al. Open notes in teaching clinics: a multisite survey of residents to identify anticipated attitudes and guidance for programs. J Grad Med Educ 2018;10(3):292-300 [FREE Full text] [doi: 10.4300/JGME-D-17-00486.1] [Medline: 29946386]
27. Niedermier VE. Teaching electronic health record documentation to medical students. J Grad Med Educ 2017;9(1):135 [FREE Full text] [doi: 10.4300/JGME-D-16-00628.1] [Medline: 28261413]
28. Gagliardi JP, Turner DA. The electronic health record and education: rethinking optimization. J Grad Med Educ 2016 Jul;8(3):325-327 [FREE Full text] [doi: 10.4300/JGME-D-15-00275.1] [Medline: 27413432]
29. Lyles C, Schillinger D, Sarkar U. Connecting the dots: health information technology expansion and health disparities. PLoS Med 2015;12(7):e1001852 [FREE Full text] [doi: 10.1371/journal.pmed.1001852] [Medline: 26172977]
30. Lyles CR, Sarkar U. Health literacy, vulnerable patients, and health information technology use: where do we go from here? J Gen Intern Med 2015;30(3):271-272 [FREE Full text] [doi: 10.1007/s11606-014-3166-5] [Medline: 25588688]
31. Tieu L, Schillinger D, Sarkar U, Hoskote M, Hahn KJ, Ratanawongsa N, et al. Online patient websites for electronic health record access among vulnerable populations: portals to nowhere? J Am Med Inform Assoc 2017;24(e1):e47-e54 [FREE Full text] [doi: 10.1093/jamia/ocw098] [Medline: 27402138]

Abbreviations

AKI: acute kidney injury
HFpEF: heart failure with preserved ejection fraction
HFrEF: heart failure with reduced ejection fraction

Edited by B Lesselroth; submitted 08.04.24; peer-reviewed by K Gray, L Shapiro, M Figueroa Gray, S Walker; comments to author 25.07.24; revised version received 11.09.24; accepted 23.11.24; published 20.01.25.

Please cite as:
Vanka A, Johnston KT, Delbanco T, DesRoches CM, Garcia A, Salmi L, Blease C
Guidelines for Patient-Centered Documentation in the Era of Open Notes: Qualitative Study
JMIR Med Educ 2025;11:e59301
URL: https://mededu.jmir.org/2025/1/e59301
doi:10.2196/59301
PMID:

XSL•FO
RenderX

©Anita Vanka, Katherine T Johnston, Tom Delbanco, Catherine M DesRoches, Annalays Garcia, Liz Salmi, Charlotte Blease. Originally published in JMIR Medical Education (https://mededu.jmir.org), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Implementing the H&P 360 in Three Medical Institutions: Usability Study

Rupinder Hayer[1*], MPH; Joyce Tang[2*], MPH, MD; Julia Bisschops[3*], MD; Gregory W Schneider[4*], MD; Kate Kirley[1*], MS, MD; Tamkeen Khan[1*], MA, PhD; Erin Rieger[5*], MD; Eric Walford[6*], MD; Irsk Anderson[2*], MD; Valerie Press[2*], MPH, MD; Brent Williams[6*], MD

[1]American Medical Association, Improving Health Outcomes, Chicago, IL, United States
[2]Section of Hospital Medicine, University of Chicago Pritzker School of Medicine, Chicago, IL, United States
[3]Department of Humanities, Health and Society, Herbert Wertheim College of Medicine, Florida International University., Miami, FL, United States
[4]Florida International University, Miami, FL, United States
[5]Columbia University Medical Center, New York, NY, United States
[6]University of Michigan, Ann Arbor, MI, United States
[*]all authors contributed equally

Corresponding Author:
Rupinder Hayer, MPH
American Medical Association
Improving Health Outcomes
330 North Wabash Avenue
Chicago, IL, 60611
United States
Phone: 1 6308499232
Email: rupinder.hayer@ama-assn.org

Abstract

Background: The traditional history and physical (H&P) provides the basis for physicians' data gathering, problem formulation, and care planning, yet it can miss relevant behavioral or social risk factors. The American Medical Association's "H&P 360," a modified H&P, has been shown to foster information gathering and patient rapport in inpatient settings and objective structured clinical examinations. It prompts students to explore 7 domains, as appropriate to the clinical context: biomedical problems, psychosocial problems, patients' priorities and goals, behavioral history, relationships, living environment and resources, and functional status.

Objective: This study aims to examine the perceived usability of the H&P 360 outside standardized patient settings.

Methods: The H&P 360 was implemented in various clinical settings across 3 institutions. Of the 207 student participants, 18 were preclerkship, 126 were clerkship, and 63 were postclerkship; 3-8 months after implementation, we administered a student survey consisting of 14 Likert-type items (1=strongly disagree to 5=strongly agree) and 3 free-text response items to assess usability.

Results: Of the 207 students, 61 responded to the survey (response rate was 29.5%). Among all students, mean ratings on the 3 usability survey items ranged from 4.03 to 4.24. The 5 items assessing the impact on patient care had mean ratings ranging from 3.88 to 4.24. The mean ratings for the 2 student learning items were 4.10 and 4.16. Students' open-ended comments were generally positive, expressing a perceived value in obtaining a more complete contextual picture of patients' conditions and supporting the usability of the H&P 360. Survey response patterns varied across institutions and learner levels.

Conclusions: Our findings suggest that using the H&P 360 may enhance information gathering critical for chronic disease management, particularly regarding social drivers of health. As a potential new standard, the H&P 360 may have clinical usability for identifying and addressing health inequities. Future work should assess its effects on patient care and outcomes.

(JMIR Med Educ 2025;11:e66221) doi:10.2196/66221

KEYWORDS

history and physical; medical education; social drivers; social determinants of health

XSL•FO
RenderX

Introduction

The traditional history and physical (H&P) structure is central to the patient-physician interaction and remains a foundational element of medical education. Through medical history, physicians elicit 60%-80% of the information relevant to diagnosis and treatment [1]. Medical students are typically required to master the skill of gathering, synthesizing, and documenting patient information early in their training. The traditional H&P, used in most medical education settings and routine clinical practice, is primarily structured to diagnose acute medical conditions and has not evolved for generations—despite the growing prevalence of chronic diseases and the increasing influence of social and behavioral drivers of health [2,3].

The social determinants or social drivers of health (SDOH) heavily influence the health of patients and populations [4]. The World Health Organization (WHO), in its conceptual framework for action on SDOH, defines "social determinants of health" as the full set of social conditions in which people live and work [4,5]. We will use the phrases social determinants and social drivers interchangeably. We will refer to social risk factors, meanwhile, as individual-level adverse SDOH, such as housing instability or low education level, and social needs as social factors that take into account people's individual preferences and priorities in identifying and guiding social interventions [6,7].

Health systems and providers are increasingly exploring ways to better integrate health care delivery with reforms aimed at addressing the SDOH, identifying patients' social risk factors, and meeting patients' social needs [8,9]. At the upstream level, laws, policies, and regulations can be used to create community conditions that foster health. At the midstream level, providers and health systems can include screening questions to identify social risk factors and offer services that connect patients to resources to meet their social needs. At the downstream level, clinicians can tailor medical interventions to acknowledge individual social conditions [6,8,9]. To significantly improve the health of all, it is critical to emphasize addressing the broader SDOH inequities—those created and sustained by structural racism and the marginalization of specific groups, including women, Black, Hispanic/Latino, LGBTQ+ (lesbian, gay, bisexual, transgender, queer or questioning, and other diverse sexual orientations and gender identities) individuals, people living with disabilities, and other populations [4,5,10-12].

In addition to increased attention to the SDOH, both globally and particularly in the United States, chronic diseases are the largest contributors to disease burden, accounting for 90% of health care costs in the United States [13]. Health systems are increasingly addressing upstream factors, such as behavioral health, and social and environmental circumstances, to prevent and manage chronic diseases and their consequences [14,15]. Against this multidimensional backdrop, individual clinicians, residents, and medical students typically rely on the H&P examination as their principal method of information gathering. However, when using the traditional H&P to frame and organize this process, learners at all levels are not prompted to collect relevant biopsychosocial data, including social needs and health behaviors, which are key to preventing and managing chronic diseases. Recent research has shown that inaccurate and incomplete patient histories are among the leading causes of diagnostic errors [16].

The H&P 360, a modified version of the traditional H&P, was developed by the American Medical Association's (AMA) Chronic Disease Prevention and Management interest group in May 2017, building on earlier work by medical educators at the University of Michigan (UM) [12,17]. This new approach was designed to more explicitly acknowledge the SDOH, the prevalence of chronic diseases, and the importance of patients' preferences and priorities in clinical decision-making. It intentionally incorporates the WHO conceptual framework for addressing the SDOH at the micro-level of individual interaction [4,5], the Centers for Disease Control and Prevention (CDC) framework for addressing chronic diseases at the health system and community-clinic levels [12,18], and contemporary models of shared decision-making [1,19].

The H&P 360 is grounded in the idea that the central, standardized written template in medical education (ie, the traditional H&P) plays a significant role in both enhancing and constraining information-seeking related to medical decision-making. This understanding of information gathering aligns with Structuration Theory and Cognitive Load Theory. Structuration Theory posits that social practice both shapes and is shaped by the structures, such as learning templates, within which it occurs [20]. Cognitive Load Theory, meanwhile, asserts that cognitive capacity is limited and that learning is enhanced when key information is presented in manageable blocks, such as the 7 domains [21].

At a deeper level, the H&P may play an important role in shaping physicians' professional identity and role expectations, as suggested by Social Learning Theory [22,23], which posits that individuals' agency and role identities are critically influenced by the social and institutional contexts in which they develop. The H&P 360 prompts students to collect relevant biopsychosocial information, particularly social risk factors and needs, using a systematic yet flexible framework. While it retains the basic structure of the traditional H&P for eliciting biomedical information, the H&P 360 also includes general prompts for 6 additional domains: patients' priorities and goals, psychosocial problems, behavioral history, relationships, living environment and resources, and functional status. The 6 nonbiomedical domains were identified through a literature review as those consistently represented in comprehensive clinical assessment settings, including geriatrics, and care for homeless and chronically mentally ill persons, as well as in the categories of the Diagnostic and Statistical Manual of Mental Disorders (4th edition) [24], and have been applied in numerous clinical and teaching settings since 2010 [17]. See Multimedia Appendix 1 for the H&P 360 template.

When using the H&P 360, students are encouraged to ask a few questions from each of the domains as part of the standard history. These additional questions help students gain a more comprehensive understanding of a patient's biopsychosocial condition and support the development of an appropriate

XSL•FO
RenderX

treatment and management plan. In follow-up encounters, continued exploration of the 7 H&P 360 domains can foster a deeper understanding of the patient, informing chronic disease management. A previous randomized trial conducted at 4 medical schools found that medical students using the H&P 360 in a standardized patient setting collected significantly more biopsychosocial information compared with students using the traditional H&P [3]. Another study found that students who applied the H&P 360 using templated notes in the electronic health record reported improved elicitation of patient goals and perspectives, as well as identification of contextual factors and patient needs critical to preventing rehospitalization [25]. In addition to enhanced data gathering, the H&P 360 has been shown to encourage multidisciplinary team care planning [17] and to improve patient rapport (unpublished data).

The goal of this study was to examine the perceived usability of the H&P 360 by both faculty and students across a variety of clinical settings and learner levels in routine clinical teaching contexts. To assess usability in different clinical teaching environments, the AMA launched a grant program for institutions willing to implement the H&P 360 in student clinical encounters and administer a standardized postintervention survey to faculty and students across sites. We hypothesized that students and faculty across sites would appreciate the usability of the approach, but that barriers to engagement would vary by site and learner level.

Methods

Site Selection

The AMA offered funding for projects to implement the H&P 360 within clinical settings at academic institutions. Priority was given to projects aimed at developing additional supporting materials and gathering student and faculty feedback during the implementation phase. Following a call for proposals, 4 academic institutions received grants from the AMA to implement the H&P 360 across a diverse range of clinical settings and undergraduate medical education learner levels. The grant period began in January 2020 and ended in June 2021. Because of the pandemic, only 3 of the institutions were able to implement their grants. These institutions were the UM School of Medicine, the University of Chicago Pritzker School of Medicine (UC), and the Herbert Wertheim College of Medicine at Florida International University (FIU). The fourth institution was unable to implement its grant project but still incorporated the H&P 360 with its students.

The 3 grant-funded institutions implemented the H&P 360 across a variety of clinical settings and learner levels, described in detail in Table 1. Clinical settings included inpatient, outpatient, virtual, community-based clinics, and longitudinal outpatient clinics. Learner levels ranged from preclerkship to clerkship and postclerkship. The approaches each school used to introduce students and faculty to the H&P 360 are also detailed in Table 1. At all 3 sites, students were introduced to the H&P 360 through a nonstandardized 1- to 2-hour seminar. One site (UC) implemented standardized note templates within the electronic health record to facilitate documentation of the H&P 360. Faculty orientation to the H&P 360 varied across sites, ranging from emailed communications with attached introductory materials and a teaching guide (UM and UC) to virtual orientation sessions over Zoom (Zoom Communications, Inc) and some in-person sessions (FIU and UC).

XSL•FO
RenderX

Table 1. Learner level, clinical setting, and teaching context by institution.

Approaches to introducing H&P[a] 360	Learner level			Students, n	Setting	Duration of use	How the H&P 360 was introduced to students and topics covered	How the H&P 360 was introduced to faculty and topics covered
	Preclerkship	Clerkship	Postclerkship					
The University of Chicago Pritzker School of Medicine	✓	N/A[b]	N/A	18	• Preclerkship students were encouraged to utilize the H&P 360 in a longitudinal patient-partnered clinical experience (about 6 face-to-face and 2 virtual clinical sessions).	9 months	Preclerkship students attended a presentation on the format and components and received training materials including examples and the interview guide.	Faculty precepting preclerkship students attended a presentation or received an email.
The University of Chicago Pritzker School of Medicine	N/A	✓	N/A	8	• Clerkship students were encouraged to utilize the H&P 360 during COVID-19 follow-up virtual visits.	1 month	Clerkship students attended a 1-hour virtual training session with background about the H&P 360 and details on using a COVID-19–specific note template. The supporting interview guide and pocket card were leveraged as needed.	Faculty attended 2 or more presentations on the H&P 360.
The University of Chicago Pritzker School of Medicine	N/A	N/A	✓	24	• Postclerkship students were encouraged to utilize the H&P 360 in an internal medicine subinternship for 1 admission per call cycle.	1 month	Postclerkship students received an email from their course director with background about the H&P 360 and instructions to access H&P 360 templates. The supporting interview guide and pocket card were leveraged as needed.	Faculty supervising postclerkship students received an email.
University of Michigan School of Medicine	N/A	N/A	✓	39	• All students utilized the H&P 360 in an outpatient setting. • Of the 39 students, 8 used it in a community-based elective and 31 used it in a longitudinal clinic setting. • Students were encouraged to apply the H&P 360 in every encounter.	Elective (1 month) and longitudinal clinics (9 months)	A 2-hour interactive in-person seminar with case examples. The supporting interview guide and pocket card were leveraged as needed.	Email introduction and follow-up, which included teaching tips, pocket cards, profiles, and cases.

XSL•FO
RenderX

Approaches to introducing H&P[a] 360	Learner level			Students, n	Setting	Duration of use	How the H&P 360 was introduced to students and topics covered	How the H&P 360 was introduced to faculty and topics covered
	Preclerkship	Clerkship	Postclerkship					
Herbert Wertheim College of Medicine at Florida International University	N/A	✓	N/A	118	• All students utilized the H&P 360 in a virtual and longitudinal, interprofessional, home-based service-learning program.	12 months/1 academic year as part of a longitudinal program.	Introduced during an interactive didactic session on chronic disease management; the self-directed video was also available for students.	One in-person faculty development session before COVID-19; a self-directed video for faculty; and a conference presentation by Dr. Brent Williams from the University of Michigan.

[a]H&P: history and physical.

[b]N/A: not applicable.

Settings

At FIU, the investigators had to alter their initial implementation strategy due to the COVID-19 pandemic. The project team completed 1 in-person faculty orientation; subsequent sessions were delivered virtually as an online module tailored to both faculty and students. For students, the team relied solely on the online module, as the planned in-person session was canceled due to the pandemic.

At UC, initial implementation plans were also disrupted by the COVID-19 pandemic, which led to medical students being removed from traditional clinical settings. During this period, an innovative program was developed in which clerkship students conducted phone outreach to patients newly testing positive for COVID-19. Taking advantage of this novel opportunity, the H&P 360 was used to help structure these outreach calls. Because of the small number of students and faculty involved, an in-depth, interactive training program for both faculty and students was offered via Zoom meetings. In the later implementation of the H&P 360 for preclerkship students, a virtual orientation over Zoom was incorporated into their Clinical Skills course. Because of the large number of faculty serving as preceptors for this course, only new faculty preceptors—who were required to attend a mandatory orientation session—received virtual training on the H&P 360 framework. Preceptors who were not new to the program and for whom orientation attendance was not required received emailed communication about the H&P 360. Implementation for postclerkship students was further modified to email communication only. These students had rolling start times each month, and there was no formal orientation session during the clerkship to integrate separate training. The number of clinical faculty preceptors for postclerkship students was quite large, also with rolling start times every 2 weeks, making email communication regarding the H&P 360 the most feasible approach.

At UM, the H&P 360 was implemented during the postclerkship period in 2 settings: a 1-month clinical elective focused on underserved populations before the pandemic (8 students over 2 months), taught by 1 of the authors (BW), and longitudinal weekly clinics in primary care settings over 9 months during the pandemic (31 students). The longitudinal clinic rotation was chosen due to the faculty coordinator's interest in implementing the H&P 360 and its suitability for continuity settings, where the domains can be explored with patients over time. For the longitudinal rotation, students received a 2-hour introduction to the H&P 360, including case examples. Precepting faculty were sent introductory materials, teaching tips, and written case examples via email both at the start and several months into the longitudinal clinics. Many longitudinal clinics transitioned to telemedicine visits during the pandemic. In both rotations, students were encouraged—but not required—to use the H&P 360, or portions of it, in every encounter.

Survey Structure

Data collection consisted of a student survey on using the H&P 360 in undergraduate medical education settings. As the survey focused specifically on the use of the H&P 360, previously published surveys were not applicable. Theoretical frameworks guiding survey development included Bloom's taxonomy of learning objectives [26], which emphasizes synthesis and application of knowledge rather than factual recall, and the Expectancy-Value Theory of Motivation [27], which posits that learner motivation is influenced by the perceived value of new information.

The survey consisted of an initial section asking for examples of a question relevant to each of the 5 domains from the H&P 360, followed by 14 Likert-type items (response scale: 1=strongly disagree to 5=strongly agree) and 3 open-ended questions. The Likert-type items were developed using a "blueprint" of 7 potential impact areas of the H&P 360, designed by the authors. Source items were either adapted from a 10-item version used in a previous study [17] or newly created. To minimize the response burden, the survey was limited to 15

XSL•FO
RenderX

items or fewer. Items were reviewed for sensibility by small groups of medical students and residents not involved in the study, resulting in minor modifications. The final instrument included 14 Likert-type items. Two items were modified or omitted at some sites and thus were not included in the analyses. The analyzed items are shown in Tables 2 and 3. The 7 areas from the "blueprint" and their corresponding item numbers were perceived usability of the H&P 360 (items 1, 2, and 3); impact on history-taking (item 4); perceived clinical value added (items 5 and 6); promotion of understanding patients' goals (item 7); enhancement of patient-provider relationships (item 8); facilitation of care planning (item 9); and promotion of inclusion of other health professionals (item 10). Two additional items were included as global measures of educational and clinical value, respectively (items 11 and 12). By covering a broad range of topics, results from individual items could be used independently by educators to inform a wide spectrum of educational and research activities.

The 3 open-ended questions were designed to elicit specific feedback about the H&P 360: "Name two (or more) aspects of the H&P 360 you found helpful"; "Name two (or more) aspects of the H&P 360 you found challenging"; and "What changes would you recommend for the H&P 360?" A systematic review of the open-ended comments is not included in this paper. Instead, a subset of comments reflecting students' perceived value and limitations of the H&P 360 is provided in Multimedia Appendix 2.

Table 2. Mean student survey scores (Likert scale 1-5)[a] by school.

Student survey scores by school	All students (N=49) (N=61)[b], mean (SD)	All FIU[c] students (N=17), mean (SD)	All UM[d] students (N=13), mean (SD)	All UC[e] students (N=19) (N=31)[b], mean (SD)
Usability				
1. The H&P[f] 360 was easy to use	4.12 (0.78)	4.18 (0.64)	4.15 (0.69)	4.05 (0.97)
2. Elements of the H&P 360 are potentially useful in all patient interactions	4.24 (0.90)	4.35 (0.61)	4.62 (0.51)	3.89 (1.20)
3. I plan to use the H&P 360 during other rotations[b]	4.03 (0.84)	3.59 (0.94)	4.31 (0.75)	4.16 (0.73)
Impact on patient care				
4. The H&P 360 changed some of the questions I ask patients during the encounter	4.08 (0.76)	3.59 (0.87)	4.46 (0.52)	4.26 (0.56)
5. The H&P 360 helped create a more comprehensive problem list	3.88 (0.95)	4.06 (0.90)	4.15 (0.69)	3.53 (1.07)
6. The H&P 360 added valuable information that I would not otherwise know about the patient[b]	4.18 (0.79)	3.82 (1.01)	4.46 (0.78)	4.26 (0.58)
7. The H&P 360 helped me better understand patients' goals[b]	4.15 (0.68)	4.12 (0.78)	4.23 (0.44)	4.13 (0.72)
8. Using the H&P 360 facilitated a stronger provider-patient relationship[b]	4.24 (0.67)	4.12 (0.70)	4.08 (0.64)	4.35 (0.66)
9. I was able to develop management plans that incorporated information from the H&P 360	3.88 (0.83)	4.00 (0.87)	4.00 (0.41)	3.68 (1.00)
Overall impact on student learning				
10. The H&P 360 helped me learn to be a better clinician[b]	4.16 (0.64)	4.06 (0.66)	4.31 (0.63)	4.16 (0.64)
11. The H&P 360 helped improve the care I provided to my patients	4.10 (0.71)	4.06 (0.75)	4.31 (0.63)	4.00 (0.75)

[a]1=strongly disagree to 5=strongly agree.

[b]These are items with a greater number of respondents because an abbreviated version of the survey was completed by preclinical students at UC.

[c]FIU: Florida International University.

[d]UM: University of Michigan.

[e]UC: University of Chicago.

[f]H&P: history and physical.

XSL•FO
RenderX

Table 3. Mean student survey scores (Likert scale 1-5)[a] by clerkship status.

Mean student survey scores	Preclerkship students (N=15), mean (SD)	Clerkship students (N=25), mean (SD)	Postclerkship students (N=24), mean (SD)
Usability			
1. The H&P[b] 360 was easy to use	N/A[c]	4.32 (0.63)	3.92 (0.88)
2. Elements of the H&P 360 are potentially useful in all patient interactions	N/A	4.40 (0.76)	4.08 (1.02)
3. I plan to use the H&P 360 during other rotations	4.33 (0.49)	3.84 (0.90)	4.08 (0.88)
Impact on patient care			
4. The H&P 360 changed some of the questions I ask patients during the encounter	N/A	3.92 (0.91)	4.25 (0.53)
5. The H&P 360 helped create a more comprehensive problem list	N/A	4.08 (0.86)	3.67 (1.21)
6. The H&P 360 added valuable information that I would not otherwise know about the patient	4.25 (0.45)	4.04 (0.93)	4.29 (0.75)
7. The H&P 360 helped me better understand patients' goals	4.25 (0.45)	4.12 (0.83)	4.13 (0.61)
8. Using the H&P 360 facilitated a stronger provider-patient relationship	4.17 (0.58)	4.36 (0.70)	4.13 (0.68)
9. I was able to develop management plans that incorporated information from the H&P 360	N/A	4.08 (0.81)	3.67 (0.82)
10. The H&P 360 facilitated care planning that included other health professionals	3.75 (0.97)	4.08 (0.86)	4.00 (1.02)
Overall impact on student learning			
11. The H&P 360 helped me learn to be a better clinician	4.25 (0.45)	4.20 (0.65)	4.08 (0.12)
12. The H&P 360 helped improve the care I provided to my patients	N/A	4.20 (0.71)	4.00 (0.72)

[a]1=strongly disagree to 5=strongly agree.

[b]H&P: history and physical.

[c]N/A: not applicable.

The survey was administered online by all 3 sites approximately 3-8 months after implementation. Six items that presumed experience in clinical care were not administered to preclinical students participating in this study; this omission applied only to a subset of students at 1 site. Data were aggregated across all sites to calculate mean scores and SDs for each survey item, allowing comparisons by institution and by clerkship status. Because of the small number of respondents in each subgroup, we were limited to analyzing descriptive statistics and were unable to conduct psychometric analyses or hypothesis testing to statistically compare subgroups. However, the descriptive analysis was still useful for aggregating data across multiple sites and generating hypotheses. Data analysis was conducted using STATA version 13.0 (StataCorp). See Multimedia Appendix 3 for the supporting CHERRIES (Checklist for Reporting Results of Internet E-Surveys) document.

Ethics Considerations

The UM received exempt institutional review board status from the Institutional Review Boards of the UM Medical campus. FIU received exempt institutional review board status from The FIU Office of Research Integrity. The UC received exempt institutional review board status from the BSD/UCMC Institutional Review Boards at the UC. Lastly, the AMA confirmed that this study was not deemed to be research by the University of Illinois Chicago Institutional Review Board. All 4 institutions confirmed that all methods were carried out in accordance with relevant guidelines and regulations.

Results

Summary of the Survey Findings

The Likert-type survey items were organized by consensus among the authors into 3 sets to identify patterns and facilitate discussion: Usability (3 items); Impact on Patient Care (7 items); and Overall Impact on Student Learning (2 items). Results are presented for all student respondents by institutional site in Table 2 and by learner-level subgroups in Table 3. Of the 207 students, 61 (29.5%) responded to the survey. Institutional response rates were as follows: FIU, 17 out of 118 (14.4%) students; UM, 13 out of 39 (33.3%) students; and UC, 31 out of 50 (62.0%) students.

Among all students, mean ratings on the 3 survey items related to usability (ease of use, use in all encounters, and intention to use in other rotations) were high, with mean (SD) scores ranging from 4.03 (0.84) to 4.24 (0.90) (Table 2). Some students' comments suggested that efficiently using the H&P 360 requires practice. One postclerkship student commented: "(The H&P 360)...is quite long so it was challenging to hit aspects of each domain while attempting to time manage. However, hitting 1-item from each domain, chosen on a case-by-case basis, seems

XSL•FO
RenderX

quite doable." Several students raised concerns about the awkwardness of asking some questions, particularly during virtual outreach calls to patients who had newly tested positive for COVID-19. One clerkship student commented: "It did not always feel natural to fit into the conversation with every patient. Some were not very open to conversation, which is understandable since we were strangers calling them out of the blue."

Students also found that the H&P 360 positively affected patient care by expanding the range of information available for clinical decision-making and promoting stronger patient-clinician relationships. Mean ratings across the 5 related items ranged from 3.88 (SD 0.95) to 4.24 (SD 0.67; see Table 3 for details). Student feedback on clinical impact emphasized the benefits of the H&P 360 in building rapport. One clerkship student commented: "It helped me build rapport with my patient and have a better [understanding] of their life and how it affects their health." Others mentioned that the H&P 360 helped build trust and identify high-risk situations. See Multimedia Appendix 2 for additional relevant student comments.

Students also found that the H&P 360 facilitated their learning and development as clinicians, with mean ratings of 4.10 (SD 0.71) and 4.16 (SD 0.64) for the items "the H&P 360 helped me...improve the care I provided to my patients" and "…be a better clinician," respectively. One postclerkship student commented: "The H&P 360 was helpful in...[r]eturning the humanity to medicine: patients are people first—Helping to understand some of the barriers to health and disease prevention that might not otherwise be apparent."

Site-Specific Survey Findings

Some variation in student survey responses was observed across institutions (Table 2). For 2 items related to using the H&P 360 to develop problem lists and management plans, student ratings at UC were lower than those at UM or FIU. For 3 items—related to using the H&P 360 in other rotations and its role in changing some questions and adding valuable information—student ratings at FIU were lower compared with UC and UM.

Survey Findings by Learner Levels

Across learner levels, some variation in student survey responses was noted for a minority of items (Table 3). For example, preclinical students gave relatively low ratings on the item related to facilitating care planning that included other health professionals compared with their responses on other items. While clerkship students valued the H&P 360 in all patient interactions and for facilitating stronger patient relationships, their ratings were relatively low for items related to the H&P 360 changing the questions they asked and their plans to use it in future rotations. Postclerkship students gave high ratings for 9 of the 12 questions. Lower ratings were observed for items related to ease of use, creating a more comprehensive problem list, and the ability to develop management plans incorporating information from the H&P 360. The phrasing of the item on time burden evolved over time and was therefore not administered consistently across or within institutions.

Discussion

Principal Findings

Previous work has documented the advantages of the H&P 360 over the traditional H&P during single inpatient encounters [17] and with standardized patients [3]. This study examined the use of the H&P 360 across a broad range of routine, longitudinal clinical teaching settings. Medical students at 3 institutions, spanning different levels of training and diverse ambulatory, inpatient, community, and virtual settings, found the H&P 360 useful and reported a positive impact on patient care and their own learning. The perceived benefits of the H&P 360 include helping students gather relevant information on patients' goals and circumstances, as well as potential barriers and facilitators of health. It also enhances patient-provider relationships and encourages interprofessional care planning. Compared with the traditional H&P, student feedback suggests that the H&P 360 made them better clinicians. We can further speculate that by using the H&P 360, students develop a more complete picture of the patient—not just signs, symptoms, and diagnoses—but also the social and human narrative context that critically influences the presentation, management, and ultimately the outcomes of disease conditions. We suspect that gathering this more complete picture of patients' lives is one factor contributing to students' perception that the provider-patient relationship was enhanced by using the H&P 360. An important area for future investigation is the mechanism behind this enhanced relationship. Perhaps it is this more complete understanding of the patient, combined with improved patient rapport, that prompted the student comment that the H&P 360 "...return(ed) the humanity to medicine."

Implication of Findings

The observed variation across institutions and learner levels—though limited by small sample sizes and collinearity between institutions and learner levels—may offer insights into factors influencing medical students' perceived value of the H&P 360. Here, we present 3 speculative observations based on these data to encourage future research and application of the H&P 360. First, the teaching setting for the FIU students included in this study was a community-based, longitudinal, interprofessional environment with an established strong emphasis on comprehensive assessment and interprofessional care. As such, FIU students may have been less likely to perceive that the H&P 360 changed questions asked or added valuable information beyond their prior practice. Additionally, during the study period, FIU students conducted visits virtually due to COVID-19. Second, UC provided a relatively unique perspective on implementation efforts by including preclinical students and applying the H&P 360 in telehealth settings. Although the very high perceived value of the H&P 360 among both preclinical students and clinical-year students in telehealth settings is striking and promising, further investigation in other settings and institutions is needed to contextualize these findings. Finally, variation among institutions may reflect differences in overall emphasis on SDOH; the role of faculty in promoting or minimizing the study findings and application of the H&P 360; or differences in its use across inpatient, outpatient, and virtual settings.

XSL•FO
RenderX

Synthesizing evidence on the H&P 360 from this and previous studies, along with input from faculty and students and our own clinical teaching experience, we suggest that incorporating the H&P 360 into routine clinical practice involves at least 4 dimensions of learning. The general patterns and variations observed across different learner levels and clinical settings in this study support and shed light on each of these dimensions.

First, learners and educators using the H&P 360 will need to *integrate domain-based thinking alongside checklist-based approaches in data gathering.* Currently, early medical students are taught to take a history by following a memorized list of specific questions. Over time, an implicit process develops, where clinicians tailor questions, diagnoses, and management plans based on patient-specific information [28]. The direction a clinician takes for follow-up inquiry is likely influenced by many factors, including training experiences, knowledge and clinical skills in managing a wide range of issues (eg, emotional well-being, food insecurity, or safe housing), and local practice norms. Consequently, this approach is likely to vary widely among clinicians. Some naturally explore psychosocial dimensions, while others remain more narrowly focused on biomedical factors. To reduce this variation and better address the role of psychological and social factors in patients' health, the H&P 360 provides uniform, systematic prompts that help clinicians recognize social and psychological determinants of health. The H&P 360 represents a fundamental shift in learning to gather patient information by introducing 6 domains as general reference points alongside the traditional checklist focused on biomedical information.

For early learners, balancing domain-based thinking with a checklist approach can be disconcerting as they decide which specific content to include or exclude within the nonbiomedical domains. This challenge aligns with student feedback that the H&P 360 initially feels long and overwhelming when seen as a checklist of individual items, but becomes manageable and useful when viewed as a set of domain-based prompts that can be selectively explored—or revisited over multiple patient encounters. This is also consistent with findings that senior medical students using the H&P 360 identify and apply significantly more psychosocial information in their care planning than those using the traditional H&P [17]. Additionally, our finding that preclinical students found the H&P 360 added valuable information, helped them understand patients' goals, and facilitated stronger patient-provider relationships suggests that early learners can successfully incorporate domain-based thinking into routine data gathering. The interview guide that accompanies the H&P 360 can be a valuable resource in this regard. It helps students decide which domain to focus on and also supports faculty in navigating these domains during classroom teaching. See Multimedia Appendix 4 for the H&P 360 interview guide. Further exploration of domain-based thinking among medical learners is warranted to identify the best ways to provide a data-gathering framework that is both accessible in the early stages and comprehensive in the later stages of learning. We are particularly interested in methods for—and the implications of—incorporating patients' values, priorities, and goals into every clinical encounter [29,30].

Second, once familiar with domain-based thinking, medical students need to *develop skills in deciding which specific information within a domain is most relevant to a given patient encounter.* The finding that clerkship and postclerkship students from both inpatient and outpatient settings found the H&P 360 added valuable information and enhanced patient care suggests that students perceive tailoring domain-based questions to individual patients and clinical contexts as useful and facilitative for patient care.

Importantly, many students' comments revealed the emergence of skills in "modularizing" components of the H&P 360—using only those most relevant to a particular clinical context without feeling compelled to cover every domain.

Third, students need to *manage the emergent information through further inquiry or redirection.* As reflected in some student comments, medical students can feel compelled to fully elucidate or address the complex behavioral or social drivers of patients' health once identified. They also recognized that some behavioral or social needs uncovered during this process are important but do not require immediate action. Students should then redirect the interview to address matters of immediate concern (eg, potentially serious symptoms or a plan for initial hospital treatment), while simultaneously developing a plan to address longer-term issues. This process of identifying, prioritizing, and guiding the interview to optimize both disease-specific and contextual information has been demonstrated in the area of diagnostic reasoning, where a clinician listens and generates hypotheses, gathers data to test these hypotheses, and, depending on the results, offers treatment or pursues further diagnostic action [31]. We suggest that the domain-based framework of the H&P 360 facilitates the application of advanced interview skills not only to diagnostic assessment but also to management and care planning that better account for patients' psychosocial and environmental realities.

Finally, the *new information elicited with the H&P 360 must be applied to clinical management planning.* Our data suggest that, while learners generally found that information from the H&P 360 enhanced care planning, ratings in this area were lower than for other measures and lower than those observed in previous studies [3,17], particularly among early learners. We believe these findings highlight the complexity of incorporating social and behavioral information into care planning—for example, in discharging a patient facing homelessness or supporting medication adherence limited by insurance, income, transportation, or behavioral factors. By bringing these "background" issues to the forefront early in training, students can develop skills to mobilize interprofessional teams and utilize local resources as part of routine patient care. We also anticipate that by directly addressing SDOH—rooted in racial, ethnic, and socioeconomic inequities—learners will be better equipped to recognize and manage systemic and personal implicit biases that negatively impact care.

Although not emphasized in the student-oriented results presented here, our study suggested that faculty play an important role in promoting the effective use of the H&P 360. Faculty development and feedback methods varied across participating sites, ranging from interactive seminars to entirely

XSL•FO
RenderX

email-based communication, and faculty "buy-in" likely varied both within and across sites. At all sites, faculty were provided with information on the purpose, content, and suggested best practices for using the H&P 360. Anecdotally, however, learners reported little awareness or receptivity among teaching faculty toward using the H&P 360 domain framework in teaching and clinical management—except at 1 site where in-person faculty development was conducted before the COVID-19 pandemic. As professional development is influenced by cues from influential social sources, as well as practice resources and norms [22], effective application of the H&P 360 will likely require its incorporation into local teaching and clinical practices. Faculty development and the use of the 7-domain framework in teaching and clinical practice represent important areas for future investigation.

Limitations

Our study was limited by small sample sizes, which prevented more rigorous statistical analyses of the Likert-type scale data and further qualitative analysis. However, both in this work and in previous studies, student responses to the H&P 360 have been primarily positive. Additionally, it is important to acknowledge that some students experienced difficulties implementing the H&P 360 during virtual interactions and in specific clinical encounters. More implementation training on how to utilize the H&P 360 in different scenarios might be helpful for students. The survey also had a low response rate at some institutions, which may be attributable to several factors. The survey was optional and not required at all 3 sites. In some cases, faculty were unable to administer the survey immediately after course completion due to time constraints.

Lastly, this project took place at the beginning of the pandemic, which introduced many competing priorities and adjustments to the overall learning environment. The overall impact of response rates on the results is difficult to estimate, as rates varied by institution and were likely influenced by additional local factors. Information on factors that could promote or limit the effective application of the H&P 360 was not explored beyond the data collected from the student surveys. For example, curricular content encountered by students before the H&P 360, as well as organizational culture, could influence its application at each institution. Exploration of these factors was outside the scope of this study. Survey data were limited by too few observations in relevant substrata (eg, inpatient vs outpatient; longitudinal vs short-term; virtual vs face-to-face; and institutional vs community-based clinical settings) to permit meaningful subgroup analyses exploring additional variables that may impact the implementation of the H&P 360 in different settings.

Conclusions

The H&P 360 provides an enhanced template for data gathering that includes general prompts addressing key dimensions of human health not captured by the traditional H&P, such as patients' values, priorities, and goals. Our findings support the usability of the H&P 360 as a more comprehensive approach for medical students to gather patient information. Among early learners, it may be best to include a few specific illustrative items under each domain to familiarize students with the domains without requiring higher-order clinical knowledge or skills. Among later learners, the now-familiar domains can be used to promote more complete data gathering and to develop skills in integrating patients' goals, psychosocial and behavioral factors, and interprofessional teams into care planning. The H&P 360 may be particularly useful for making health inequities and their root causes more visible in routine clinical encounters, while guiding management planning to address them. Future work should measure its effects on patient care and outcomes.

Relevant topics for future investigation related to the H&P 360 include influences on students' use of the H&P 360 at different developmental stages; its use to identify and address SDOH; and methods and outcomes of faculty development to promote routine incorporation of domain-based thinking into clinical teaching and practice. To facilitate further investigation and implementation of the H&P 360 among medical schools, a set of tools and resources is available on the AMA website or authors may be contacted directly for further information.

Acknowledgments

The project was performed with financial support from the Medical Education Department within the American Medical Association. The authors thank their partners at the University of Michigan (contract ID AMA 36456), the University of Chicago (contract ID AMA 36457), and Florida International University (contract ID AMA 36572). The authors express gratitude for the planning and logistical support provided by several team members at each institution who led and coordinated the on-site study activities. They also deeply appreciate the contributions of numerous current AMA team members, notably Kevin Heckman, Kathryn Pajak, and Kimberly Loomis. The collaborating group authors include Jonathan Lio, MD; Vineet Arora, MD; Rachel Clarke, PhD, Onelia Lage, MD; and Ebony Whisenant, MD.

Authors' Contributions

RH and KK were the project leads for the AMA. JB was the lead investigator at FIU. GS was a coinvestigator at FIU. TK was the lead on the planning and conduct of analyses. VP led data analyses, and BW was the lead investigator at UM and developed the first draft of the H&P 360. All authors provided substantial contributions to the conception and design; acquisition of data; analysis and interpretation of data; drafting of the article and revising it critically for important intellectual content; final approval of the version to be published; and agreement to be accountable for all aspects of the work, ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
The H&P 360 template.
[DOCX File , 570 KB - mededu_v11i1e66221_app1.docx]

Multimedia Appendix 2
Student comments.
[DOCX File , 15 KB - mededu_v11i1e66221_app2.docx]

Multimedia Appendix 3
The CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist.
[PDF File (Adobe PDF File), 283 KB - mededu_v11i1e66221_app3.pdf]

Multimedia Appendix 4
Interview guide to support the H&P 360.
[DOCX File , 37 KB - mededu_v11i1e66221_app4.docx]

References

1. Keifenheim KE, Teufel M, Ip J, Speiser N, Leehr EJ, Zipfel S, et al. Teaching history taking to medical students: a systematic review. BMC Med Educ 2015 Sep 28;15(1):159 [FREE Full text] [doi: 10.1186/s12909-015-0443-x] [Medline: 26415941]
2. H. Kenneth W. Clinical methods: the history, physical, and laboratory examinations. Ann Intern Med 1990 Oct 01;113(7):563-563. [doi: 10.7326/0003-4819-113-7-563_2]
3. Kirley K, Hayer R, Khan T, Johnson E, Sanchez ES, Kosowicz L, et al. Expanding the traditional history and physical examination to address chronic diseases and social needs: a multisite randomized control trial of 4 medical schools. Acad Med 2020 Dec;95(11S Association of American Medical Colleges Learn Serve Lead: Proceedings of the 59th Annual Research in Medical Education Presentations):S44-S50. [doi: 10.1097/ACM.0000000000003640] [Medline: 32769457]
4. Preda A, Voigt K. The social determinants of health: why should we care? Am J Bioeth 2015 Mar 18;15(3):25-36. [doi: 10.1080/15265161.2014.998374] [Medline: 25786009]
5. World Health Organization (WHO). A Conceptual Framework for Action on the Social Determinants of Health. In: WHO. Geneva, Switzerland: WHO; 2010:1-76.
6. Thornton RLJ, Glover CM, Cené CW, Glik DC, Henderson JA, Williams DR. Evaluating strategies for reducing health disparities by addressing the social determinants of health. Health Aff (Millwood) 2016 Aug 01;35(8):1416-1423 [FREE Full text] [doi: 10.1377/hlthaff.2015.1357] [Medline: 27503966]
7. Alderwick H, Gottlieb LM. Meanings and misunderstandings: a social determinants of health lexicon for health care systems. Milbank Q 2019 Jul 08;97(2):407-419 [FREE Full text] [doi: 10.1111/1468-0009.12390] [Medline: 31069864]
8. Berwick DM. Getting serious about producing health. JAMA 2022 May 17;327(19):1865-1866. [doi: 10.1001/jama.2022.6921] [Medline: 35482356]
9. Castrucci B, Auerbach J. Meeting individual social needs falls short of addressing social determinants of health. Health Affairs. 2019 Jan 16. URL: https://www.healthaffairs.org/content/forefront/meeting-individual-social-needs-falls-short-addressing-social-determinants-health [accessed 2025-01-01]
10. World Health Organization (WHO). The Asia-Pacific Perspective: Redefining Obesity and Its Treatment. Geneva, Switzerland: WHO; 2000:1-55.
11. Maybank A, De Maio F, Lemos D, Derige DN. Embedding racial justice and advancing health equity at the American Medical Association. Am J Med 2022 Jul;135(7):803-805. [doi: 10.1016/j.amjmed.2022.01.058] [Medline: 35245496]
12. Penman-Aguilar A, Talih M, Huang D, Moonesinghe R, Bouye K, Beckles G. Measurement of health disparities, health inequities, and social determinants of health to support the advancement of health equity. Journal of Public Health Management and Practice 2016;22:S33-S42. [doi: 10.1097/phh.0000000000000373]
13. Benavidez GA, Zahnd WE, Hung P, Eberth JM. Chronic disease prevalence in the US: sociodemographic and geographic variations by zip code tabulation area. Prev Chronic Dis 2024 Mar 29;21:E14-E10. [doi: 10.5888/pcd21.230267] [Medline: 38426538]
14. Berg S. How these health systems are transforming chronic disease care. American Medical Association. 2024. URL: https://www.ama-assn.org/delivering-care/diabetes/how-these-health-systems-are-transforming-chronic-disease-care [accessed 2025-01-01]
15. Willis CD, Riley BL, Herbert CP, Best A. Networks to strengthen health systems for chronic disease prevention. Am J Public Health 2013 Nov;103(11):e39-e48. [doi: 10.2105/ajph.2013.301249]

XSL•FO
RenderX

16. Faustinella F, Jacobs RJ. The decline of clinical skills: a challenge for medical schools. Int J Med Educ 2018 Jul 13;9:195-197 [FREE Full text] [doi: 10.5116/ijme.5b3f.9fb3] [Medline: 30007951]
17. Williams BC, Ward DA, Chick DA, Johnson EL, Ross PT. Using a six-domain framework to include biopsychosocial information in the standard medical history. Teach Learn Med 2019 Sep 14;31(1):87-98. [doi: 10.1080/10401334.2018.1480958] [Medline: 30216097]
18. Bauer UE, Briss PA, Goodman RA, Bowman BA. Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. The Lancet 2014 Jul;384(9937):45-52. [doi: 10.1016/s0140-6736(14)60648-6]
19. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012 Oct 23;27(10):1361-1367. [doi: 10.1007/s11606-012-2077-6] [Medline: 22618581]
20. Hardcastle MR, Usher KJ, Holmes CA. An overview of structuration theory and its usefulness for nursing research. Nursing Philosophy 2005 Aug 30;6(4):223-234. [doi: 10.1111/j.1466-769x.2005.00230.x]
21. de Jong T. Cognitive load theory, educational research, and instructional design: some food for thought. Instr Sci 2009 Aug 27;38(2):105-134. [doi: 10.1007/s11251-009-9110-0]
22. Bandura A. Social Learning Theory. Englewood Cliffs, NJ: Prentice Hall; 1977.
23. McLeod S. Albert Bandura's social learning theory. In: Simply Psychology. London, UK: Prentice Hall; 2011.
24. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM). Washington, DC: American Psychiatric Association; 2000.
25. Rieger EY, Anderson IJ, Press VG, Cui MX, Arora VM, Williams BC, et al. Implementation of a biopsychosocial history and physical exam template in the electronic health record: mixed methods study. JMIR Med Educ 2023 Feb 21;9:e42364 [FREE Full text] [doi: 10.2196/42364] [Medline: 36802337]
26. Adams NE. Bloom's taxonomy of cognitive learning objectives. J Med Libr Assoc 2015 Jul;103(3):152-153 [FREE Full text] [doi: 10.3163/1536-5050.103.3.010] [Medline: 26213509]
27. Wigfield A, Eccles JS. Expectancy-value theory of achievement motivation. Contemp Educ Psychol 2000 Jan;25(1):68-81. [doi: 10.1006/ceps.1999.1015] [Medline: 10620382]
28. Carraccio CL, Benson BJ, Nixon LJ, Derstine PL. From the educational bench to the clinical bedside: translating the Dreyfus developmental model to the learning of clinical skills. Academic Medicine 2008;83(8):761-767. [doi: 10.1097/acm.0b013e31817eb632]
29. Vermunt NPCA, Harmsen M, Westert GP, Olde Rikkert MGM, Faber MJ. Collaborative goal setting with elderly patients with chronic disease or multimorbidity: a systematic review. BMC Geriatr 2017 Jul 31;17(1):167-112 [FREE Full text] [doi: 10.1186/s12877-017-0534-0] [Medline: 28760149]
30. McEwan D, Harden SM, Zumbo BD, Sylvester BD, Kaulius M, Ruissen GR, et al. The effectiveness of multi-component goal setting interventions for changing physical activity behaviour: a systematic review and meta-analysis. Health Psychol Rev 2016 Nov 13;10(1):67-88. [doi: 10.1080/17437199.2015.1104258] [Medline: 26445201]
31. Sox HC, Higgins MC, Owens DK, Schmidler GS. Medical Decision Making. Hoboken, NJ: Wiley Blackwell; 2024.

Abbreviations

AMA: American Medical Association
CDC: Centers for Disease Control and Prevention
FIU: Florida International University
H&P: history and physical
LGBTQ+: lesbian, gay, bisexual, transgender, queer or questioning, and other diverse sexual orientations and gender identities
SDOH: social determinants or social drivers of health
UC: University of Chicago
UM: University of Michigan
WHO: World Health Organization

Edited by B Lesselroth; submitted 16.09.24; peer-reviewed by F Yang, LE Eberman; comments to author 03.11.24; revised version received 10.12.24; accepted 06.03.25; published 05.06.25.

Please cite as:
Hayer R, Tang J, Bisschops J, Schneider GW, Kirley K, Khan T, Rieger E, Walford E, Anderson I, Press V, Williams B
Implementing the H&P 360 in Three Medical Institutions: Usability Study
JMIR Med Educ 2025;11:e66221
URL: https://mededu.jmir.org/2025/1/e66221
doi:10.2196/66221
PMID:

XSL•FO
RenderX

©Rupinder Hayer, Joyce Tang, Julia Bisschops, Gregory W Schneider, Kate Kirley, Tamkeen Khan, Erin Rieger, Eric Walford, Irsk Anderson, Valerie Press, Brent Williams. Originally published in JMIR Medical Education (https://mededu.jmir.org), 05.06.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Resilience Training Web App for National Health Service Keyworkers: Pilot Usability Study

Joanna Burrell[1*], BA, PGCert, DClinPsy; Felicity Baker[2*], BSc, MPhil, DClinPsy; Matthew Russell Bennion[3,4*], BEng, MSc, MBA, PhD

[1]Department of Psychology, University of Sheffield, Cathedral Court, Sheffield, United Kingdom

[2]Ultimate Resilience Ltd, Nottingham, United Kingdom

[3]Department of Computer Science, The University of Sheffield, Sheffield, United Kingdom

[4]Digital Innovation Unit, NHS Midlands and Lancashire Commissioning Support Unit, Stoke on Trent, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Joanna Burrell, BA, PGCert, DClinPsy
Department of Psychology, University of Sheffield, Cathedral Court, Sheffield, United Kingdom

Abstract

Background: It is well established that frontline health care staff are particularly at risk of stress. Resilience is important to help staff to manage daily challenges and to protect against burnout.

Objective: This study aimed to assess the usability and user perceptions of a resilience training web app developed to support health care keyworkers in understanding their own stress response and to help them put into place strategies to manage stress and to build resilience.

Methods: Nurses (n=7) and other keyworkers (n=1), the target users for the resilience training web app, participated in the usability evaluation. Participants completed a pretraining questionnaire capturing basic demographic information and then used the training before completing a posttraining feedback questionnaire exploring the impact and usability of the web app.

Results: From a sample of 8 keyworkers, 6 (75%) rated their current role as “sometimes” stressful. All 8 (100%) keyworkers found the training easy to understand, and 5 of 7 (71%) agreed that the training increased their understanding of both stress and resilience. Further, 6 of 8 (75%) agreed that the resilience model had helped them to understand what resilience is. Many of the keyworkers (6/8, 75%) agreed that the content was relevant to them. Furthermore, 6 of 8 (75%) agreed that they were likely to act to develop their resilience following completion of the training.

Conclusions: This study tested the usability of a web app for resilience training specifically targeting National Health Service keyworkers. This work preceded a larger scale usability study, and it is hoped this study will help guide other studies to develop similar programs in clinical settings.

(JMIR Med Educ 2025;11:e51101) doi:10.2196/51101

KEYWORDS

resilience; workplace stress; National Health Service; NHS keyworker; digital learning; digital health; usability; feasibility; mental health; pilot study; learning; training; exercise; primary care provider; health care professional; occupational health; worker; hospital; emergency; survey; questionnaire; mobile phone

Introduction

Resilience allows individuals to manage everyday challenges and changes. For health care professionals who are working in highly emotive and stressful situations, resilience skills are particularly important [1]. It is well established that frontline staff such as nurses are particularly at risk of stress due to factors such as long shifts, organizational pressures, and the emotional impact of their work [2]. During the COVID-19 pandemic, there were high rates of mental health problems among health care staff. For example, a survey of 255 nurses working with respiratory patients found 21% to be experiencing moderate or severe symptoms of anxiety and 17.2% to be experiencing depression. A total of 18.9% scored low or very low on a measure of resilience [3]. A study of 1106 physicians also reported high levels of anxiety and low levels of resilience during the pandemic [4]. Building emotional resilience is therefore imperative to prevent burnout in health care staff, to keep them healthy both physically and mentally, to improve well-being, and to ensure job retention in the workplace [1].

One way that employers can meet this need is through the provision of resilience training. Key benefits of resilience training include improvements in mental health and well-being, social support, self-efficacy, and coping. Further benefits include

XSL•FO
RenderX

improved ability to adapt to pressures and demands in the workplace and other areas of life [5]. In health care settings, nurse resilience interventions have been highlighted as a potential way of enhancing staff coping and well-being, job satisfaction, and retention [2]. Greater nurse resilience has also been associated with better work performance [6].

A constraint of traditional resilience training programs is the time required to attend in person, which can exclude certain staff groups, such as nurses, from participation. Smartphone apps have the potential to offer training in resilience to large numbers of people while overcoming barriers, such as stigma, time, and acceptability, and can be integrated easily into the wider organizational well-being strategy [7-11].

The aim of this study was to evaluate whether health care keyworkers would be willing to carry out resilience training via an online platform specifically designed to enable them to understand their own stress response and put in place strategies to manage stress, build emotional resilience, and maintain well-being. The data collected would generate important information for future implementation, while contributing feedback for a more refined usability study with this population.

Methods

Participants

The recruitment process was carried out by the Medical Devices Testing and Evaluation Centre (MD-TEC) team. A study sample was recruited from the University Hospitals Birmingham National Health Service (NHS) Foundation Trust, and participation was incentivized with Continuing Professional Development credits for participation. This study was advertised over the web via the internal trust-wide communications bulletin and targeted emails. There was not an enforced inclusion criterion, but this study requested for participants who were a nurse or health care professional and worked in either the emergency department, intensive care unit, or critical care.

Ethical Considerations

This study was run as a formative usability study by MD-TEC with human participants. The University of Sheffield Re-Use of Existing Data Questionnaire was completed, and the Psychology Research Ethics Committee deemed this study exempt from ethical approval because the data were fully anonymized. A short self-declaration form was submitted. This application went to the Psychology Departments Ethics Administrator for a final check before a letter of confirmation was issued.

Informed consent was derived through the sharing of a recruitment flyer with potential participants. This explained the research and its function as part of medical device usability testing for further development.

All the data were received fully anonymized from MD-TEC post study. The participants were not personally identifiable by the researchers. Research participants were offered Continuing Professional Development credits for their participation. They self-selected for and undertook the research voluntarily.

The training web app was developed by the third author without funding. The content of the training web app drew on the Skills-Based Model of Personal Resilience [12] and included a selection of evidence-based skills and exercises to regulate distress emotions and build positive emotions, such as slow rhythmic breathing and mindfulness practice. The selection of skills and exercises were chosen for their capacity to provide maximum benefit to participants, calming stress, and facilitating improved coping, in the context of this brief trial.

Bennion et al [8] highlight four key indicators of quality drawn from effective digital psychotherapy approaches. These include clinician involvement, academic involvement, research, or other evidence and use of specific psychological approach or theory. The intervention followed these recommendations, drawing on academic and clinical theory [13] and involving clinicians, academics, and computer scientists in its development to ensure greater quality and effectiveness.

The web-based training was published using Articulate Storyline (Articulate Global, LLC) and accessed via a web browser. It consisted of both written and spoken content on a series of slides, short videos, and experiential exercises which could be moved through at participants' own pace using "previous" and "next" buttons. The estimated time to complete the training was 20 minutes.

Pretraining Questionnaire

The pretraining questionnaire captured basic demographic information: gender, age range, job role, current area of work, current band, and years of nursing experience. Current job stress was rated on a 5-point scale (never, hardly ever, occasionally, sometimes, and very). Participants were also asked whether they had heard of or previously undertaken any resilience training.

Posttraining Feedback Questionnaire

The posttraining feedback questionnaire focused on 6 areas: app design and navigation, app content, app impact, app training exercises, app relevance, and app access. Each question was posed on a Likert scale with five possible answer options to allow the user to respond to each statement on a range from "strongly agree" to "strongly disagree."

Procedure

Upon contacting the MD-TEC team to participate, interested individuals were provided the opportunity to ask any questions about participation in this study. If willing to consent, participants were then sent the link and password to access the training. The training and surveys were hosted on the MD-TEC Software Usability Testing Site (MD-TEC), and thus could be completed on any device with internet access. Once logged in on an internet browser, individuals were presented with the precourse survey before completing the full training module.

Once the training module was completed, participants were taken to a landing page and requested to click a link to take them to the feedback survey. They were reminded at this point that no identifiable information would be collected from them. As the surveys were completed anonymously, participants who completed the training and survey were asked to inform the

XSL•FO
RenderX

MD-TEC team via email once they had done so. They were then sent a certificate toward Continuing Professional Development for their time contributed to research, which they could add to their personal records. The total time for each participant to complete the training module and feedback survey was approximately 45 minutes.

Statistical Analysis

This study did not use a specific sample size calculation as it was focused on app usability. It instead aimed to achieve at least 5 participants which is deemed an optimal number to reveal 77% to 85% of problems [14]. Data were analyzed using IBM SPSS (version 26; IBM Corp). The pretraining and posttraining feedback questionnaires were summarized as a mix of continuous variables with medians and categorical ordinal variables with percentages.

Results

Participants

The age of participants ranged from 25 to 64 years. The total sample (N=8) was comprised of 8 (100%) females, 7 (87.5%) nurses, and 1 (12.5%) keyworker of other professions. Grades ranged from 5 to 7 with a median of 6 (IQR 5-6). Five of the participant's had over 15 years of nursing experience. All 8 participants completed baseline measures and posttraining measures.

Pretraining Questionnaire

Sample Overview

Of the 8 participants who completed this study, 3 (37.5%) worked in the hospital's intensive care unit, 1 (12.5%) worked in the emergency department, and 4 (50%) worked in other undisclosed areas of the hospital.

Current Role Stress

Most participants (6/8, 75%) rated their current role stress as "sometimes" stressful, while 1 of 8 (12.5%) said "occasionally" stressful and 1 of 8 (12.5%) said "very" stressful.

Awareness and Knowledge of Resilience Training

Most participants (6/8, 75%) had heard of resilience training, and those that had taken part (4/8, 50%) had done so in a face-to-face setting.

Posttraining Feedback Questionnaire

App Design and Navigation

Feedback regarding the design of the training was predominantly positive. All participants found the training easy to navigate, 6 of 8 (75%) deemed the default speed at which the training progressed to be acceptable, and 7 of 8 (97.5%) thought the appearance of the buttons was OK.

App Content

Feedback for the content indicated that all participants (8/8, 100%) found the training easy to understand, 6 of 8 (75%) felt there was enough text content, 4 of 8 (50%) felt there was enough spoken content, and 5 of 8 (62.5%) felt there were enough interactive exercises.

App Impact

A large number of the participants (5/7, 71%) agreed that the training increased their understanding of both stress and resilience, while 6 of 8 (75%) agreed that the resilience model had helped them to understand what resilience is.

App Training Exercises

The training exercise feedback was positive but varied. For the breathing and positive tips exercises, 6 of 8 (75%) participants agreed they were likely to try the exercises again in the future. The mindfulness exercise had 4 of 8 (50%) participants agree they were likely to try the exercise again.

App Relevance

There was a high level of agreement that the training was relevant to nurses, with 6 of 8 (75%) participants agreeing that the content was relevant to them.

Furthermore, 6 of 8 (62.5%) participants agreed that they were likely to act to develop their resilience following completion of the training.

Access to Training

All the participants indicated a different personal preference to how they would prefer to access the training. Participants felt the package should be made available across all platforms to allow the training to be completed where and when it was most convenient to them. When asked their preferred location of access, 5 of 8 (62.5%) indicated their preference as being "at home."

Discussion

Principal Findings

We explored the perceived usability and feasibility of a resilience training web app created for NHS health care keyworkers. Data collected covered a number of areas: design and navigation, content, impact, and relevance. The results showed that 100% (8/8) of participants found the training easy to understand and agreed that it had increased their understanding of both stress (5/7, 71%) and resilience (6/8, 75%). Three-quarters of participants agreed that the content was relevant to them, and this corresponded with the number of participants rating their current role as "sometimes" stressful. Furthermore, three-quarters of participants agreed that they were likely to take action to develop their resilience following completion of the training. This information was used to inform the design of a larger usability study.

A total of 8 participants were recruited, with 7 being from the target population. All participants completed the process from start to finish. Participants successfully carried out what was required of them based on this study's protocol, although some participants did not complete all the questions asked on the posttraining questionnaire. There was no indication given as to why this was the case. In a follow up usability study [13] validation checks were put in place within the surveys to stop questions from being missed by mistake.

XSL•FO
RenderX

The findings of this study indicated that participants found the training app design and navigation acceptable and usable. However, the measure used was not a standard model of system usability (eg, International Organization for Standardization, 2018). This study's design was updated to use two validated measures (the System Usability Scale and the Usability Metric for User Experience) to strengthen the robustness of a follow-up usability study [14]. Adding these two additional validation measures to this study's design helped to strengthen assessment of the training app's usability.

Participants indicated that the training was easy to understand and that there was enough text content; however, they also indicated that there was a need for the training to have more spoken and interactive content. This fits with a recent study [15] in which nurses' interactive behavior was identified as an influencing aspect of nurse satisfaction with online learning. Based on these findings, we recommend the training's interactive content be revisited in its next design iteration.

Most participants perceived that the training increased their understanding of both stress and resilience and that the resilience model had helped them to understand what resilience is. A more robust method of measurement was required to further explore the impact of the training and this study's design was updated to incorporate ratings of perceived knowledge regarding stress and resilience. These new scales were used in a follow up usability study [14] and found to increase significantly between pre- and postapp training.

The training exercise feedback was positive but varied. Both the breathing and positive tips exercises were well received, with participants agreeing they were likely to try the exercises again in the future. However, only half of the participants agreed that they would try the mindfulness exercise again. This may have been due to the difficulty in carrying out the exercise in a busy work environment.

Many participants agreed that the training was relevant to them and believed that they were likely to act to develop their resilience following completion of the training.

Limitations

Recognized limitations of usability studies include that testing is conducted in an artificial situation and personal preferences of the participants are not representative of the wider user population [16]. The digital training app used in our study is an early prototype. This may need multiple design developments to create a smartphone app that can be used to deliver the resilience training. The aim of this formative usability study was to assess the acceptability and user perceptions of the current version of the training program. As such, this study is part of the iterative product development process and is different to a summative usability study, conducted for validation and regulatory purposes [17].

This study had a single-group design and advertised for a specific group; however, anyone employed by the trust who contacted MD-TEC regarding this study could be involved. This was done primarily to allow anyone employed by the trust to gain access to training that could benefit them. Potential participants who were unaware of this clause may have been lost because of this decision. The initial training materials were designed with nurses in mind but were not specifically tailored for the demographic. This may have changed participants' initial perception of the suitability of the training to them personally. A single-group design can limit the ability to draw definitive conclusions about the effectiveness of the training due to its lack of a control or comparison group [18]. However, since this study was focused on the usability of the training and not the effectiveness, and it was not seeking to make a comparative analysis, a single-group design was appropriate.

This study limited its evaluation to perceived usability, which was not obtained through laboratory-based observations. As such, the positive ratings reported may not be representative of true user experience. A heuristic evaluation of the training to detect usability problems was not carried out, due to pandemic restrictions making this problematic to implement. This study used quantitative scales and measures to collect data but did not use qualitative measures to gain deeper insight into what NHS health care staff felt about the training. A measure of time spent using the training was not collected. This could have also given an indication of acceptability. This study used two single Likert scales to measure perceived increases in knowledge about stress and resilience. Studies have shown that perceptions of learning may not reflect knowledge gains, when compared with evidence of actual learning [19]. A more robust method of measuring knowledge retention would have benefitted this study. This could have been achieved by having a pre- and postquiz based on the content of the training to see what knowledge was retained.

While the majority of participants gave positive responses in the evaluation of this study, the generalizability of these outcomes is limited due to the disproportionate number of female participants and participants from a nursing background. Only 1 (12.5%) participant was from a different professional group. This limits the inferences that can be drawn about usability and acceptability of the training to male participants and those with other keyworker roles. It is recommended that future studies recruit a more representative sample to enhance generalizability of the results.

Conclusions

Overall, the resilience training module was well received by the participants. The participants felt the package was easy to navigate. There was a high level of agreement that the visual delivery of the training was acceptable, as well as the speed at which this was delivered.

A number of techniques demonstrated during the training were also well received, with 6 of 8 participants agreeing that they would use them in future stressful situations. Mindfulness was the only exercise that received more varied feedback, with half agreeing on its utility in the work environment.

Health care staff participating in this study largely agreed that the training was relevant to their group and that the tone of the delivery was appropriate. No clear preference regarding how to access the training was identified, highlighting the need for accessibility via computer, tablet, and smartphone. Participants

XSL•FO
RenderX

expressed a wish to to access the training when they have a moment of need and the opportunity in their busy working day.

Future Directions

As one of the first NHS web-based resilience programs to be tested, this first usability study aimed to understand whether web-based training for resilience is deemed usable and acceptable by health care staff. The results of this study will be used to expand and build upon the initial prototype to make a more interaction enriched version of the training.

This study also provided an understanding of the program's limitations and highlighted some aspects which require further adaptation for delivery via a new medium. Future research would aim to evaluate the impact of including greater interactivity on engagement and learning. It would also aim to extend the accessibility and acceptability of the program to a wider audience by developing an effective prototype for a smartphone app.

This study was run externally by MD-TEC, who had their own processes for running usability studies of this nature. This study's design covered some of the key factors required for an effective online survey, but it could have been further improved by seeking acknowledgment with MD-TEC regarding the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist [20].

It is clear from the results that there is a need for future research to evaluate how skills-based learning using web-based training impacts long term resilience. A larger scale study would allow for more in-depth investigation of the impact of such training on participants' levels of stress and resilience as well as their perspectives on acceptability.

Given the diversity of NHS staff, it will be important for any future study to gather a wide set of demographic information to investigate acceptability and generalizability across diverse populations. With increasing awareness (ie, gained through the COVID-19 pandemic) of the pressures faced by all NHS staff, across a breadth of ethnic and socioeconomic groups, a larger scale study would allow for a wider inclusion criterion covering all NHS staff groups.

Acknowledgments

This study was supported by a European Regional Development Fund (ERDF) awarded to FB and JB. We thank the National Institute for Health and Care Research Trauma Management MedTech Co-Operative for facilitating the usability study.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JB did the conceptualization, methodology, writing of the original draft, review and editing of the writing, visualization, supervision, project administration, and funding acquisition. FB handled the conceptualization, methodology, writing of the original draft, review and editing of the writing, visualization, and supervision. MRB worked on the conceptualization, methodology, software, validation, formal analysis, resources, data curation, writing of the original draft, review and editing of the writing, visualization, and supervision.

Conflicts of Interest

JB and FB are employees of Ultimate Resilience LTD, creators of the Skills-Based Model of Personal Resilience applied to the web app. MRB developed the web app.

References

1. Yılmaz EB. Resilience as a strategy for struggling against challenges related to the nursing profession. Chin Nurs Res 2017;4(1):9-13. [doi: 10.25164/cnr201701003]
2. Henshall C, Davey Z, Jackson D. Nursing resilience interventions-a way forward in challenging healthcare territories. J Clin Nurs 2020 Oct;29(19-20):3597-3599. [doi: 10.1111/jocn.15276] [Medline: 32237252]
3. Roberts NJ, McAloney-Kocaman K, Lippiett K, Ray E, Welch L, Kelly C. Levels of resilience, anxiety and depression in nurses working in respiratory clinical areas during the COVID pandemic. Respir Med 2021 Jan;176:106219. [doi: 10.1016/j.rmed.2020.106219] [Medline: 33248362]
4. Mosheva M, Hertz-Palmor N, Dorman Ilan S, et al. Anxiety, pandemic-related stress and resilience among physicians during the COVID-19 pandemic. Depress Anxiety 2020 Oct;37(10):965-971. [doi: 10.1002/da.23085] [Medline: 32789945]
5. Helmreich I, Kunzler A, Chmitorz A, et al. Psychological interventions for resilience enhancement in adults. Cochrane Database Syst Rev 2017;2017(2):CD012527. [doi: 10.1002/14651858.CD012527]
6. Walpita YN, Arambepola C. High resilience leads to better work performance in nurses: evidence from South Asia. J Nurs Manag 2020 Mar;28(2):342-350. [doi: 10.1111/jonm.12930] [Medline: 31845421]

XSL•FO
RenderX

7. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7. [doi: 10.2196/mental.4984] [Medline: 26932350]
8. Bennion MR, Hardy GE, Moore RK, Kellett S, Millings A. E-Therapies in England for stress, anxiety or depression: how are apps developed? A survey of NHS e-therapy developers. BMJ Health Care Inform 2019 Jun;26(1):e100027. [doi: 10.1136/bmjhci-2019-100027] [Medline: 31171556]
9. Blake H, Bermingham F, Johnson G, Tabner A. Mitigating the psychological impact of COVID-19 on healthcare workers: a digital learning package. Int J Environ Res Public Health 2020 Apr 26;17(9):2997. [doi: 10.3390/ijerph17092997] [Medline: 32357424]
10. Rich A, Aly A, Cecchinato ME, et al. Evaluation of a novel intervention to reduce burnout in doctors-in-training using self-care and digital wellbeing strategies: a mixed-methods pilot. BMC Med Educ 2020 Sep 9;20(1):294. [doi: 10.1186/s12909-020-02160-y] [Medline: 32907573]
11. Golden EA, Zweig M, Danieletto M, et al. A resilience-building app to support the mental health of health care workers in the COVID-19 era: design process, distribution, and evaluation. JMIR Form Res 2021 May 5;5(5):e26590. [doi: 10.2196/26590] [Medline: 33872189]
12. Baker FRL, Baker KL, Burrell J. Introducing the skills‐based model of personal resilience: drawing on content and process factors to build resilience in the workplace. J Occup Organ Psyc 2021 Jun;94(2):458-481 [FREE Full text] [doi: 10.1111/joop.12340]
13. Bennion MR, Baker F, Burrell J. An unguided web-based resilience training programme for NHS keyworkers during the COVID-19 pandemic: a usability study. J Technol Behav Sci 2022;7(2):125-129. [doi: 10.1007/s41347-021-00225-3] [Medline: 35317264]
14. Nielsen J. Estimating the number of subjects needed for a thinking aloud test. Int J Hum Comput Stud 1994 Sep;41(3):385-397. [doi: 10.1006/ijhc.1994.1065]
15. Lv K, Zhou N. Influencing aspects of clinical nurses' interactive continuous learning behaviour based on in-service online network videos: a grounded theory approach. Nurse Educ Today 2023 Mar;122:105726. [doi: 10.1016/j.nedt.2023.105726] [Medline: 36736040]
16. Rubin J, Chisnell D. Handbook of Usability Testing, 2nd edition: Wiley Publishing; 2008.
17. FDA draft guidance: human factors studies and related clinical study considerations in combination product design and development. : FDA; 2016.
18. Knapp TR. Why is the one-group pretest-posttest design still used? Clin Nurs Res 2016 Oct;25(5):467-472. [doi: 10.1177/1054773816666280] [Medline: 27558917]
19. Persky AM, Lee E, Schlesselman LS, Psych E. Perception of learning versus performance as outcome measures of educational research. Am J Pharm Educ 2020 Jul;84(7):ajpe7782. [doi: 10.5688/ajpe7782] [Medline: 32773832]
20. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34. [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
MD-TEC: Medical Devices Testing and Evaluation Centre
NHS: National Health Service

Edited by TDA Cardoso; submitted 22.07.23; peer-reviewed by B Rosen, S Kaur; revised version received 10.11.24; accepted 12.11.24; published 06.01.25.

Please cite as:
Burrell J, Baker F, Bennion MR
Resilience Training Web App for National Health Service Keyworkers: Pilot Usability Study
JMIR Med Educ 2025;11:e51101
URL: https://mededu.jmir.org/2025/1/e51101
doi:10.2196/51101

© Joanna Burrell, Felicity Baker, Matthew Russell Bennion. Originally published in JMIR Medical Education (https://mededu.jmir.org), 6.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluation of a Simulation Program for Providing Telenursing Training to Nursing Students: Cohort Study

Ola Ali-Saleh*, PhD; Layalleh Massalha*, MA; Ofra Halperin*, PhD

Department of Nursing, Max Stern Yezreel Valley College, Emek Yezreel, Israel

*all authors contributed equally

Corresponding Author:
Ofra Halperin, PhD
Department of Nursing, Max Stern Yezreel Valley College, Emek Yezreel, Israel

Abstract

Background: Telenursing has become prevalent in providing care to diverse populations experiencing different health conditions both in Israel and globally. The nurse-patient relationship aims to improve the condition of individuals requiring health services.

Objectives: This study aims to evaluate nursing graduates' skills and knowledge regarding remote nursing care prior to and following a simulation-based telenursing training program in an undergraduate nursing degree.

Methods: A cohort study assessed 114 third-year nursing students using comprehensive evaluation measures of knowledge, skills, attitudes, self-efficacy, and clinical skills regarding remote nursing care. Assessments were conducted at 2 critical time points: prior to and following a structured simulation-based training intervention.

Results: Participant demographics revealed a predominantly female sample (101/114, 88.6%), aged 20 - 50 years (mean 25.68, SD 4.59 years), with moderate to advanced computer and internet proficiency. Notably, 91.2% (104/114) had no telenursing exposure, yet 75.4% (86/114) expressed training interest. Statistical analyses demonstrated significant improvements across all measured variables, characterized by moderate to high effect sizes. Key findings included substantial increases in telenursing awareness, knowledge, skills, attitudes and self-efficacy; significant reduction in perceived barriers to remote care delivery; and complex interrelation dynamics between variables. A multivariate analysis revealed nuanced correlations: higher awareness and knowledge were consistently associated with more positive attitudes and increased self-efficacy. Positive attitudes correlated with enhanced self-efficacy and reduced perceived barriers. Change score analyses further indicated that increased awareness and knowledge facilitated more positive attitudinal shifts, while heightened awareness and positive attitudes corresponded with decreased implementation barriers.

Conclusions: The study underscores the critical importance of integrating targeted telenursing training into nursing education. By providing comprehensive preparation, educational programs can equip students to deliver optimal remote care services. The COVID-19 pandemic has definitively demonstrated that remote nursing will be central to future health care delivery, emphasizing the urgent need to prepare nursing students for this emerging health care paradigm.

(JMIR Med Educ 2025;11:e67804) doi:10.2196/67804

KEYWORDS

simulation-based training program; telenursing; simulation; program; training; nursing student; nursing care; Israel; nurse-patient relationship; telehealth nursing; remote nursing care; undergraduate; cohort study; knowlege; self efficacy; skills; attitudes

Introduction

Background

In recent decades, telehealth—the use of information and communication technology in health care—has become a global priority [1]. Telenursing, a subset of telehealth defined as "the use of technology to deliver nursing care and conduct nursing practice" [2], has emerged as a significant health care option. Indeed, telenursing enables patients to access health care providers remotely through various technologies, including mobile devices, computers, and videoconferencing [3]. The American Nurses Association defines telenursing as the use of "technology to deliver nursing care and conduct nursing practices" [4].

Telenursing offers numerous benefits, including improved access to care, savings in time and resources, and enhanced self-care opportunities [5]. Telenursing has been found effective in reducing the number of outpatient and emergency room visits, shortening hospital stays, and lowering health care costs [6]. It has also proved beneficial in educating patients, promoting self-care competence, and providing cost-effective mental health support [7], as well as in providing care for chronically ill patients [5], oncology patients [6] and palliative care [8].

XSL•FO
RenderX

The COVID-19 pandemic accelerated the adoption of telenursing by emphasizing its crucial role in disaster and public health emergency responses [9]. This shift highlighted the need for integrating telenursing concepts into nursing education at both undergraduate and graduate levels [3,10].

Despite the growing importance of telenursing, it is often underrepresented in nursing curricula. Poreddi et al [11] found that while nurse interns generally hold positive perceptions of telenursing, their knowledge of the subject is limited. This gap underscores the need to incorporate telenursing concepts into nursing education in order to prepare future health care providers for an increasingly digital environment. Nurses play an indispensable role in telehealth implementation, with their skills and attitudes serving as supportive factors [12].

Nevertheless, telenursing also entails challenges. Indeed, to provide optimal care without physical contact nurses must possess high-level clinical and interpersonal skills [13,14]. A lack of sufficient knowledge and skills constitutes the main obstacle in telenursing [15]. Previous studies have reported that telehealth education in nursing programs is inadequate [16]. To acquire the skills and develop the competencies required for telenursing, students must practice the use of screen technology and virtual access to remote patients, and telenursing should be introduced early in the nursing curriculum [17].

Simulation-based learning has been identified as an effective method for teaching telenursing skills. This approach allows students to practice in a safe and realistic environment in which they can improve their cognitive, emotional, and psychomotor abilities [18,19]. Studies have shown that simulations positively affect self-efficacy, academic motivation, and the acquisition of clinical skills [20,21]. Moreover, experiential education can be used to augment such crucial factors as perceived usefulness, self-efficacy, and innovativeness, thus enhancing our understanding of the effectiveness and implications of telenursing [22].

Glinkowski et al [23] examined telehealth among college nursing students. They found that 67% (207/308) of participants were willing to engage in telehealth services and 69.49% (214/308) agreed that telehealth should be included as part of the nursing education curriculum. Indeed, enhancing understanding of telenursing and establishing a robust human infrastructure among future nursing professionals are of critical importance. Receptiveness and adaptability have emerged as crucial factors in shaping the quality of health care services provided through telenursing [24]. Studies that examined nursing and medical students in the United States and Poland revealed positive perspectives and attitudes toward telehealth and telenursing. Yet, these studies also revealed knowledge gaps among these students, as well as erroneous beliefs regarding the advantages and possibilities of using telenursing in health practice [25-28]. Adequate education is needed to overcome this lack of knowledge and improve students' attitudes, including more telenursing content [29].

According to Assaye et al [30], the most significant factors influencing perceptions of telenursing among health care providers include technology availability, web access, and lack of telemedicine training. Indeed, nurses with insufficient education and training in the use of technology face difficulties in implementing telenursing [31]. To overcome these difficulties, students must develop a positive and unprejudiced attitude, while acquiring comprehensive knowledge and acknowledging the limitations of these technologies [32].

Despite the importance of integrating telenursing in nursing study programs, in Israel this topic has not been incorporated into the nursing director's core program and is not taught in practical nursing training programs. Although studies have been conducted on the use of telenursing, very few have examined the issue of training nursing students to use it. Implementing and evaluating such a training program have the potential to help integrate telenursing into the nursing director's core program.

In conclusion, despite the increasing importance of telenursing in health care delivery, particularly in view of recent global events, its integration into nursing education remains limited, particularly in certain countries, such as Israel. Further research, educational initiatives, and pilot training programs are needed to bridge this gap and ensure that future nurses are adequately prepared for the evolving landscape of health care delivery.

This study evaluates the skills and knowledge of third-year nursing students regarding remote nursing care before and after participation in a simulation-based training program on telenursing as part of their undergraduate nursing degree. To the best of our knowledge, very few studies to date have evaluated programs that use simulation to train nursing students in the provision of nursing care from a distance (telenursing).

Hypotheses

This study tests 3 hypotheses. The first hypothesis posits that participation in telenursing training will lead to increased awareness, knowledge, and understanding of required skills in telenursing practice. We expect to observe improved attitudes toward telenursing and enhanced self-efficacy, while simultaneously seeing a reduction in perceived barriers to telenursing implementation. The second hypothesis suggests that there will be significant positive correlations between nursing students' self-efficacy in telenursing and their awareness, knowledge, skill perceptions, and attitudes toward telenursing. Conversely, we anticipate that barriers to telenursing will demonstrate a negative correlation with self-efficacy levels. The third hypothesis proposes that nursing students' awareness, knowledge, skill perceptions, and attitudes toward telenursing will serve as significant predictors of both initial self-efficacy levels prior to training and the magnitude of change in self-efficacy following the intervention.

Methods

Participants

Participants in this study included 114 nursing students in their third year of studies. Most participants were female (101/114, 88.6%), between the ages of 20 and 50 years (mean 25.68, SD 4.59 years), and studying for a first undergraduate degree (107/114, 93.9%) (rather than attempting a career change).

XSL•FO
RenderX

Instruments

The questionnaire included 6 sections:

1. The Awareness of Telenursing questionnaire included 6 items, scored on a scale of 1-3 (1=know about it, 2=heard about it, and 3=know nothing about it). Sample item: "Telenursing is the most advanced service provided in nursing." Internal consistency was acceptable in the pretest (α=.70) but low in the posttest (α=.48). A higher mean score reflects higher awareness of telenursing.
2. The Knowledge of Telenursing questionnaire included 10 dichotomous items, scored 0 or 1. Sample item: "Epidemiological patient surveys can be conducted via telenursing." Low to acceptable internal consistencies were found in the pretest (α=.75) and the posttest (α=.54). A higher summary score reflects a higher level of knowledge of telenursing.
3. The Skills Required for Telenursing questionnaire included 10 items, scored on a scale of 1-5, with 1 indicating a low level and 5 indicating a very high level. Respondents were asked to rate the extent to which they felt that nurses need specific skills for using telenursing. Sample item: "High listening skills and high question asking skills are required for telenursing." High internal consistencies were found in the pretest (α=.91) and in the posttest (α=.90). A higher mean score reflects a higher level of skills required for telenursing.
4. The Attitudes About Telenursing questionnaire included 13 items, scored on a scale of 1-5, with 1 indicating that the respondent does not agree at all and 5 indicating a very high level of agreement. Sample item: "I believe that telenursing facilitates the provision of equitable service to all patients." High internal consistencies were found in the pretest (α=.89) and the posttest (α=.89). A higher mean score reflects more positive attitudes about telenursing.
5. The Barriers to Telenursing questionnaire included 9 items, scored on a scale of 1-5, with 1 indicating that the respondent does not agree at all and 5 indicating a very high level of agreement. Sample item: "I did not invest so many years studying just to work in front of a computer. I will miss personal contact with patients and meeting them face to face." Good internal consistencies were found in the pretest (α=.79) and the posttest (α=.82). A higher mean score reflects a higher level of barriers to using telenursing.
6. The Self-Efficacy in Telenursing questionnaire included 10 items, scored on a scale of 1-5, with 1 indicating that the respondent does not feel certain and 5 indicating that the respondent feels very certain in being able to help the patient follow instructions given over the telephone and understand complex cases presented in that manner. High internal consistencies were found in the pretest (α=.91) and the posttest (α=.95). A higher mean score reflects higher self-efficacy.

Procedure

This study is a cohort intervention study conducted among all third-year nursing students in the college, which examined level of knowledge, skills and attitudes regarding self-efficacy and clinical skills for telenursing, and willingness to use this method at 2 points in time—prior to and following training.

The training took place in two stages: (1) the students were taught by the course lecturer, who works in this field. Topics of study included diverse nursing practices, ethical aspects of telehealth, clinical skills including communications skills, challenges in telenursing, and tools for coping with complex issues arising from telenursing. (2) Students practiced telenursing through simulations in various nursing areas. During the simulations they practiced treating patients using the telenursing tools and communications skills they had learned and conducted virtual patient assessments and physical examinations.

The participating students answered a questionnaire assessing the research variables prior to and following training.

Ethical Considerations

The study was approved by the College Ethics Committee, Emek Yezreel Academic College (approval no. YVC EMEK-2023‐87). Students were recruited via the researcher's research assistant, who asked for volunteers. Before completing the questionnaires, participants were told that participation was voluntary and that they could drop out of the study at any time. They were informed that their opinions were important for constructing the departmental training program and were therefore encouraged to express them. Participants signed informed consent forms prior to participation. All students in the cohort agreed to participate in the study with no compensation provided. The participants' privacy and identity were protected, and confidentiality was assured in that no identifying information was asked. The study objectives were explained to the participants and the study was conducted according to the academic ethical code.

Data Analysis

The data were analyzed with SPSS (version 29; IBM Corp). Descriptive statistics were used for the participants' demographic characteristics and study variables. As the variable of skills required for telenursing (pre and post) was negatively skewed (preskewness −1.74, SE 0.23; postskewness −3.20, SE 0.23), it underwent exponential transformation. Time differences were assessed with 2-tailed paired *t* tests, using Cohen *d* for effect sizes. Change scores were computed as residual gain scores between the pre- and posttests, and Pearson correlations were calculated between the study variables regarding the pretest scores and the change scores. Multiple linear regressions were calculated for self-efficacy in telenursing, using pretest scores and change scores. Awareness, knowledge, skills, attitudes, and barriers to telenursing were defined as predictors.

Results

Descriptive Results

Most participants have initially reported a moderate (54/114, 47.4%) or advanced (58/114, 50.9%) level of knowledge in using computers and the web. Most have not been exposed to telenursing (104/114, 91.2%) but were interested in training in it (86/114, 75.4%). Participants' age was generally not

XSL•FO
RenderX

associated with the study variables (P=.07 to P=.94) and was thus not controlled for. Other demographic variables had low variance. Thus, the first hypothesis was assessed with a series of 2-tailed paired t tests. Significant changes were noted in all variables with moderate to high effect sizes (Table 1). Awareness of telenursing, knowledge of telenursing, skills, attitudes, and self-efficacy in telenursing have all significantly increased following participation in the training program, and barriers to telenursing have significantly decreased.

Table . Means, SDs, t values, and Cohen d values for the study variables by time (N=114)[a].

	Pretest, mean (SD)	Posttest, mean (SD)	t_{113} (P value)	Cohen d
Awareness of telenursing	1.75 (0.48)	2.54 (0.36)	15.18 (<.001)	1.85
Knowledge of telenursing	6.84 (2.48)	8.42 (1.55)	6.18 (<.001)	0.76
Skills required for telenursing	4.60 (0.55)	4.77 (0.38)	3.80 (<.001)	0.36
Attitudes about telenursing	3.41 (0.62)	4.06 (0.62)	9.78 (<.001)	1.04
Barriers to telenursing	2.94 (0.61)	2.63 (0.74)	−4.52 (<.001)	0.45
Self-efficacy in telenursing	3.69 (0.79)	4.04 (0.73)	4.88 (<.001)	0.46

[a]Ranges: awareness of telenursing 1-3; knowledge of telenursing 0-10; and skills, attitudes, barriers for telenursing, and self-efficacy in telenursing 1-5.

Pearson correlations were calculated among the study variables regarding the pretest and change scores. Significant associations were found (Table 2). In the pretest, higher awareness of telenursing, higher knowledge of telenursing, and perception of the higher skills required for telenursing were associated with more positive attitudes and higher self-efficacy regarding telenursing. Furthermore, more positive attitudes about telenursing were associated with higher self-efficacy in telenursing and with lower barriers to it.

XSL•FO
RenderX

Table . Pearson correlations between the study variables for the pretest scores and the change scores (N=114).

	1	2	3	4	5	6
Pretest						
1. Awareness of telenursing						
r	1	0.13	0.12	0.19	–0.13	0.20
P value	—[a]	.18	.19	.04	.16	.04
2. Knowledge of telenursing						
r	0.13	1	0.15	0.38	–0.06	0.23
P value	.18	—	.11	<.001	.53	.02
3. Skills required for telenursing						
r	0.12	0.15	1	0.22	0.03	0.19
P value	.19	.11	—	.02	.75	.04
4. Attitudes about telenursing						
r	0.19	0.38	0.22	1	–0.27	0.20
P value	.04	<.001	.02	—	.004	.04
5. Barriers to telenursing						
r	–0.13	–0.06	0.03	–0.27	1	0.06
P value	.16	.53	.75	.004	—	.51
6. Self efficacy in telenursing						
r	0.20	0.23	0.19	0.20	0.06	1
P value	.04	.02	.04	.04	.51	—
Change scores						
1. Awareness of telenursing						
r	1	0.25	–0.11	0.23	–0.22	0.14
P value	—	.006	.26	.01	.02	.14
2. Knowledge of telenursing						
r	0.25	1	0.01	0.23	–0.08	0.02
P value	.006	—	.95	.01	.42	.85
3. Skills required for telenursing						
r	–0.11	0.01	1	0.12	0.01	0.21
P value	.26	.95	—	.20	.93	.02
4. Attitudes about telenursing						
r	0.23	0.23	0.12	1	–0.38	0.37
P value	.01	.01	.20	—	<.001	<.001
5. Barriers to telenursing						
r	–0.22	–0.08	0.01	–0.38	1	–0.23
P value	.02	.42	.93	<.001	—	.02

XSL•FO
RenderX

	1	2	3	4	5	6
6. Self efficacy in telenursing						
r	0.14	0.02	0.21	0.37	–0.23	1
P value	.14	.85	.02	<.001	.02	—

[a]Not applicable.

Regarding the change scores, higher awareness of telenursing was associated with higher knowledge of telenursing and both were associated with more positive attitudes regarding it. Furthermore, higher awareness of telenursing and more positive attitudes regarding it were associated with lower barriers to telenursing. Finally, the higher skills required for telenursing, more positive attitudes about it, and lower barriers were associated with higher self-efficacy in telenursing.

Associations With and Change in Self-Efficacy

Two multiple linear regressions were calculated to evaluate the associations between awareness, knowledge, skills, attitudes and barriers to telenursing, and self-efficacy in telenursing regarding the pretest and the change between the pretest and the posttest. Level of knowledge in using the computer and the web (0: moderate and low, 1: advanced) was entered first, and the study variables or change in the study variables was entered second (Table 3).

Table . Multiple linear regressions for self-efficacy in telenursing with awareness, knowledge, skills, and attitudes and barriers to telenursing (N=114).

	Pretest scores[a]			Change scores[b]		
	B (SE)	β	*P* value	*B* (SE)	β	*P* value
Level of computer knowledge (advanced)	0.36 (0.14)	.23	.01[c]	−0.10 (0.18)	−.05	.58
Awareness of telenursing	0.32 (0.14)	.19	.03[c]	0.03 (0.10)	.03	.73
Knowledge of telenursing	0.07 (0.03)	.22	.02[c]	−0.10 (0.09)	−.10	.27
Skills required for telenursing	0.01 (0.01)	.06	.54	0.19 (0.09)	.19	.04[c]
Attitudes about telenursing	0.21 (0.12)	.17	.09	0.36 (0.10)	.36	<.001[c]
Barriers to telenursing	0.07 (0.11)	.06	.53	−0.09 (0.10)	−.09	.34

[a]R^2=0.23, P<.001; $F_{6, 107}$=5.20, P<.001.

[b]R^2=0.21, P<.001; $F_{6, 107}$=4.64, P<.001.

[c]These values are significant.

Both regression models were found significant, with 23% and 21% of the variance in the pretest score and in the change score, respectively, being explained by them. Regarding the pretest score, higher awareness and more knowledge of telenursing were associated with the perception of higher self-efficacy in telenursing. Regarding the change score, greater improvement in the perceived skills required for telenursing and a higher positive change in the attitudes regarding telenursing were associated with a greater improvement in the perception of self-efficacy in telenursing.

Discussion

This study aimed to examine how a simulation training program on telenursing affected awareness, knowledge, skills, attitudes, self-efficacy, and perceived barriers regarding telenursing among third-year nursing students. The results demonstrate significant improvements across all measured variables, with moderate to high effect sizes, suggesting that the implemented training program was effective. Moreover, the higher skills required for telenursing, more positive attitudes regarding it, and lower barriers were associated with higher self-efficacy in telenursing. These findings emphasize that the simulated experiences served as effective interventions, providing students with innovative learning opportunities [33].

The substantial increase in participants' awareness and knowledge of telenursing reflects the growing recognition that it is a critical component of modern health care delivery [1,3]. This increase is particularly noteworthy, given that most participants (104/114, 91.2%) had no prior exposure to telenursing, despite their initial moderate to advanced levels of computer and web proficiency. Findings regarding the posttest score of awareness of telenursing and its change score may be biased in unknown ways and should be regarded with caution.

XSL•FO
RenderX

Vaidya [34] recently emphasized the need to offer simulation telehealth education to undergraduate, graduate and health care practitioners in an effort to achieve a more effective remote diagnosis and treatment management for patients in need, such as those living with chronic disease.

The findings of this study are also in line with those of Mun et al [35], which indicated that nursing students lacked substantial awareness regarding telenursing. Nevertheless, the results also portrayed a positive outlook. Indeed, according to Kazawa et al [36], engaging in telenursing helps students enhance their understanding of telehealth practices, develop critical thinking skills, and broaden their knowledge of how to manage and address patient needs in a virtual care setting. Chang et al [37] also found that nurses with telehealth experience have significantly higher perceptions of its usefulness than those with no such experience, and these perceptions correlated positively with attitudes and behavioral intentions. These findings imply that providing nursing students with telenursing education can help them understand and harness this method [28]. Moreover, telenursing education was shown to have a significant impact on nurses' knowledge, attitudes, and awareness of future work [8,11]. Telehealth simulation was shown to improve nursing students' professional skills [38].

This study goes a step further by demonstrating that targeted training can significantly improve attitudes. This finding is crucial, as positive attitudes are likely to translate into greater willingness to engage with and implement telenursing practices in future professional roles. Nurses with prior telehealth knowledge had more positive attitudes toward telenursing than those who had never encountered telehealth-related information, and their attitudes toward telenursing correlated positively with their intentions to engage in telehealth [37]. Moreover, nursing students' attitudes toward telenursing demonstrated a significant correlation with telenursing experience, observation of telenursing during clinical practice, and exposure to telenursing education [35].

The decrease in perceived barriers to telenursing following the training program is a particularly encouraging outcome of this study. It suggests that the program has effectively addressed common concerns and obstacles associated with telenursing implementation and has the potential to offer a smoother integration of these practices in future health care settings.

Indeed, to prepare for their future roles, nursing students need telenursing education [35]. Previous studies examining nursing students reported positive prospects for telenursing, alongside negative perceptions associated with a lack of awareness [28,39]. Furthermore, Mun et al [35] found that nursing students who had a negative outlook regarding telenursing noted its impracticality as compared with face-to-face nursing, as well as the lack of patient contact, challenges faced by older individuals, and accessibility issues for low-income or rural residents. The assumption is that these limited perceptions derive from a lack of formal education in telenursing [40]. This type of education has the potential to enhance knowledge and attitudes regarding telenursing [41]. Indeed, significant improvements in understanding the use and role of telenursing were found among individuals who had undergone telehealth education [42].

This study found that skills and self-efficacy improved following the intervention. These findings are in line with a previous study that found a significant enhancement in skills and self-efficacy following training [37], thus indicating that telenursing education plays a crucial role in improving the type of specialized knowledge required for clinical telenursing. These findings suggest that nursing students need formal education in telenursing. Such education will enhance their competency and nurture a positive attitude, facilitating the seamless integration of telenursing into the digital health care era [35]. Our findings also corroborate those of Reierson et al [17], who emphasized the importance of introducing and practicing telenursing at the beginning of nursing curricula.

Moreover, our study found differences in the predictors of self-efficacy. Prior to the intervention, self-efficacy was a function of awareness and knowledge, whereas following it self-efficacy correlated with a change in skills and attitudes. To the best of our knowledge, almost no intervention studies have been conducted on the role of education in promoting telenursing. Kazawa et al [36] found that telenursing education is essential in expanding nursing students' knowledge and skills. Moreover, Mun et al [35] found that self-efficacy regarding telenursing among nursing students was associated with telenursing experience, education, and attitudes toward telenursing. Knowledge regarding advance care planning was also found to be associated with self-efficacy [43,44]. Bandura [45,46] also found a relationship between knowledge and skills, which translates into action by increasing self-efficacy to overcome barriers, and Mata et al [47] found that skills can improve health professionals' performance and self-efficacy.

Former studies of simulation-based instruction in nursing and telehealth were done by Parmeter et al [48], using a posttest-only design. Following peer-to-peer telehealth simulation scenarios via Zoom (Zoom Video Communications, Inc), the students demonstrated a high score of confidence and telehealth performance. These results align with the findings of this study.

These findings have significant implications for nursing education and practice. First, they strongly support the integration of telenursing into nursing curricula, as advocated by Asimakopoulou [3] and Puro and Feyereisen [10]. Moreover, the success of the simulation-based training program in this study is in line with previous research highlighting the effectiveness of simulations in nursing education [20,49]. Simulation can also play an important role in helping students acquire and improve their self-efficacy and nursing skills [50]. Our findings suggest that similar approaches may be valuable in preparing nursing students for the growing prevalence of telenursing in health care delivery.

This study presents 3 key limitations. First, the absence of a control group limits the ability to conclusively attribute observed changes solely to the training program. Future research should incorporate a parallel control group design to enable a more rigorous comparative analysis and establish clearer causal relationships. Second, the focus on immediate posttraining outcomes prevents understanding of long-term telenursing

XSL•FO
RenderX

competency retention and clinical application. Implementing a longitudinal study design with multiple follow-up assessments at 3, 6, and 12 months posttraining would provide insights into knowledge and skill sustainability. Third, the study's recruitment from a single college limits the generalizability of the findings across nursing student populations in Israel. Expanding the research to multiple educational institutions, potentially including diverse geographic and institutional contexts, would enhance the external validity of the results. Cronbach α value for awareness of telenursing was acceptable at pretest (α=0.70) but low at posttest (α=0.48). This finding indicates that the posttest score of awareness of telenursing has a low reliability, and the relevant findings should be regarded with caution. That is, the findings regarding the posttest score of awareness of telenursing and its change score may be biased in unknown ways and should be regarded with caution. Future studies are advised to validate a modified version of this questionnaire or use a different one.

Future research should address these methodological limitations by integrating control groups, longitudinal designs, and multi-institutional sampling to comprehensively evaluate telenursing training programs and their broader implementation potential. In addition, research exploring the implications of telenursing training on patient outcomes and health care system efficiency would provide critical empirical evidence to support broader program implementation.

In conclusion, this study provides compelling evidence for the effectiveness of a simulation-based telenursing training program in enhancing nursing students' competencies across multiple domains of telenursing. The findings underscore the importance of integrating telenursing education into nursing curricula in an effort to prepare future health care providers for the evolving landscape of health care delivery.

To implement these findings effectively, health care organizations should provide hands-on telenursing training through structured workshops and regular skill refresher seminars for practicing nurses. The Ministry of Health should establish standardized telenursing protocols and mandate their adoption across all health maintenance organizations to ensure consistent quality of care. Medical schools should integrate practical telenursing simulations into their core curriculum, while practicing health care professionals should complete required continuing education modules which include hands-on simulation training to maintain their competency in virtual care delivery.

Acknowledgments

The authors thank the students for their participation in the course and their responses to the questionnaire before and after it.

Data Availability

All data generated or analyzed during this study are included in this published article.

Authors' Contributions

OAS, LM, and OH contributed to conceptualization, methodology, and formal analysis and investigation; OAS and LM contributed to writing—original draft preparation; OH contributed to writing—review and editing; and OAS and OH contributed to resources and supervision.

Conflicts of Interest

None declared.

References

1. Global diffusion of eHealth: making universal health coverage achievable: report of the third global survey on eHealth. World Health Organization. 2016. URL: https://www.who.int/publications/i/item/9789241511780 [accessed 2025-02-17]
2. Schlachta L, Sparks S. Definitions of telenursing, telemedicine. In: Fitzpatrick J, editor. Encyclopedia of Nursing Research: Springer Publishing Inc; 1998:558-559.
3. Asimakopoulou E. Telenursing in clinical practise and education [Editorial]. Int J Caring Sci 2020;13(2):781-782 [FREE Full text]
4. Developing Telehealth Protocols: A Blueprint for Success: American Nurses Association; 2001.
5. Souza-Junior VD, Mendes IAC, Mazzo A, Godoy S. Application of telenursing in nursing practice: an integrative literature review. Appl Nurs Res 2016 Feb;29:254-260. [doi: 10.1016/j.apnr.2015.05.005]
6. Kamei T. Telenursing and artificial intelligence for oncology nursing. Asia Pac J Oncol Nurs 2022 Dec;9(12):100119. [doi: 10.1016/j.apjon.2022.100119] [Medline: 36276880]
7. Ghoulami-Shilsari F, Esmaeilpour Bandboni M. Tele-nursing in chronic disease care: a systematic review. Jundishapur J Chronic Dis Care 2019;8(2):1-8. [doi: 10.5812/jjcdc.84379]
8. Walton L, Courtright K, Demiris G, Gorman EF, Jackson A, Carpenter JG. Telehealth palliative care in nursing homes: a scoping review. J Am Med Dir Assoc 2023 Mar;24(3):356-367. [doi: 10.1016/j.jamda.2023.01.004] [Medline: 36758619]

XSL•FO
RenderX

9. Alipour J, Hayavi-Haghighi MH. Opportunities and challenges of telehealth in disease management during COVID-19 pandemic: a scoping review. Appl Clin Inform 2021 Aug;12(4):864-876. [doi: 10.1055/s-0041-1735181] [Medline: 34528234]
10. Puro NA, Feyereisen S. Telehealth availability in US hospitals in the face of the COVID-19 pandemic. J Rural Health 2020 Sep;36(4):577-583. [doi: 10.1111/jrh.12482] [Medline: 32603017]
11. Poreddi V, Bidadi Veerabhadraiah K, Reddy S, Narayana M, Channaveerachari N, BadaMath S. Nursing interns' perceptions of telenursing: implications for nursing education. Telehealth Med Today 2021;6(2). [doi: 10.30953/tmt.v6.258]
12. Koivunen M, Saranto K. Nursing professionals' experiences of the facilitators and barriers to the use of telehealth applications: a systematic review of qualitative studies. Scand J Caring Sci 2018 Mar;32(1):24-44. [doi: 10.1111/scs.12445] [Medline: 28771752]
13. Jácome M, Rego N, Veiga P. Potential of a nurse telephone triage line to direct elderly to appropriate health care settings. J Nurs Manag 2019 Sep;27(6):1275-1284. [doi: 10.1111/jonm.12809]
14. Sexton V, Dale J, Bryce C, Barry J, Sellers E, Atherton H. Service use, clinical outcomes and user experience associated with urgent care services that use telephone-based digital triage: a systematic review. BMJ Open 2022 Jan 3;12(1):e051569. [doi: 10.1136/bmjopen-2021-051569] [Medline: 34980613]
15. Edirippulige S, Smith AC, Young J, Wootton R. Knowledge, perceptions and expectations of nurses in e-health: results of a survey in a children's hospital. J Telemed Telecare 2006 Nov;12(3_suppl):35-38. [doi: 10.1258/135763306779380255]
16. Ali NS, Carlton KH, Ali OS. Telehealth education in nursing curricula. Nurse Educ 2015;40(5):266-269. [doi: 10.1097/NNE.0000000000000149] [Medline: 25689080]
17. Reierson I, Solli H, Bjørk IT. Nursing students' perspectives on telenursing in patient care after simulation. Clin Simul Nurs 2015 Apr;11(4):244-250. [doi: 10.1016/j.ecns.2015.02.003]
18. Lee J, Oh PJ. Effects of the use of high-fidelity human simulation in nursing education: a meta-analysis. J Nurs Educ 2015 Sep;54(9):501-507. [doi: 10.3928/01484834-20150814-04] [Medline: 26334336]
19. Oh PJ, Jeon KD, Koh MS. The effects of simulation-based learning using standardized patients in nursing students: a meta-analysis. Nurse Educ Today 2015 May;35(5):e6-e15. [doi: 10.1016/j.nedt.2015.01.019] [Medline: 25680831]
20. Filomeno L, Minciullo A. High-fidelity simulation for the education of nursing students: a scoping review of the literature. Prof Inferm 2021;74(2):180-189. [doi: 10.7429/pi.2021.7423180] [Medline: 35084162]
21. Lestander Ö, Lehto N, Engström Å. Nursing students' perceptions of learning after high fidelity simulation: effects of a three-step post-simulation reflection model. Nurse Educ Today 2016 May;40:219-224. [doi: 10.1016/j.nedt.2016.03.011] [Medline: 27125176]
22. van Houwelingen CTM, Moerman AH, Ettema RGA, Kort HSM, Ten Cate O. Competencies required for nursing telehealth activities: a Delphi-study. Nurse Educ Today 2016 Apr;39:50-62. [doi: 10.1016/j.nedt.2015.12.025] [Medline: 27006033]
23. Glinkowski W, Pawłowska K, Kozłowska L. Telehealth and telenursing perception and knowledge among university students of nursing in Poland. Telemedicine and E-Health 2013 Jul;19(7):523-529. [doi: 10.1089/tmj.2012.0217]
24. Honey M, Collins E, Britnell S. Education into policy: embedding health informatics to prepare future nurses—New Zealand case study. JMIR Nurs 2020;3(1):e16186. [doi: 10.2196/16186] [Medline: 34345779]
25. Malhotra P, Ramachandran A, Chauhan R, Soni D, Garg N. Assessment of knowledge, perception, and willingness of using telemedicine among medical and allied healthcare students studying in private institutions. Telehealth Med Today 2020;5(4). [doi: 10.30953/tmt.v5.228]
26. Abraham C, Jensen C, Rossiter L, Dittman Hale D. Telenursing and remote patient monitoring in cardiovascular health. Telemed J E Health 2024 Mar;30(3):771-779. [doi: 10.1089/tmj.2023.0187] [Medline: 37682280]
27. El- Said Abd Ellatif A, Mohamed Sobhy Elsayed D, Hamido Abosree T. Knowledge and attitude of faculty of nursing students regarding telenursing. J Nurs Sci 2023 Jan 1;4(1):677-689. [doi: 10.21608/jnsbu.2023.278954]
28. Khraisat OMA, Al-Bashaireh AM, Alnazly E. Telenursing implications for future education and practice: nursing students' perspectives and knowledge from a course on child health. PLoS One 2023;18(11):e0294711. [doi: 10.1371/journal.pone.0294711] [Medline: 38011137]
29. Kurtović B, Hošnjak AM, Ledinski S, Smrekar M, Babić J, Čukljek S. Nursing students' knowledge and attitudes towards telenursing. Croat Nurs J (Online) 2024;8(1):5-16. [doi: 10.24141/2/8/1/1]
30. Assaye BT, Belachew M, Worku A, et al. Perception towards the implementation of telemedicine during COVID-19 pandemic: a cross-sectional study. BMC Health Serv Res 2023;23(1):1186. [doi: 10.1186/s12913-023-09927-1]
31. Ftouni R, AlJardali B, Hamdanieh M, Ftouni L, Salem N. Challenges of telemedicine during the COVID-19 pandemic: a systematic review. BMC Med Inform Decis Mak 2022 Aug 3;22(1):207. [doi: 10.1186/s12911-022-01952-0] [Medline: 35922817]
32. Ghaddaripouri K, Mousavi Baigi SF, Abbaszadeh A, Mazaheri Habibi MR. Attitude, awareness, and knowledge of telemedicine among medical students: a systematic review of cross-sectional studies. Health Sci Rep 2023 Mar;6(3):e1156. [doi: 10.1002/hsr2.1156] [Medline: 36992712]
33. Dziurka M, Machul M, Ozdoba P, et al. Clinical training during the COVID-19 pandemic: experiences of nursing students and implications for education. Int J Environ Res Public Health 2022 May 23;19(10):6352. [doi: 10.3390/ijerph19106352] [Medline: 35627889]

XSL•FO
RenderX

34. Vaidya A. Developing simulation-based telehealth training for Next-Gen nurses. TechTarget. 2024. URL: https://www.techtarget.com/virtualhealthcare/answer/Developing-Simulation-Based-Telehealth-Training-for-Next-Gen-Nurses [accessed 2025-01-23]
35. Mun M, Choi S, Woo K. Investigating perceptions and attitude toward telenursing among undergraduate nursing students for the future of nursing education: a cross-sectional study. BMC Nurs 2024 Apr 8;23(1):236. [doi: 10.1186/s12912-024-01903-2] [Medline: 38589885]
36. Kazawa K, Teramoto C, Azechi A, Satake H, Moriyama M. Undergraduate nursing students' learning experiences of a telehealth clinical practice program during the COVID-19 pandemic: a qualitative study. Nurse Educ Today 2022 Apr;111:105297. [doi: 10.1016/j.nedt.2022.105297] [Medline: 35182935]
37. Chang MY, Kuo FL, Lin TR, Li CC, Lee TY. The intention and influence factors of nurses' participation in telenursing. Informatics (MDPI) 2021;8(2):35. [doi: 10.3390/informatics8020035]
38. Moore J, Jairath N, Montejo L, O'Brien S, Want D. Using a telehealth simulation to prepare nursing students for intraprofessional collaboration. Clin Simul Nurs 2023 May;78:1-6. [doi: 10.1016/j.ecns.2023.02.007]
39. Emikpe BO, Asare DA, Emikpe AO, Folitse RD, Botchway LN. Knowledge and perception of veterinary students in Ghana on telemedicine. Niger J Physiol Sci 2021 Jun 30;36(1):115-121. [Medline: 34987249]
40. Chike-Harris KE, Durham C, Logan A, Smith G, DuBose-Morris R. Integration of telehealth education into the health care provider curriculum: a review. Telemed J E Health 2021 Feb 1;27(2):137-149. [doi: 10.1089/tmj.2019.0261]
41. Pit SW, Velovski S, Cockrell K, Bailey J. A qualitative exploration of medical students' placement experiences with telehealth during COVID-19 and recommendations to prepare our future medical workforce. BMC Med Educ 2021 Aug 16;21(1):431. [doi: 10.1186/s12909-021-02719-3] [Medline: 34399758]
42. Wong CJ, Nath JB, Pincavage AT, et al. Telehealth attitudes, training, and preparedness among first-year internal medicine residents in the COVID-19 era. Telemed J E Health 2022 Feb 1;28(2):240-247. [doi: 10.1089/tmj.2021.0005]
43. ten Koppel M, Onwuteaka-Philipsen BD, van der Steen JT, et al. Care staff's self-efficacy regarding end-of-life communication in the long-term care setting: results of the PACE cross-sectional study in six European countries. Int J Nurs Stud 2019 Apr;92:135-143. [doi: 10.1016/j.ijnurstu.2018.09.019]
44. Evenblij K, ten Koppel M, Smets T, Widdershoven GAM, Onwuteaka-Philipsen BD, Pasman HRW. Are care staff equipped for end-of-life communication? A cross-sectional study in long-term care facilities to identify determinants of self-efficacy. BMC Palliat Care 2019 Dec;18(1):1-11. [doi: 10.1186/s12904-018-0388-z]
45. Bandura A. Self-Efficacy: The Exercise of Control: W H Freeman; 1997.
46. Bandura A. Social Foundations of Thought and Action: Prentice Hall; 1986.
47. Mata ÁDS, de Azevedo KPM, Braga LP, et al. Training in communication skills for self-efficacy of health professionals: a systematic review. Hum Resour Health 2021 Mar 6;19(1):30. [doi: 10.1186/s12960-021-00574-3] [Medline: 33676515]
48. Parmeter S, Foronda C, Lee J. Improving telenursing skills through simulation-based education. J Dr Nurs Pract 2023 Jun 27;2:93-101. [doi: 10.1891/JDNP-2022-0021] [Medline: 37369454]
49. Weissblueth E, Linder I. The effects of simulations in a simulation center on principals' training and professional self-efficacy. Int J Educ Policy Leadersh 2020;16(14):1-13. [doi: 10.22230/ijepl.2020v16n14a965]
50. Hayden JK, Smiley RA, Alexander M, Kardong-Edgren S, Jeffries PR. The NCSBN national simulation study: a longitudinal, randomized, controlled study replacing clinical hours with simulation in prelicensure nursing education. J Nurs Regul 2014 Jul;5(2):S3-S40. [doi: 10.1016/S2155-8256(15)30062-4]

Edited by J Gentges; submitted 21.10.24; peer-reviewed by LH Schroeder, MA Zoubi; revised version received 29.01.25; accepted 30.01.25; published 12.03.25.

Please cite as:
Ali-Saleh O, Massalha L, Halperin O
Evaluation of a Simulation Program for Providing Telenursing Training to Nursing Students: Cohort Study
JMIR Med Educ 2025;11:e67804
URL: https://mededu.jmir.org/2025/1/e67804
doi:10.2196/67804

© Ola Ali-Saleh, Layalleh Massalha, Ofra Halperin. Originally published in JMIR Medical Education (https://mededu.jmir.org), 12.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Distance Learning During the COVID-19 Lockdown and Self-Assessed Competency Development Among Radiology Residents in China: Cross-Sectional Survey

Peicheng Wang[1,2*], MSc; Ziye Wu[1*], MPH; Jingfeng Zhang[3], MD; Yanrong He[1], MPH; Maoqing Jiang[3], MD; Jianjun Zheng[3], MSc; Zhenchang Wang[4], MD; Zhenghan Yang[4], MD; Yanhua Chen[1,2,5], PhD; Jiming Zhu[1,6], DPhil

[1]Vanke School of Public Health, Tsinghua University, Haidian District, Beijing, China
[2]School of Medicine, Tsinghua University, Beijing, China
[3]Department of Radiology, Ningbo No. 2 Hospital, Ningbo, China
[4]Department of Radiology, Beijing Friendship Hospital, Capital Medical University, Beijing, China
[5]Department of Public Health, Policy and Systems, University of Liverpool, Liverpool, United Kingdom
[6]Institute for Healthy China, Tsinghua University, Beijing, China
[*]these authors contributed equally

Corresponding Author:
Jiming Zhu, DPhil
Vanke School of Public Health, Tsinghua University, Haidian District, Beijing, China

Abstract

Background: During the COVID-19 lockdown, it was difficult for residency training programs to conduct on-site, hands-on training. Distance learning, as an alternative to in-person training, could serve as a viable option during this challenging period, but few studies have assessed its role.

Objective: This study aims to investigate the impact of distance learning during the lockdown on residents' self-assessed competency development and to explore the moderating effect of poor mental health on the associations. It is hypothesized that radiology residents who were trained through distance learning during the lockdown were more likely to report higher self-assessed competency compared to those who did not receive organized, formal training.

Methods: A cross-sectional survey was conducted in 2021 among all of the radiology residents in 407 radiology residency programs across 31 provinces of China. To estimate the long-term outcomes of radiology residents' training after the initial COVID-19 outbreak, this study measured 6 core competencies developed by the US Accreditation Council for Graduate Medical Education reported by radiology residents. Multiple linear regression and moderating effect analysis were conducted to examine the associations between distance learning, mental health status, and self-assessed competencies. Mental health status moderated the association between distance learning and self-assessed competency of radiology residents.

Results: A total of 2381 radiology residents (29.7% of the 8,008 nationwide) met the inclusion criteria and were included in the analysis. Among them, 71.4% (n=1699) received distance learning during the COVID-19 lockdown, and 73.2% (n=1742) reported mental health struggles ranging in severity from slight to extremely severe. Radiology residents who were trained through distance learning (β=0.35, 90% CI 0.24 - 0.45) were more likely to report higher self-assessed competencies. This was particularly true for the competency of "interpersonal and communication skills" (β=0.55, 90% CI 0.39 - 0.70). Whereas, the competency of "patient care and technical skills" (β=0.14, 90% CI 0.01 - 0.26) benefited the least from distance learning. Poor mental health significantly moderated the relationship between distance learning and competency (β=−0.15, 90% CI −0.27 to −0.02).

Conclusions: Distance learning, a means of promoting enabling environments during the COVID-19 lockdown, serves its purpose and helps generally improve residents' self-assessed competencies, though different competency domains benefit unequally. The impact of mental health status calls for special attention so that distance learning can fulfill its potential.

(JMIR Med Educ 2025;11:e54228) doi:10.2196/54228

KEYWORDS

radiology residents; distance learning; mental health status; self-assessed competency; ACGME competencies; Accreditation Council of Graduate Medical Education

XSL•FO
RenderX

Introduction

The COVID-19 pandemic threatens the health of people globally and has brought unprecedented pressure to health systems [1,2]. The national public health system plays a vital role in fighting against pandemics by taking measurements such as surveillance and epidemiological investigations [3], case finding and management [4], and collective quarantine of close contacts [5]. However, potential challenges including insufficient alerts, low efficiency of reporting to higher authorities, and workforce shortages still exist [1]. Of note, the training of health care providers and the improvement of their professional skills have been underscored for their great significance in medical service delivery and health systems resilience [6-8].

Residency training systems serve the purpose of cultivating a qualified health workforce [9]. Standardized residency training (SRT) was initiated in 2013 in China, aiming to train doctors to meet the needs of population health [10]. With the increasing trend of competency-based medical training in global medical education, the assessment of competencies has gained ground in practice [11]. The US Accreditation Council of Graduate Medical Education (ACGME) identified 6 core competencies for physicians (ie, patient care [PC], medical knowledge [MK], system-based practice [SBP], practice-based learning and improvement [PBLI], professionalism [PROF], and interpersonal communication skill [ICS]) [11] and implemented milestones by the Next Accreditation System initiative in July 2013 [12]. Residency education and competency-based practice assessed by the milestones are common requirements of ACGME and have been used in residency training in China [9,13].

COVID-19 has changed medical education dramatically, especially during the lockdown period. The impacts of COVID-19 on medical education in radiology, surgery, and emergency medicine have gained attention [14-16]. Radiology is related to other medical specialties and all levels of health care delivery [17]. Radiology residents were typically required to rotate between different specialties to obtain knowledge and clinical skills [18]. The mandatory social distancing challenged the traditional training on radiology trainee approaches such as teaching at workstations [19]. In China, SRT in radiology spans 3 years and involves workstation-based training throughout rotations in various specialties and departments [9]. In the first year, residents undergo rotations in the departments of radiology, ultrasound medicine, nuclear medicine, pathology, and relevant clinical departments. In the second and third years, they receive advanced rotational training within radiology subspecialties such as computed tomography, magnetic resonance imaging, x-ray, and interventional radiology [9]. Due to COVID-19, the mode of residency training has been switched from traditional in-person classes to distance learning [20], posing challenges to the effectiveness of rotational training and the developing competency of residents.

Numerous benefits have been found for distance learning. For instance, residents can schedule more flexibly and access the courses more easily [21]. They can learn at their own pace with the help of recorded lectures and communicate with professionals and peers on the web at their own convenience [22]. The positive acceptance and a higher level of satisfaction with distance learning have been reported by residents in Canada and the United States [21,23]. However, the practice of digital readout in distance learning is similar to the experience of in-person reading, in addition to the difficulties of gauging body language during practical operations or in the use of medical instruments [24], which may lead to unsatisfactory outcomes in radiology education. Meanwhile, the COVID-19 pandemic has impacted the mental health status of health care professionals dramatically [25]. Students who experienced distance learning during the pandemic had been found to have psychological distress [26,27]. According to the Job Demands-Resources model, mental health is a personal resource that helps residents to deal with job challenges by moderating their performance [28-31]. Accordingly, it is of great importance to take care of the mental health of residents who have experienced distance learning [32]. To date, during the COVID-19 lockdown, when workstation-based training was difficult to deliver, the role of distance learning remains unclear. It is also uncertain whether psychological status affects the effectiveness of distance learning.

In sum, the COVID-19 lockdown had brought substantial challenges to radiology training programs, which had to transition from face-to-face instruction to remote learning. In the meantime, mental distress caused by factors such as social distancing may make residency training rather difficult. Given that distance learning remains a primary alternative when traditional teaching is not feasible (such as during pandemic outbreaks and lockdowns), yet few studies have explored its effects, we aimed to investigate the impact of distance learning on the development of self-assessed competencies as well as the moderating effect of mental health status. We hypothesize that radiology residents who received distance learning during the COVID-19 lockdown were more likely to report higher self-assessed competencies compared to those who did not receive organized, formal training during the same period (ie, nondistance learners) and that this association was moderated by poor mental health. To test this, we used a nationwide survey dataset of radiology residents in China, which collected information on distance learning and mental health status during the lockdown (January-May 2020), and self-assessed competencies 6 months later. Previous studies have shown a strong positive correlation between the assessments by Clinical Competency Committees and residents' self-evaluations using the milestones. This suggests that residents are generally able to accurately assess their own competencies, which in turn supports the validity of using milestone assessments as an effective measure of self-assessed competency in this study [33].

Methods

Ethical Considerations

The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the institution review board of Tsinghua University, China (20210140). Informed consent was obtained before the research started, and data were deidentified to ensure participant privacy. Participation in this

XSL•FO
RenderX

survey was voluntary, participants did not receive any incentives to take part in the study.

Study Setting and Population

A nationwide retrospective cross-sectional survey was conducted on the web by the Chinese Association of Radiologists (CAR) during December 1, 2020, to April 30, 2021, targeting all the radiology residents in 407 radiology residency programs across 31 provinces of China.

To complete the distribution of the questionnaire and to ensure the participation of the radiology residents, we contacted the directors of the targeted hospitals' radiology departments by email or telephone to inform them of the purpose and details of the survey initiated by the CAR. The directors were then instructed to share the link of the questionnaire posted on the popular web-based survey platform "Wenjuanxing" with radiology residents. Anonymous responses to the questionnaire were submitted. All participants were informed that the questionnaire could only be filled out once, that participation was completely voluntary, and that they could withdraw at any time without penalty. During the 5-month survey period, the research team cooperated well with the CAR for the monitoring of the participation. If there was a low rate of submission, the CAR would require the hospital to improve the quality and quantity of their survey response. This proactive approach could help increase the number of responses and the representativeness of the population. Residents who did not undergo residency training during the COVID-19 epidemic between January and May 2020 would be excluded.

Measures and Outcomes

Distance Learning

Distance learning was asked "Was your training institution changed the teaching arrangements to the form of distance learning to reduce the negative impact during the initial COVID-19 outbreak?" The response options were either "yes or no." In contrast, nondistance learning participants (ie, nondistance learners) were those who did not receive organized, formal training during the same period. They primarily stayed at home on leave, reported daily health status, and engaged in delayed teaching plans or self-directed learning.

Mental Health Status

Mental health status was measured by the question "Did you suffer psychological distress during the initial COVID-19 outbreak? (from January to May 2020 in China)." A 5-point Likert scale was used to measure the degree of mental health struggles: 1=no impact, 2=mild impact, 3=moderate impact, 4=severe impact, and 5=extremely severe impact. Radiology residents can choose any rating between 1 and 5 (single choice). The variable is correlated with long-term mental health (depression and burnout) measured by the Depression and Anxiety Stress Scale—Depression and Maslach Burnout Inventory scales (Multimedia Appendix 1). The Cronbach α reliability coefficient for depression and burnout was 0.930 and 0.957, respectively.

Self-Assessed Competency

Milestone-based assessment of competencies for residence is one of the common requirements of the ACGME [11]. Self-assessment plays a key role in this process by fostering reflection on professional actions, identifying learning needs, and enabling residents to develop and refine personalized improvement plans [34]. Moreover, residents' self-assessments showed a strong alignment with the Clinical Competency Committee evaluations across postgraduate year levels [33,35,36]. To estimate the long-term outcomes of radiology residents' training after the initial COVID-19 outbreak, our study measured 6 core competencies developed by the ACGME that were assessed by radiology residents themselves. As is suggested by the experts, we selected 9 subcompetencies from diagnostic radiology milestones to represent the 6 ACGME core competencies: 2 PC subcompetencies, 2 MK subcompetencies, 2 PBLI subcompetencies, 1 SBP subcompetency, 1 PROF subcompetency, and 1 ICS subcompetency.

A dedicated section of the questionnaire is designed to assess 9 subcompetencies with 9 single-choice questions. Radiologists are able to select a score ranging from 0 to 9 for each competency. Examples of milestone sets for each subcompetency are shown in Figure 1. The primary outcome was self-evaluation milestone (SEM) scores (range 0 - 9 scores) for 9 subcompetencies and the average SEM scores.

XSL•FO
RenderX

Figure 1. Milestone sets for patient care 1 (image interpretation) and professionalism (self-awareness and help-seeking).

(Clinical context) Patient care 1: image interpretation				
Level 1	Level 2	Level 3	Level 4	Level 5
Identifies primary imaging findings	Identifies secondary and critical imaging findings and formulates differential diagnoses	Prioritizes differential diagnoses and recommends management options	Provides a single diagnosis with integration of current guidelines to recommend management, when appropriate	Demonstrates expertise and efficiency at a level expected of a subspecialist
Scores: 1 2 3 4 5 6 7 8 9				
Comments:		Not yet completed level 1		
		Not yet assessable		
(Nonclinical context) Professionalism: self-awareness and help-seeking				
Level 1	Level 2	Level 3	Level 4	Level 5
Recognizes status of personal and professional well-being, with assistance, and is aware of available resources	Independently recognizes status of personal and professional well-being using available resources when appropriate	With assistance, proposes a plan to optimize personal and professional well-being	Independently develops a plan to optimize personal and professional well-being	Coaches others when emotional responses or limitations in knowledge or skills do not meet professional expectations
Recognizes limits in the knowledge or skills of self or team, with assistance	Independently recognizes limits in the knowledge or skills of self or team and demonstrates appropriate help-seeking behaviors	With assistance, proposes a plan to remediate or improve limits in the knowledge or skills of self or team	Independently develops a plan to remediate or improve limits in the knowledge or skills of self or team	
Scores: 1 2 3 4 5 6 7 8 9				
Comments:		Not yet completed level 1		

Sociodemographic Characteristics

The sociodemographic information included age (≤27 or >27 years), sex (male or female), educational level (bachelor's degree or master's degree or above), training year (the second year or the third year), training sites level (grade-a tertiary general hospital, grade-a tertiary specialized hospital, grade-b tertiary general hospital, or others), undergraduate major (clinical medicine, medical imaging, or others), working hours per week (≤40, 40 - 48, or >48), annual after-tax income in 2020 (analyzed as continuous variable), and types of residents (professional master or nonprofessional master).

Statistical Analysis

The SEM scores of radiology residents for 9 subcompetencies were reported by means and SDs). The differences in SEM scores of residents between distance learning and nondistance learning were compared using independent samples 2-tailed *t* test. To explore the association between distance learning and competencies, multiple linear regression (MLR) models were constructed. The dependent variables were SEM scores of 9 subcompetencies, and the key explanatory variable was distance learning. The moderating effect of mental health on distance learning was explored by the MLR model. The significance of the moderating effect was tested by simple slope analysis. All models were controlled for participants' characteristics. A variance inflation factor was used to detect the multicollinearity of independent variables for all models (variance inflation factor scores <3). A *P* value of <.05 was considered statistically significant for 2-tailed tests (*t* test or chi-square test). A conservative level of *P* value of <.10 was used to assess potential moderators in the regression, which was reported by a coefficient (β) and 90% CIs [37,38]. All statistical analyses were performed by STATA (version 17.0; StataCorp LLC).

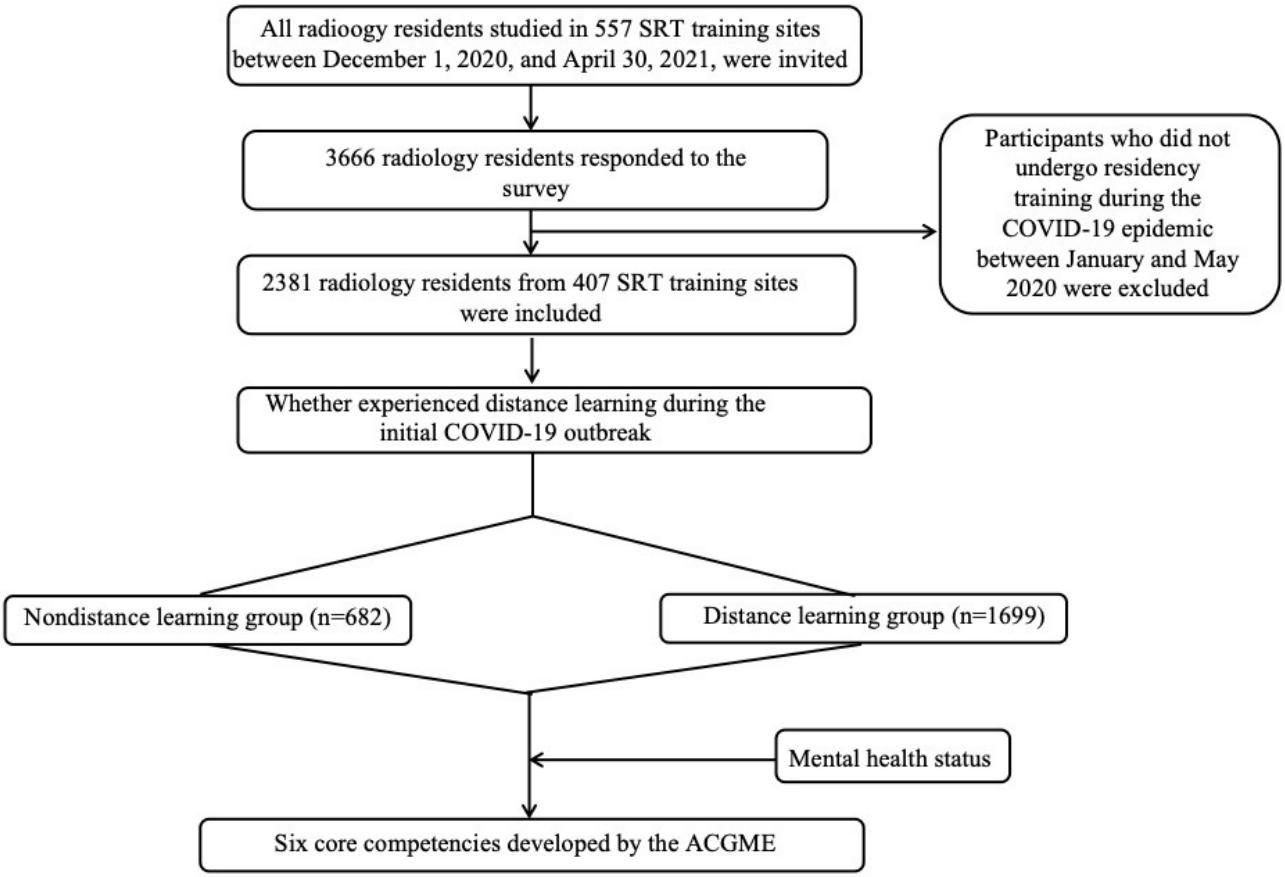
Results

Participants' Characteristics

Of the 8008 targeted radiology residents, 2381 (overall effective response rate: 29.7%) participated in this survey (Figure 2). As is shown in Table 1, the mean age of the participants was 27.8 (SD 2.4) years. In total, 58.5% (n=1392) of them were female, and 50.5% (n=1202) were in the third-year training. The majority of the participants received training in a grade-a tertiary hospital (n=2310, 97%), had a bachelor's degree (n=2187, 91.9%), and their undergraduate major was medical imaging (n=2016, 84.7%). The median annual after-tax income was 40,000 RMB (IQR 10,000 - 60,000; a currency exchange rate of 1 RMB=US $0.145 is applicable), with 25.8% (n=614) of them earning more than 60,000 RMB (about US $ 8698.9). The average working hours per week was 44.3 (SD 12.5) hours, and 23.1% (n=551) of the participants worked more than 48 hours per week. During the initial COVID-19 outbreak from January 2020 to May 2020, 71.4% (n=1699) of the radiology residents participated in distance learning, and 73.2% (n=1742) of them reported slight or severe mental health struggles. In total, 35.8% (n=853) of the participants contributed to the prevention and control of COVID-19.

XSL•FO
RenderX

Figure 2. Flowchart of the study design. ACGME: Accreditation Council for Graduate Medical Education; SRT: standardized residency training.



XSL•FO
RenderX

Table . Characteristics of participants in China.

Variables	Total (N=2381)	Distance learning		*P* value
		Yes (n=1699)	No (n=682)	
Region, n (%)				.36
East	945 (36.7)	663 (70.2)	282 (29.9)	
Central	496 (20.8)	358 (72.2)	138 (27.8)	
West	788 (33.1)	561 (71.2)	227 (28.8)	
Northeast	152 (6.4)	117 (77)	35 (23)	
Age (years), mean (SD)	27.8 (2.4)	27.8 (2.4)	27.7 (2.4)	.28
≤27, n (%)	1293 (54.3)	915 (70.8)	378 (29.2)	.49
>27, n (%)	1088 (45.7)	784 (72.1)	304 (27.9)	.49
Sex, n (%)				.02
Male	989 (41.5)	680 (68.8)	309 (31.2)	
Female	1392 (58.5)	1019 (73.2)	373 (26.8)	
SRT[a] training years, n (%)				.63
Second year	1179 (49.5)	836 (70.9)	343 (29.1)	
Third year	1202 (50.5)	863 (71.8)	339 (28.2)	
SRT sites level, n (%)				.34
Grade-A tertiary general hospital	2310 (97)	1649 (71.4)	661 (28.6)	
Grade-A tertiary specialized hospital	51 (2.1)	39 (76.5)	12 (23.5)	
Grade-B tertiary general hospital	15 (0.6)	8 (53.3)	7 (46.7)	
Others	5 (0.2)	3 (60)	2 (40)	
Education level, n (%)				.55
Bachelor's degree	2187 (91.9)	1557 (71.2)	630 (28.8)	
Master's or doctoral degree	194 (8.2)	142 (73.2)	52 (26.8)	
Undergraduate major, n (%)				.12
Clinical medicine	346 (14.5)	238 (68.8)	108 (31.2)	
Medical imaging	2016 (84.7)	1444 (71.6)	572 (28.4)	
Others	19 (0.8)	17 (89.5)	2 (10.5)	
Type of residents, n (%)				.007
Professional master	774 (32.5)	580 (74.9)	194 (25.1)	
Nonprofessional master	1607 (67.5)	1119 (69.6)	488 (30.4)	
Annual after-tax income (RMB[b]), median (IQR)	43,800 (10,000-67,000)	42,700 (9600-60,000)	46,600 (10,000-70,000)	.02
≤10,000, n (%)	691 (29)	513 (74.2)	178 (25.8)	.12
10,001 - 40,000, n (%)	565 (23.7)	400 (70.8)	165 (29.2)	.12
40,001 - 60,000, n (%)	511 (21.5)	367 (71.8)	144 (28.2)	.12
>60,000, n (%)	614 (25.8)	419 (68.2)	195 (31.8)	.12
Working hours per week (hours), mean (SD)	44.3 (12.5)	43.9 (11.8)	45.3 (14.0)	.01
≤40, n (%)	1311 (55.1)	948 (72.3)	363 (27.7)	.43
41 - 48, n (%)	519 (21.8)	369 (71.1)	150 (28.9)	.43

XSL•FO
RenderX

Variables	Total (N=2381)	Distance learning		*P* value
		Yes (n=1699)	No (n=682)	
>48, n (%)	551 (23.1)	382 (69.3)	169 (30.7)	.43
Mental health impact during the initial COVID-19 outbreak, n (%)				<.001
No impact	639 (26.8)	469 (73.4)	170 (26.6)	
Mild impact	1249 (52.5)	903 (72.3)	346 (27.7)	
Moderate impact	397 (16.7)	276 (69.5)	121 (30.5)	
Severe impact	79 (3.3)	45 (57)	34 (43)	
Extremely severe impact	17 (0.7)	6 (35.3)	11 (64.7)	
COVID-19–related work participation, n (%)				.90
Yes	853 (35.8)	610 (71.5)	243 (28.5)	
No	1528 (64.2)	1089 (71.3)	439 (28.7)	

[a]SRT: standardized residency training.

[b]A currency exchange rate of 1 RMB=US $0.145 is applicable.

SEM Scores of Radiology Residents Between Distance Learning and Nondistance Learning

The mean score of competencies and a comparison of subcompetencies scores were presented in Table 2. The overall average score of radiology residents' competency was 3.37 (SD 1.47). The average score of participants who received distance learning was 3.46 (SD 1.49), higher than those who did not (mean 3.13, SD 1.39; *P*<.001). Residents who received distance learning outperformed in all subcompetencies than those without distance learning during the initial COVID-19 outbreak (PC-1: *P*=.007; MK-1: *P*=.004; and others: *P*<.001), except for PC-2 (*P*=.09; Table 2). Due to the low response rate, Mann-Whitney tests were performed. The results were similar (Multimedia Appendix 2). For radiology residents who had not participated in COVID-19–related activities (1528/2381; Multimedia Appendix 3), the differences between residents' competencies or subcompetencies showed the same trends, except for the PC-1 (P=.10).

XSL•FO
RenderX

Table . Self-evaluation milestone scores for radiology residents between distance learning and nondistance learning.

Diagnostic radiology sub-competencies	Total (N=2381), mean (SD)	Distance learning		*P* value
		Yes (n=1699), mean (SD)	No (n=682), mean (SD)	
PC[a]				
PC-1: image interpretation	3.90 (1.69)	3.96 (1.68)	3.75 (1.72)	.007
PC-2: competence in procedures	2.25 (1.77)	2.28 (1.81)	2.16 (1.65)	.09
MK[b]				
MK-1: diagnostic knowledge	3.75 (1.75)	3.82 (1.76)	3.59 (1.73)	.004
MK-2: imaging technology and image acquisition	3.53 (1.90)	3.62 (1.92)	3.29 (1.81)	<.001
SBP[c]				
SBP-1: system navigation for patient-centered care	2.86 (1.88)	2.96 (1.90)	2.61 (1.80)	<.001
SBP-2: contrast agent safety	3.57 (1.95)	3.71 (1.98)	3.21 (1.80)	<.001
PBLI[d]				
PBLI: evidence-based and informed practice	3.25 (1.84)	3.34 (1.86)	3.02 (1.78)	<.001
PROF[e]				
PROF: self-awareness and help-seeking	3.49 (1.90)	3.61 (1.92)	3.20 (1.83)	<.001
ICS[f]				
ICS: patient- and family-centered communication	3.72 (2.10)	3.87 (2.11)	3.33 (2.03)	<.001
Average (all subcompetencies)	3.37 (1.47)	3.46 (1.49)	3.13 (1.39)	<.001

[a]PC: patient care.

[b]MK: medical knowledge.

[c]SBP: system-based practice.

[d]PBLI: practice-based learning and improvement.

[e]PROF: professionalism.

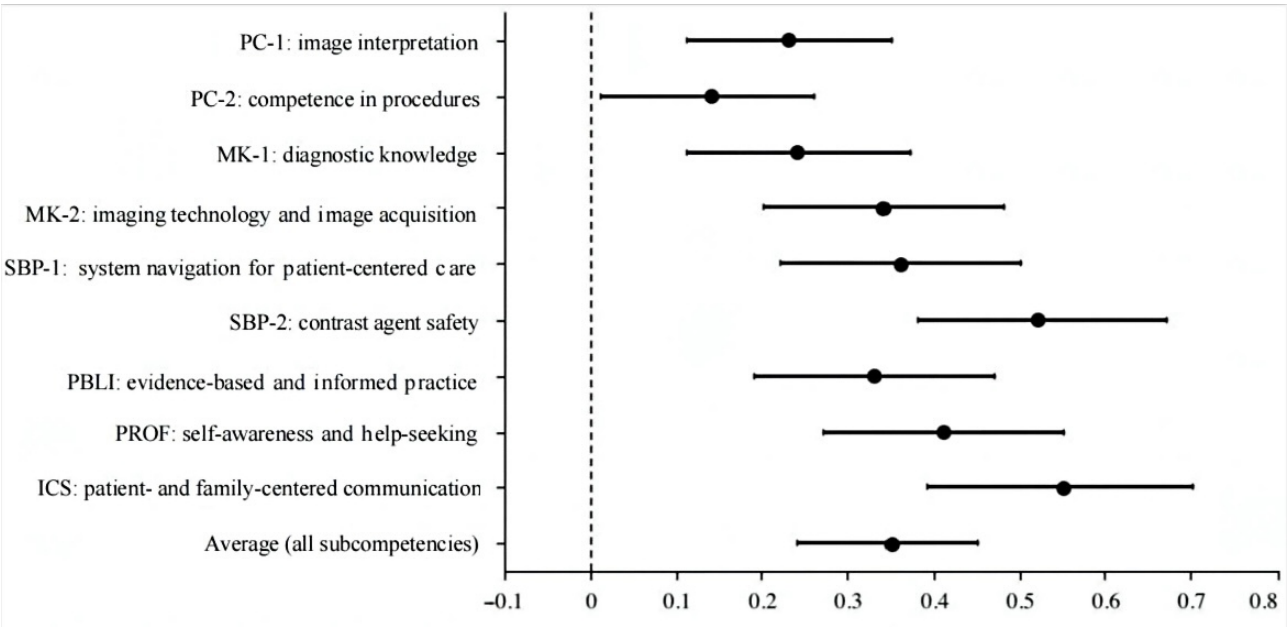
[f]ICS: interpersonal communication skill.

Association Between Distance Learning and Competencies Among Radiology Residents

As is shown in Figure 3 (see also Multimedia Appendix 4), MLR analyses showed that radiology residents who were trained through distance learning were more likely to report high competencies (β=0.35, 90% CI 0.24 - 0.45) after adjusted by participants' characteristics, including age, sex, educational level, training years, working hours per week, annual after-tax income in 2020, and types of residents. This was particularly evident in the competencies of "interpersonal and communication skills" (β=0.55, 90% CI 0.39 - 0.70) and "contrast agent safety" (β=0.52, 90% CI 0.38 - 0.67). Whereas, the competency of "patient care and technical skills" benefited the least from distance learning (β=0.14, 90% CI 0.01 - 0.26). The effect of each explanatory variable on the overall average SEM score is shown in Multimedia Appendix 5. Factors associated with higher competencies included older age (>27 years: β=0.12, 90% CI 0.01 - 0.22), being male (β=0.28, 90% CI 0.18 - 0.37), and having a longer SRT training year (third year: β=0.51, 90% CI 0.41 - 0.61).

XSL•FO
RenderX

Figure 3. Associations between distance learning and competencies among residents. ICS: interpersonal communication skill; MK: medical knowledge; PBLI: practice-based learning and improvement; PC: patient care; PROF: professionalism; SBP: system-based practice.



The Moderating Effect of Mental Health Status Between Distance Learning and Competencies

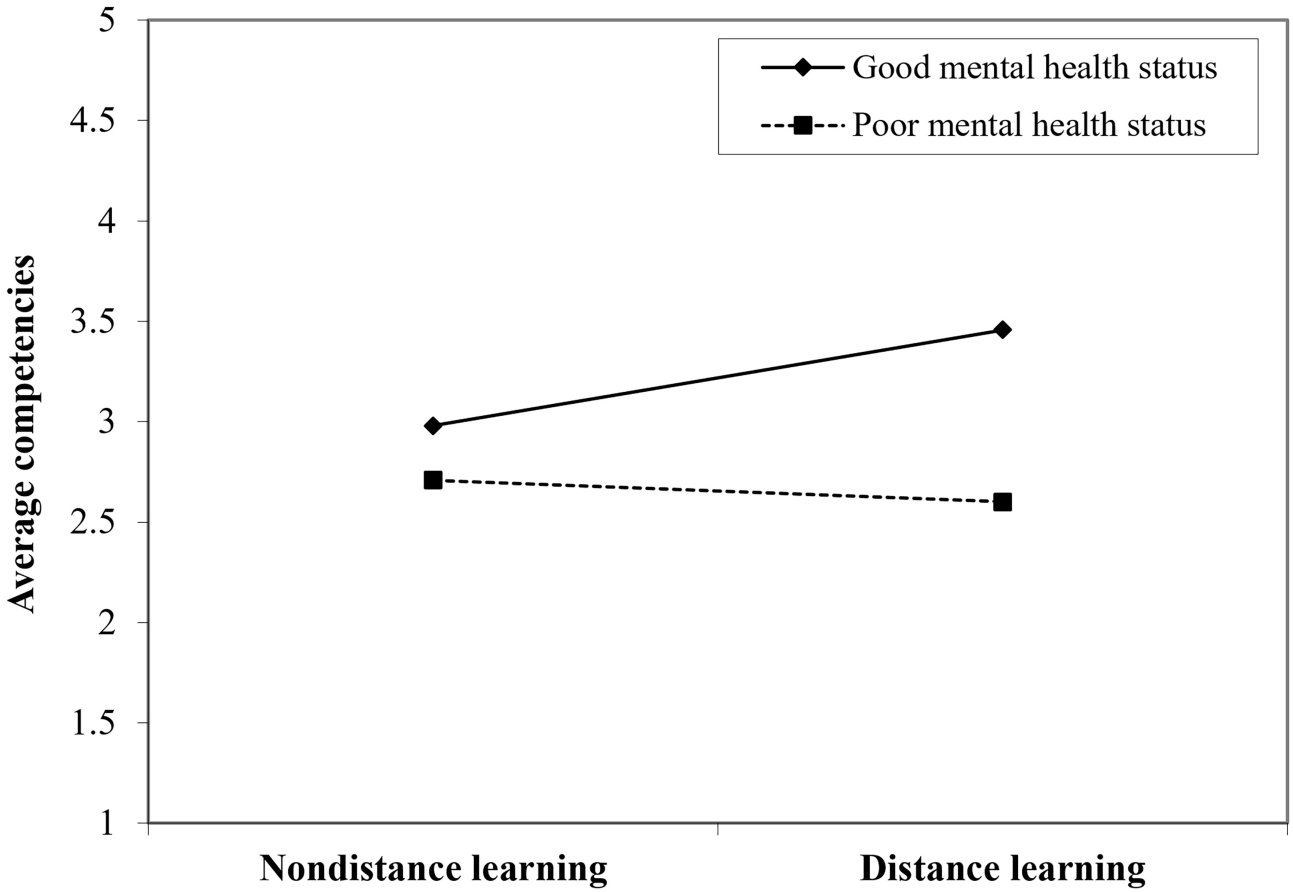
The association between mental health status and competencies is shown in Multimedia Appendix 6. We controlled the covariates mentioned earlier to investigate the moderating effect of mental health on the association between distance learning and competencies. As is shown in Table 3, there was a significant association between distance learning and radiology residents' competencies (β=0.63, 90% CI 0.34 - 0.91; *P*<.001) and was moderated by mental health status (β=–0.15, 90% CI −0.27 to −0.02; *P*=.06). The relationship between distance learning and competencies at low and high (mean−SD and mean + SD, respectively) mental health scores is shown in Figure 4. Poor mental health caused by the pandemic may offset the positive effect of distance learning on residents' competencies. The highest level of competencies was found in individuals who reported less mental distress and adopted distance learning. Furthermore, the moderating effect of poor mental health on 4 subcompetencies (ie, MK-1, MK-2, SBP-2, and ICS) was similar to it on total competencies (*P*<.10; Multimedia Appendix 7).

XSL•FO
RenderX

Table . The moderating effects of mental health status on the association between competencies and distance learning.

Variables	Multiple linear regression models		
	β (SE)	90% CI	*P* value
Distance learning	0.63 (0.17)	0.34 to 0.91	<.001
Mental health status	–0.07 (0.06)	–0.17 to 0.04	.28
Interaction (distance learning*mental health status)	–0.15 (0.08)	–0.27 to –0.02	.06
Age (years) (reference ≤27)			
>27	0.12 (0.06)	0.01 to 0.22	.07
Sex (reference=male)			
Female	–0.27 (0.06)	–0.37 to –0.17	<.001
Education (reference=bachelor's degree)			
Master's or doctoral degree	0.19 (0.11)	0.01 to 0.38	.10
Training year (reference=second year)			
Third year	0.51 (0.06)	0.41 to 0.61	<.001
Working hours per week (reference ≤40 hours per week)			
40 - 48	0.01 (0.07)	–0.12 to 0.13	.93
>48	0.01 (0.07)	–0.11 to 0.13	.86
Income	0.01 (0.01)	–0.01 to 0.03	.26
Type of residents (reference=nonprofessional master)			
Professional master	–0.03 (0.08)	–0.17 to 0.10	.70

Figure 4. The moderating effect of mental health on the relationship between distance learning and average competencies. Two lines are the visual representation of different moderation effects of distance learning on competencies when the mental health status is at its +1SD and –1SD.



XSL•FO
RenderX

Discussion

Principal Findings

Based on a national survey of radiology residents in China, our study found that radiology residents who received distance learning were more likely to report high proficiency in key competencies after 1 year. This was particularly true for learning knowledge and communication skills but was less evident in obtaining technical skills. In addition, we found a significant moderating effect of mental health status on the association between distance learning and competencies during COVID-19.

Web-based training programs are proposed to mitigate the loss of learning from clinical rotations during the pandemic [39,40]. Consistent with the results of previous findings that distance learning has the potential to improve learners' academic performance, skill development, and engagement [39,41], our study observed a positive impact of distance learning on radiology residents' self-assessed competencies. However, it should be noted that the impact of distance learning on professional competency varies among radiology residents. Our study enriches the understanding of distance learning by using milestones of 9 subcompetencies for 6 core competencies. We also assessed the long-term outcomes of distance learning among radiology residents in the initial stage of the COVID-19 pandemic (between January and May 2020). These findings can be used to help inform evidence-based policies to improve residency training in the future.

Knowledge gain is an important indicator of what trainees have learned during SRT training [42]. The radiology residents in our study reported a substantial gain of knowledge and communication skills during SRT, particularly in ICS competency and SBP competency that include clinical knowledge and medical humanities, with a regression coefficient >0.5. Whereas, the effect of distance learning on professional attitudes (PROF) and professional growth in clinical practice (PBLI) was moderate, with a regression coefficient between 0.3 and 0.4, which is in line with previous studies [43-45]. Distance learning was found to be supportive in the continuity of teaching and learning during COVID-19 [39], which might explain why residents who received distance learning reported a higher score in the competence of medical knowledge than nondistance learners. In addition, radiology residents are able to participate in web-based conferences and conduct digital medical consultations to enhance their competency of communication [43], though it could be hard to evaluate their body language in distance learning [46]. Our findings validated the positive role of distance learning in fulfilling the objectives of training by creating an enabling environment, especially in the domain of knowledge and communication. In other words, distance learning can help keep the consistency of residency programs during pandemics, facilitating the resilience and recovery of health systems.

Our findings showed that competence in procedures (PC-2) benefited the least from distance learning, followed by competencies in technical skills (including PC competencies that reflect radiology residents' professionalism and MK competencies that reflect mastery of professional knowledge and imaging technology). The result is consistent with previous findings that most medical students feel unable to acquire practical clinical skills through web-based teaching during COVID-19 [47]. This could be explained by the reduction in daily cases due to COVID-19, which was one of the deficiencies in distance learning [20]. To address this challenge, new technologies such as touchscreen Anatomage Table and Touch Surgery application have been used to strengthen trainings of plastic surgery [48]. These new approaches can be integrated into web-based residency training to compensate for the shortage of training in technical skills with distance learning. Other training programs, such as competence in procedures, could also be carried out among radiology residents in the post–COVID-19 era.

Health professional education involves maintaining a sense of purpose and mental well-being (eg, balance work life, stress, depression, burnout, and more) among residents [49-51]. Our study extended the prior results by identifying that poor mental health moderates the relationship between distance learning and competency among radiology residents. During the COVID-19 pandemic, physicians experienced more mental health disorders [52] with negative impacts (eg, higher turnover intention and lower work performance) [53,54]. However, good mental health may help trainees to receive distance learning consistently [55]. When trainees are in a suboptimal mental status, they tend to have negative attitudes toward learning and are discouraged from receiving distance learning continuously [55]. Health workforce is the backbone of health systems in response of pandemics [56]. In addition to the structural changes required for the improvement of health systems' resilience, it could also be necessary to provide long-term psychological support for the health workforce to help them overcome psychological distress [57]. In summary, several recommendations can be drawn from this study: (1) distance learning can be used for transferring knowledge to residents particularly when challenges exist in the traditional offline approaches; (2) clinical skills have a crucial role in offline training, which should be noted and well used by training institutions; (3) instructors could use a scientific and caring approach with a special focus on learners' psychological well-being when preparing material and organizing courses; and (4) physicians who have engaged in distance learning could enhance their learning experience through an enhancement of their offline learning environment. This promotes a better psychological state for learning to become more effective.

Health system resilience is critical in training the required competencies of the health workforce [49]. In addition to competency-based education, the COVID-19 pandemic highlighted the application of digital interprofessional education [58]. In the future, distance learning may help physicians gain knowledge and skills in public health, interprofessional communication, and teamwork. In this regard, remote residency training should be developed as a holistic educational concept rather than a mere substitution for traditional in-person learning. Distance learning enables physicians to have a flexible schedule, a feasible access to classes, and an opportunity to keep a good balance between work and life [14]. What is more, the facilitation of distance learning contributes to the sharing of

XSL•FO
RenderX

high-quality educational resources in the post–COVID-19 era. In particular, for health professionals in resource-constrained places (eg, rural areas in western China), the provision of high-quality distance learning could help them overcome geographical constraints and thus reduce inequalities in access to education [49,59]. Nevertheless, it is not possible to replace in-person teaching with distance learning completely. An integrated model of the 2 is encouraged to maximize the advantages of different teaching modalities.

Limitations

This study has several limitations. First, we used SEM for the self-evaluation of radiology residents' competencies; results may be influenced by the Dunning-Kruger effect, where participants with lower abilities tend to overestimate their competencies and the potential self-reporting bias [60]. Second, the lower response rate may be subject to selection bias, as our survey is voluntary in nature. However, based on sample size calculations [61], we obtained 1573 valid samples (precision=5%; baseline proportion=0.50), which accurately represent the characteristics of the residents. Importantly, this represents the largest nationally representative sample of radiology residents in China to date, which may help to minimize the potential bias. Third, although mental health status was asked by a single question based on the self-report psychological distress on a Likert scale, this variable is correlated with long-term mental health status (depression and burnout). Other potential moderating factors that may influence learning status, such as the design of the web-based course, courseware, and teaching styles, could be explored in future research. Fourth, the generalization of the results is another limitation of our study [49].

Future Directions

As distance learning is anticipated to be applied across various fields, additional research is warranted to substantiate our findings. Longitudinal studies are recommended for future research to fully assess the long-term effects of distance learning on competence and mental health.

Conclusions

Distance learning helps mitigate the negative impact of the COVID-19 lockdown on the education of health professionals. Meanwhile, attention should be paid to the disadvantages of distance learning and the mental health status of learners, as they may negatively influence the effectiveness and sustainability of distance learning. Our study provides insights into the role of distance learning in residency training during the pandemic.

Acknowledgments

The authors appreciate the time and effort of all the project staff and participants.

Data Availability

The datasets used and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

J Zhu, PW, Z Wu, YC and ZY conceived the study and its design and were responsible for all aspects of the study. J Zhang, MJ, J Zheng, Z Wang, and ZY collected the data set. PW and Z Wu conducted the statistical analyses. PW, Z Wu, YC, and J Zhu drafted the paper. Z Wu, PW, YC, YH, and J Zhu were involved in manuscript preparation and revisions. ZY, YC, and J Zhu are corresponding authors. All authors have approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The Spearman *r* between short- and long-term mental health status.
[PDF File, 57 KB - mededu_v11i1e54228_app1.pdf]

Multimedia Appendix 2
Self-evaluation milestone scores for radiology residents between distance learning and nondistance learning.
[PDF File, 108 KB - mededu_v11i1e54228_app2.pdf]

Multimedia Appendix 3
Distance learning efforts for radiology residents who had not participated in COVID-19–related activities.
[PDF File, 96 KB - mededu_v11i1e54228_app3.pdf]

Multimedia Appendix 4
Associations between distance learning and competencies among residents.

XSL•FO
RenderX

[PDF File, 126 KB - mededu_v11i1e54228_app4.pdf]

Multimedia Appendix 5
Association between distance learning and the overall average self-evaluation milestone score among residents.
[PDF File, 96 KB - mededu_v11i1e54228_app5.pdf]

Multimedia Appendix 6
The Spearman *r* between mental health status and competencies.
[PDF File, 123 KB - mededu_v11i1e54228_app6.pdf]

Multimedia Appendix 7
The moderating effects of mental health on the association between subcompetencies and distance learning.
[PDF File, 119 KB - mededu_v11i1e54228_app7.pdf]

References

1. Zhang P, Gao J. Evaluation of China's public health system response to COVID-19. J Glob Health 2021 Jan 16;11:05004. [doi: 10.7189/jogh.11.05004] [Medline: 33643637]
2. Liu Q, Luo D, Haase JE, et al. The experiences of health-care providers during the COVID-19 crisis in China: a qualitative study. Lancet Glob Health 2020 Jun;8(6):e790-e798. [doi: 10.1016/S2214-109X(20)30204-7] [Medline: 32573443]
3. Xu TL, Ao MY, Zhou X, et al. China's practice to prevent and control COVID-19 in the context of large population movement. Infect Dis Poverty 2020 Aug 19;9(1):115. [doi: 10.1186/s40249-020-00716-0] [Medline: 32814591]
4. Li Z, Chen Q, Feng L, et al. Active case finding with case management: the key to tackling the COVID-19 pandemic. Lancet 2020 Jul 4;396(10243):63-70. [doi: 10.1016/S0140-6736(20)31278-2] [Medline: 32505220]
5. Dong C, Tian Y, Xu W, et al. Introduction on collective quarantine of close contacts of patients with COVID-19 for medical observation in China: from the perspective of frontline staff. Biosci Trends 2020 Jul 17;14(3):215-221. [doi: 10.5582/bst.2020.03094] [Medline: 32389941]
6. Bourgeault IL, Maier CB, Dieleman M, et al. The COVID-19 pandemic presents an opportunity to develop more sustainable health workforces. Hum Resour Health 2020 Oct 31;18(1):83. [doi: 10.1186/s12960-020-00529-0] [Medline: 33129313]
7. Kuhlmann E, Dussault G, Wismar M. Health labour markets and the "human face" of the health workforce: resilience beyond the COVID-19 pandemic. Eur J Public Health 2020 Sep 1;30(Suppl 4):iv1-iv2. [doi: 10.1093/eurpub/ckaa122] [Medline: 32949241]
8. Kruk ME, Myers M, Varpilah ST, Dahn BT. What is a resilient health system? Lessons from Ebola. Lancet 2015 May;385(9980):1910-1912. [doi: 10.1016/S0140-6736(15)60755-3]
9. Zhang J, Han X, Yang Z, et al. Radiology residency training in China: results from the first retrospective nationwide survey. Insights Imaging 2021 Feb 17;12(1):25. [doi: 10.1186/s13244-021-00970-2] [Medline: 33595737]
10. National Health and Family Planning Commission, State Commission Office for Public Sector Reform (SCOPSR), National Development and Reform Commission, Ministry of Education, Ministry of Finance, Ministry of Human Resources and Social Security, State Administration of Traditional Chinese Medicine. Guiding opinions on establishing the standardised residency training system. National Health Commission of the People's Republic of China. 2013. URL: http://www.nhc.gov.cn/qjjys/wslgf/201401/032c8cdf2eb64a369cca4f9b76e8b059.shtml [accessed 2022-06-12]
11. Diagnostic radiology milestones. The Accreditation Council for Graduate Medical Education. 2019. URL: https://www.acgme.org/globalassets/pdfs/milestones/diagnosticradiologymilestones.pdf [accessed 2022-06-20]
12. Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system—rationale and benefits. N Engl J Med 2012 Mar 15;366(11):1051-1056. [doi: 10.1056/NEJMsr1200117] [Medline: 22356262]
13. Leddy R, Lewis M, Ackerman S, et al. Practical implications for an effective radiology residency quality improvement program for milestone assessment. Acad Radiol 2017 Jan;24(1):95-104. [doi: 10.1016/j.acra.2016.08.018] [Medline: 27769821]
14. Chen D, Ayoob A, Desser TS, Khurana A. Review of learning tools for effective radiology education during the COVID-19 era. Acad Radiol 2022 Jan;29(1):129-136. [doi: 10.1016/j.acra.2021.10.006] [Medline: 34799258]
15. Dedeilia A, Sotiropoulos MG, Hanrahan JG, Janga D, Dedeilias P, Sideris M. Medical and surgical education challenges and innovations in the COVID-19 era: a systematic review. In Vivo 2020 Jun;34(3 Suppl):1603-1611. [doi: 10.21873/invivo.11950] [Medline: 32503818]
16. Lo HY, Lin SC, Chaou CH, Chang YC, Ng CJ, Chen SY. What is the impact of the COVID-19 pandemic on emergency medicine residency training: an observational study. BMC Med Educ 2020 Oct 7;20(1):348. [doi: 10.1186/s12909-020-02267-2] [Medline: 33028295]
17. Brady A, Brink J, Slavotinek J. Radiology and value-based health care. JAMA 2020 Oct 6;324(13):1286-1287. [doi: 10.1001/jama.2020.14930] [Medline: 32915190]

XSL•FO
RenderX

18. Cavalieri S, Spinetta M, Zagaria D, et al. The impact of COVID-19 pandemic on radiology residents in Northern Italy. Eur Radiol 2021 Sep;31(9):7077-7087. [doi: 10.1007/s00330-021-07740-0] [Medline: 33755754]
19. Alvin MD, George E, Deng F, Warhadpande S, Lee SI. The impact of COVID-19 on radiology trainees. Radiology 2020 Aug;296(2):246-248. [doi: 10.1148/radiol.2020201222] [Medline: 32216719]
20. Patil NS, Gunter D, Larocque N. The impact of the COVID-19 pandemic on radiology resident education: where do we go from here? Acad Radiol 2022 Apr;29(4):576-583. [doi: 10.1016/j.acra.2021.11.015] [Medline: 35033451]
21. Larocque N, Shenoy-Bhangle A, Brook A, Eisenberg R, Chang YM, Mehta P. Resident experiences with virtual radiology learning during the COVID-19 pandemic. Acad Radiol 2021 May;28(5):704-710. [doi: 10.1016/j.acra.2021.02.006] [Medline: 33640229]
22. Lanier MH, Wheeler CA, Ballard DH. A new normal in radiology resident education: lessons learned from the COVID-19 pandemic. Radiographics 2021;41(3):E71-E72. [doi: 10.1148/rg.2021210030] [Medline: 33939548]
23. Chénard-Roy J, Guitton MJ, Thuot F. Online residency training during the COVID-19 pandemic: a national survey of otolaryngology head and neck surgery program directors. J Otolaryngol Head Neck Surg 2021 Nov 16;50(1):65. [doi: 10.1186/s40463-021-00546-6] [Medline: 34784978]
24. Awan OA. Virtual radiology readouts after the coronavirus disease (COVID-19) pandemic. AJR Am J Roentgenol 2021 Sep;217(3):765-766. [doi: 10.2214/AJR.21.25607] [Medline: 33594906]
25. Chen Q, Liang M, Li Y, et al. Mental health care for medical staff in China during the COVID-19 outbreak. Lancet Psychiatry 2020 Apr;7(4):e15-e16. [doi: 10.1016/S2215-0366(20)30078-X] [Medline: 32085839]
26. Giusti L, Mammarella S, Salza A, et al. Predictors of academic performance during the covid-19 outbreak: impact of distance education on mental health, social cognition and memory abilities in an Italian university student sample. BMC Psychol 2021 Sep 15;9(1):142. [doi: 10.1186/s40359-021-00649-9] [Medline: 34526153]
27. Zhao L, Hwang WY, Shih TK. Investigation of the physical learning environment of distance learning under COVID-19 and its influence on students' health and learning satisfaction. Int J Distance Educ Technol 2021 Apr;19(2):77-98. [doi: 10.4018/IJDET.20210401.oa4]
28. Williamson DL, Carr J. Health as a resource for everyday life: advancing the conceptualization. Crit Public Health 2009 Mar;19(1):107-122. [doi: 10.1080/09581590802376234]
29. Airila A, Hakanen JJ, Schaufeli WB, Luukkonen R, Punakallio A, Lusa S. Are job and personal resources associated with work ability 10 years later? The mediating role of work engagement. Work Stress 2014 Jan 7;28(1):87-105. [doi: 10.1080/02678373.2013.872208]
30. Raja U, Azeem MU, Haq IU, Naseer S. Perceived threat of terrorism and employee outcomes: the moderating role of negative affectivity and psychological capital. J Bus Res 2020 Mar;110:316-326. [doi: 10.1016/j.jbusres.2020.01.026]
31. Stirpe L, Profili S, Sammarra A. Satisfaction with HR practices and employee performance: a moderated mediation model of engagement and health. Eur Manag J 2022 Apr;40(2):295-305. [doi: 10.1016/j.emj.2021.06.003]
32. Di Malta G, Bond J, Conroy D, Smith K, Moller N. Distance education students' mental health, connectedness and academic performance during COVID-19: a mixed-methods study. Distance Educ 2022 Jan 2;43(1):97-118. [doi: 10.1080/01587919.2022.2029352]
33. Lyle B, Borgert AJ, Kallies KJ, Jarman BT. Do attending surgeons and residents see eye to eye? An evaluation of the Accreditation Council for Graduate Medical Education milestones in general surgery residency. J Surg Educ 2016;73(6):e54-e58. [doi: 10.1016/j.jsurg.2016.07.004] [Medline: 27561627]
34. Barbato KBG, de Carvalho LS, Barreira Marangoni V, de Souza F, de Vasconcelos Vaena MM. Core competencies self-assessment and patient-practitioner orientation during the first year of a Brazilian orthopedic residency. Rev Bras Ortop (Sao Paulo) 2023 Oct;58(5):e742-e749. [doi: 10.1055/s-0043-1768621] [Medline: 37908538]
35. Watson RS, Borgert AJ, O Heron CT, et al. A multicenter prospective comparison of the Accreditation Council for Graduate Medical Education milestones: clinical competency committee vs. resident self-assessment. J Surg Educ 2017 Nov;74(6):e8-e14. [doi: 10.1016/j.jsurg.2017.06.009]
36. Kwasny L, Shebrain S, Munene G, Sawyer R. Is there a gender bias in milestones evaluations in general surgery residency training? Am J Surg 2021 Mar;221(3):505-508. [doi: 10.1016/j.amjsurg.2020.12.020] [Medline: 33358140]
37. Li Z, Chen L, Li M, Cohen J. Prenatal exposure to sand and dust storms and children's cognitive function in China: a quasi-experimental study. Lancet Planet Health 2018 May;2(5):e214-e222. [doi: 10.1016/S2542-5196(18)30068-8] [Medline: 29709285]
38. Hodkinson A, Zhou A, Johnson J, et al. Associations of physician burnout with career engagement and quality of patient care: systematic review and meta-analysis. BMJ 2022 Sep 14;378:e070442. [doi: 10.1136/bmj-2022-070442] [Medline: 36104064]
39. Ahmady S, Kallestrup P, Sadoughi MM, et al. Distance learning strategies in medical education during COVID-19: a systematic review. J Educ Health Promot 2021;10:421. [doi: 10.4103/jehp.jehp_318_21] [Medline: 35071627]
40. Chick RC, Clifton GT, Peace KM, et al. Using technology to maintain the education of residents during the COVID-19 pandemic. J Surg Educ 2020;77(4):729-732. [doi: 10.1016/j.jsurg.2020.03.018] [Medline: 32253133]
41. Abdull Mutalib AA, Md Akim A, Jaafar MH. A systematic review of health sciences students' online learning during the COVID-19 pandemic. BMC Med Educ 2022 Jul 3;22(1):524. [doi: 10.1186/s12909-022-03579-1] [Medline: 35786374]

XSL•FO
RenderX

42. Ritzmann S, Hagemann V, Kluge A. The Training Evaluation Inventory (TEI)—evaluation of training design and measurement of training outcomes for predicting training success. Vocat Learn 2014 Apr;7(1):41-73. [doi: 10.1007/s12186-013-9106-4]
43. Warnica W, Moody A, Probyn L, Bartlett E, Singh N, Pakkal M. Lessons learned from the effects of COVID-19 on the training and education workflow of radiology residents—a time for reflection: perspectives of residency program directors and residents in Canada. Can Assoc Radiol J 2021 Nov;72(4):637-644. [doi: 10.1177/0846537120963649] [Medline: 33047608]
44. Biswas SS, Biswas S, Awal SS, Goyal H. Current status of radiology education online: a comprehensive update. SN Compr Clin Med 2022;4(1):182. [doi: 10.1007/s42399-022-01269-z] [Medline: 35971436]
45. Liao F, Murphy D, Wu JC, Chen CY, Chang CC, Tsai PF. How technology-enhanced experiential e-learning can facilitate the development of person-centred communication skills online for health-care students: a qualitative study. BMC Med Educ 2022 Jan 25;22(1):60. [doi: 10.1186/s12909-022-03127-x] [Medline: 35078482]
46. Reinhart A, Malzkorn B, Döing C, Beyer I, Jünger J, Bosse HM. Undergraduate medical education amid COVID-19: a qualitative analysis of enablers and barriers to acquiring competencies in distant learning using focus groups. Med Educ Online 2021 Dec;26(1):1940765. [doi: 10.1080/10872981.2021.1940765] [Medline: 34128776]
47. Dost S, Hossain A, Shehab M, Abdelwahed A, Al-Nusair L. Perceptions of medical students towards online teaching during the COVID-19 pandemic: a national cross-sectional survey of 2721 UK medical students. BMJ Open 2020 Nov 5;10(11):e042378. [doi: 10.1136/bmjopen-2020-042378] [Medline: 33154063]
48. Zingaretti N, Contessi Negrini F, Tel A, Tresoldi MM, Bresadola V, Parodi PC. The impact of COVID-19 on plastic surgery residency training. Aesth Plast Surg 2020 Aug;44(4):1381-1385. [doi: 10.1007/s00266-020-01789-w]
49. Frenk J, Chen LC, Chandran L, et al. Challenges and opportunities for educating health professionals after the COVID-19 pandemic. Lancet 2022 Oct 29;400(10362):1539-1556. [doi: 10.1016/S0140-6736(22)02092-X] [Medline: 36522209]
50. Rotenstein LS, Berwick DM, Cassel CK. Addressing well-being throughout the health care workforce: the next imperative. JAMA 2022 Aug 9;328(6):521-522. [doi: 10.1001/jama.2022.12437] [Medline: 35849383]
51. Aiken LH, Simonetti M, Sloane DM, et al. Hospital nurse staffing and patient outcomes in Chile: a multilevel cross-sectional study. Lancet Glob Health 2021 Aug;9(8):e1145-e1153. [doi: 10.1016/S2214-109X(21)00209-6] [Medline: 34224669]
52. Li W, Frank E, Zhao Z, et al. Mental health of young physicians in China during the novel coronavirus disease 2019 outbreak. JAMA Netw Open 2020 Jun 1;3(6):e2010705. [doi: 10.1001/jamanetworkopen.2020.10705] [Medline: 32478846]
53. Al-Mansour K. Stress and turnover intention among healthcare workers in Saudi Arabia during the time of COVID-19: can social support play a role? PLoS One 2021;16(10):e0258101. [doi: 10.1371/journal.pone.0258101] [Medline: 34618851]
54. Sadovyy M, Sánchez-Gómez M, Bresó E. COVID-19: how the stress generated by the pandemic may affect work performance through the moderating role of emotional intelligence. Pers Individ Dif 2021 Oct;180:110986. [doi: 10.1016/j.paid.2021.110986] [Medline: 34629581]
55. Gu Z, Li P, Zhang A, Xu X, Gu F. The role of mental health and sustainable learning behavior of students in education sector influences sustainable environment. Front Psychol 2022;13:822751. [doi: 10.3389/fpsyg.2022.822751] [Medline: 35211067]
56. Witter S, Wurie H, Chandiwana P, et al. How do health workers experience and cope with shocks? Learning from four fragile and conflict-affected health systems in Uganda, Sierra Leone, Zimbabwe and Cambodia. Health Policy Plan 2017 Nov 1;32(Suppl 3):iii3-iii13. [doi: 10.1093/heapol/czx112] [Medline: 29149313]
57. Dean L, Cooper J, Wurie H, et al. Psychological resilience, fragility and the health workforce: lessons on pandemic preparedness from Liberia and Sierra Leone. BMJ Glob Health 2020 Sep;5(9):e002873. [doi: 10.1136/bmjgh-2020-002873] [Medline: 32988928]
58. Samarasekera DD, Nyoni CN, Amaral E, Grant J. Challenges and opportunities in interprofessional education and practice. Lancet 2022 Oct 29;400(10362):1495-1497. [doi: 10.1016/S0140-6736(22)02086-4] [Medline: 36522199]
59. Huang Y. Research on online education in the midst of the COVID-19 pandemic. JAER 2020 May 2;5(2):77-80. [doi: 10.22606/jaer.2020.52005]
60. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. J Pers Soc Psychol 1999 Dec;77(6):1121-1134. [doi: 10.1037//0022-3514.77.6.1121] [Medline: 10626367]
61. Hu LP, Liu HG. Triple-type theory of statistics and its application in the scientific research of biomedicine. Zhonghua Yi Xue Za Zhi 2005 Jul 20;85(27):1936-1940. [Medline: 16255993]

Abbreviation

ACGME: Accreditation Council of Graduate Medical Education
CAR: Chinese Association of Radiologists
ICS: interpersonal communication skill
MK: medical knowledge
MLR: multiple linear regression
PBLI: practice-based learning and improvement

XSL•FO
RenderX

PC: patient care
PROF: professionalism
SBP: system-based practice
SEM: self-evaluation milestone
SRT: standardized residency training

Edited by B Lesselroth; submitted 05.11.23; peer-reviewed by S Biswas, X Sun; revised version received 05.10.24; accepted 15.02.25; published 08.05.25.

Please cite as:
Wang P, Wu Z, Zhang J, He Y, Jiang M, Zheng J, Wang Z, Yang Z, Chen Y, Zhu J
Distance Learning During the COVID-19 Lockdown and Self-Assessed Competency Development Among Radiology Residents in China: Cross-Sectional Survey
JMIR Med Educ 2025;11:e54228
URL: https://mededu.jmir.org/2025/1/e54228
doi:10.2196/54228

© Peicheng Wang, Ziye Wu, Jingfeng Zhang, Yanrong He, Maoqing Jiang, Jianjun Zheng, Zhenchang Wang, Zhenghan Yang, Yanhua Chen, Jiming Zhu. Originally published in JMIR Medical Education (https://mededu.jmir.org), 8.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Comparison of an Emergency Medicine Asynchronous Learning Platform Usage Before and During the COVID-19 Pandemic: Retrospective Analysis Study

Blake Briggs[1*], MD; Madhuri Mulekar[2*], PhD; Hannah Morales[1*], DO; Iltifat Husain[3*], MD

[1]Division of Emergency Medicine, Department of Surgery, University of Tennessee Graduate School of Medicine, 1924 Alcoa Highway, Knoxville, TN, United States

[2]Department of Mathematics, University of South Alabama, Mobile, AL, United States

[3]Department of Emergency Medicine, Wake Forest School of Medicine, Winston-Salem, NC, United States

[*]all authors contributed equally

Corresponding Author:
Blake Briggs, MD
Division of Emergency Medicine, Department of Surgery, University of Tennessee Graduate School of Medicine, 1924 Alcoa Highway, Knoxville, TN, United States

Abstract

Background: The COVID-19 pandemic challenged medical educators due to social distancing. Podcasts and asynchronous learning platforms help distill medical education in a socially distanced environment. Medical educators interested in providing asynchronous teaching should know how these methods performed during the pandemic.

Objective: The purpose of this study was to assess the level of engagement for an emergency medicine (EM) board review podcast and website platform, before and during the COVID-19 pandemic. We measured engagement via website traffic, including such metrics as visits, bounce rate, unique visitors, and page views. We also evaluated podcast analytics, which included total listeners, engaged listeners, and number of plays.

Methods: Content was designed after the American Board of EM Model, covering only 1 review question per episode. Website traffic and podcast analytics were studied monthly from 2 time periods of 20 months each, before the pandemic (July 11, 2018, to February 31, 2020) and during the pandemic (May 1, 2020, to December 31, 2021). March and April 2020 data were omitted from the analysis due to variations in closure at various domestic and international locations. Results underwent statistical analysis in March 2022.

Results: A total of 132 podcast episodes and 93 handouts were released from July 11, 2018, to December 31, 2021. The mean number of listeners per podcast increased significantly from 2.11 (SD 1.19) to 3.77 (SD 0.76; *t* test, $P<.001$), the mean number engaged per podcast increased from 1.72 (SD 1.00) to 3.09 (SD 0.62; *t* test, $P<.001$), and the mean number of plays per podcast increased from 42.54 (SD 40.66) to 69.23 (SD 17.54; *t* test, $P=.012$). Similarly, the mean number of visits per posting increased from 5.85 (SD 3.28) to 15.39 (SD 3.06; *t* test, $P<.001$), the mean number of unique visitors per posting increased from 3.74 (SD 1.83) to 10.41 (SD 2.33; *t* test, $P<.001$), and the mean number of page views per posting increased from 17.13 (SD 10.63) to 33.32 (SD 7.01; *t* test, $P<.001$). Note that, all measures showed a decrease from November 2021 to December 2021.

Conclusions: During the COVID-19 pandemic, there was an increased engagement for our EM board review podcast and website platform over a long-term period, specifically through website visitors and the number of podcast plays. Medical educators should be aware of the increasing usage of web-based education tools, and that asynchronous learning is favorably viewed by learners. Limitations include the inability to view Spotify (Spotify Technology S.A.) analytics during the study period, and confounding factors like increased popularity of social media inadvertently promoting the podcast.

(JMIR Med Educ 2025;11:e58100) doi:10.2196/58100

KEYWORDS

asynchronous learning; medical education; podcast; COVID-19; emergency medicine; online learning; engagement; web-based; online study; online class; videoconferencing; assessment; effectiveness; challenges; knowledge retention; performance; virtual learning; pre-pandemic; post-pandemic

XSL•FO
RenderX

Introduction

As the field of medical education evolves, web-based media and digital study tools are finding larger audiences each year [1]. The COVID-19 pandemic dramatically changed the landscape of medical education. Suddenly in March and April 2020, all learning was switched to remote platforms, greatly challenging educators and hastening the switch to web-based media [2-4].

Previous studies have demonstrated that podcasts have positive effects on knowledge retention and test performance [5,6]. Multiple studies have previously been published on the effectiveness of remote learning during the COVID-19 pandemic via remote learning and web-based modules [7,8]. Most recently, 1 study aimed to measure podcast and blog utilization during the early months of the COVID-19 pandemic [9]. This study found an increase in blog page views during the early months of the pandemic, but no statistical change in podcast usage. However, this study had a short measurement period (January to May 2020). In addition, the study made measuring educational content related to COVID-19 a secondary outcome. As asynchronous teaching continues to increase in popularity among students in the wake of the pandemic, medical educators should be curious about the popularity of such materials during a time in which in-person education was severely limited or paused altogether. The purpose of this study was to assess the level of engagement for an emergency medicine (EM) board review podcast and platform, comparing before COVID-19 to during the COVID-19 pandemic over a period of 34 months. Our secondary outcome was to measure important website variables that have previously not been mentioned in medical education literature, especially in the setting of the pandemic. We hypothesized that the pandemic would increase the number of website visitors, page views, and podcast episode plays.

Methods

Overview

This retrospective analysis was conducted from March 5, 2022, to April 30, 2022. Data were collected by the study authors from July 11, 2018, when the first podcast episode was released, to December 31, 2021. Emergency Medicine Board Bombs (EMBB) was launched by 2 academic EM physicians in July 2018. The goal of this asynchronous educational platform was to increase first-time pass rate among residents and attendings taking their in-service exam and boards, respectively. EMBB is a peer-reviewed resource and functions at no cost to the learner. EMBB has never been formally assigned to any formal, academic curriculum; its educational platform is entirely free and open access to all learners. The website has podcasts and printable study guides that function as summaries of various common pathologies encountered in the emergency department and on-the-board exams.

Platform Development

Each podcast episode was structured to quickly cover one multiple-choice question, a discussion of correct and incorrect answers, and the relevant subject matter. Audio-editing was conducted using Apple Garageband, a free service provided to those who own Apple hardware. The podcast was available for free streaming on a designated website, emboardbombs, as well as dedicated podcast platforms (Apple Podcasts, Soundcloud [SoundCloud Global Limited & Co KG], and Spotify [Spotify Technology S.A.]). Questions for each episode were modeled after the American Board of Emergency Medicine (ABEM) certification exam. The Model of the Clinical Practice of Emergency Medicine (EM Model), serves as the basis for ABEM content and was followed in drafting podcast episodes [10]. A peer review process was used to develop multiple-choice questions. Each question was written by an EM physician with an academic appointment and was shared with 2 other academic physicians for review before it was featured on the podcast.

Medical source material was derived from *Tintinalli's Emergency Medicine* as well as UpToDate and EB Medicine [11-13]. The educational platform was self-funded by the creators and developed on their own time and schedule. No financial support or aid was received. In terms of dedicated time and monetary investment, the cost of equipment and software totaled nearly US $400 annually. In terms of hourly commitment, approximately 5 - 10 hours weekly is needed to record, edit, and publish podcast episodes, as well as write and publish study guides. The podcast was not formally added to any curriculum. It was disseminated by word of mouth. No marketing or paid advertising was used.

Variable Definitions

Podcast analytics were derived from Apple Podcasts Connect which is a free service provided for all Apple Podcast hosts. It provides data on total listeners, engaged listeners, and number of plays [14]. Listeners were defined by Apple as the number of unique devices that played more than 0 seconds of an episode. Engaged listeners were defined as the number of devices that played at least 20 minutes or 40% of an episode within a single session. Of note, pausing or stopping an episode did not count as starting a new session. Number of episode plays was based on the number of unique devices where the play duration is more than 0 seconds. At the time of our data collection during the pandemic, Spotify did not publish podcast statistics, and therefore, their user data could not be obtained.

The website learning platform was hosted on Squarespace. Website traffic analytics were derived from Squarespace, which measured traffic using variables such as website visits, website bounce rate, website unique visitors, and website page views [15]. Visits were defined as the total number of browsing sessions per visitor on the website within a 30-minute period. A browser cookie from Squarespace was used to track views within a 30-minute period. The bounce rate was defined as the number of visitors who navigate away from the website after viewing 1 page. Unique visitors were defined as the total number of new IP addresses that visited the website. Page views were defined as the total number of views across all pages on the website. Page views count the number of times a page is viewed. Furthermore, 1 visit consists of 1 or more pages.

XSL•FO
RenderX

Data Collection

Website traffic and podcast analytics from July 11, 2018, to February 28, 2020, were compared with those from May 1, 2020, to December 31, 2021. May 1, 2020, was chosen as the transition date because, during March and April 2020, various schools and residency programs began switching to remote learning. As the pandemic evolved, medical schools and graduate medical education sites began suspending in-person rotations. The Accreditation Council for Graduate Medical Education announced in mid-March that all in-person educational activities, meetings, and site visits were to migrate to virtual occurrences only [16]. By the end of April 2020, all nonessential, in-person educational activities had ceased [17].

Statistical Analysis

All collected data were organized in a Microsoft Excel spreadsheet and analyzed using statistical software JMP Pro 16.0.0 (SAS Institute Inc) in March 2022. All numerical data were summarized using mean and SD. Variations in monthly data from before COVID-19 and during COVID-19 periods were compared using the Levene test, whereas the means per month were compared using a 2-sample *t* test after accounting for differences in variations if any [18,19]. In addition, a nonparametric Mann-Whitney *U* test was also used to compare analytics from 2 time periods. Time series plots were used to study trends in monthly data. A significance level of 0.05 was used to determine the significance of outcomes.

Ethical Considerations

The Institutional Review Board was approached for ethics approval but reported that the study did not meet the criteria for human candidates research, and therefore, no approval was required.

Results

During the study period from July 11, 2018, to December 31, 2021, a total of 132 podcast episodes and 93 study guides were created. The first podcast episode was released on July 11, 2018.

From July 11, 2018, to February 28, 2020, 68 episodes were released, along with 30 study guides. From May 1, 2020, to December 31, 2021, 59 podcasts were released, and 53 handouts were published. Note that 5 episodes and 10 handouts were released during March-April 2020, which were also available to learners during the COVID-19 pandemic. This resulted in a total of 225 postings (132 podcasts and 93 handouts) being available to learners during the COVID-19 pandemic (Table 1).

Table . Number of podcasts, handouts, and total postings before, in-between, and during COVID-19 periods.

Period	Podcasts, n	Handouts, n	Postings, n
Before COVID-19	68	30	98
In-between period	5	10	15
During COVID-19	59	53	112
Total	132	93	225

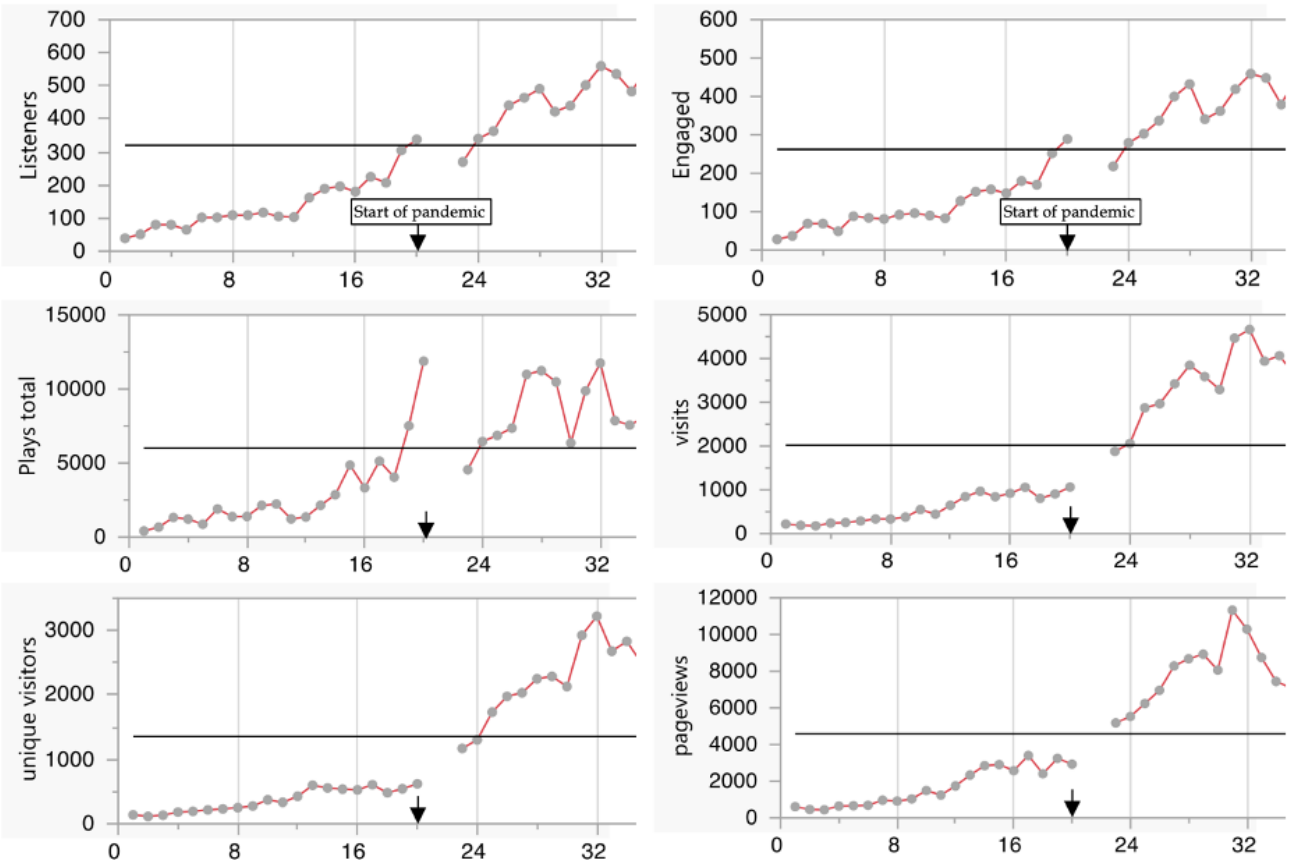
The time series presented in Figure 1 show month-to-month changes in podcast and website visit analytics before the COVID-19 and during COVID-19 periods and differences in changing patterns. Although higher outcomes were observed during the COVID-19 period in all 6 podcast and website visit measurements compared with before the COVID-19 period, not all changes showed linear patterns of increase. In fact, the number of unique visitors, visits, and page reviews showed decreasing trend after reaching a peak around the middle of the COVID-19 period. However, at the end of the 20-month period, they still remained higher than before the COVID-19 level. During the before the COVID-19 period, number of listeners per month steadily increased from 39 to 338. During the COVID-19 period, it continued to increase, reaching a maximum number of listeners at 672. A similar trend was observed for number engaged per month, increasing from 28 to 289 during the before the COVID-19 period and reaching a maximum of 555 during the COVID-19 period. Although a similar trend was observed for the total number of plays with an increase from 412 to 11,879 during the before the COVID-19 period, a sharp drop was observed during the period of uncertainty (March-April 2020). Again, during the COVID-19 period, total number of plays increased from 4547 to 14,296. Number of visits during the before the COVID-19 period increased from 218 to 1064; there was further increase in the COVID-19 period, reaching 4664 in January 2021. The number of visits started declining thereafter, reaching a low of 1879. The number of unique visitors and page views showed patterns similar to that of the number of visits. The number of unique visitors increased steadily during the before the COVID-19 period from 138 to 620. It increased to 3222 in January 2021 but started declining to a low of 2293. The number of page views also increased steadily during the before the COVID-19 period from 610 to 3405; in the COVID-19 period, it increased to 11,326 in November 2020, only to steadily decrease to a low of 5389 in December 2021. Note that all measures showed a decrease from November 2021 to December 2021.

Comparison of podcast and website visit analytics are presented in Table 2. It shows that regardless of differences in the number of podcasts and handouts available during the 2 time periods, variation in analytics from month to month did not differ significantly during the 2 time periods under study except for bounce rate and number of visitors. Significantly higher variation as measured by SD was observed in bounce rate (0.07 vs 0.05; Levene test, *P*=.036) and number of unique visitors (523.45 vs 179.62; Levene test, *P*=.0049) during COVID-19 pandemic compared with the before the COVID-19 period. Percent increase in mean analytics from before the COVID-19 period to during the COVID-19 period ranged from 24% (bounce rate, 0.55 to 0.30 per 100 postings, n=20) to 539%

XSL•FO
RenderX

(unique visitors, 3.74 to 10.41 per posting, n=20) with the mean number of unique visitors showing the highest percent increase and the bounce rate the lowest. The number of visits increased by 504% (5.85 to 15.39 per posting, n=20) whereas the number of listeners, engaged, and total plays each increased by more than 200% (listeners: 2.11 to 3.77 per podcast, n=20; engaged: 1.72 to 3.09 per podcast, n=20; total plays: 42.54 to 69.23 per podcast, n=20). Percent increases in the average monthly analytics indicate considerable increase in visits and usage of podcasts from before COVID-19 to during the COVID-19 period.

Figure 1. Monthly change in podcast and website visit analytics before COVID-19 and during COVID-19 periods. The arrowhead marks the start of the pandemic.



XSL•FO
RenderX

Table . Comparison of podcast and website visit analytics before the COVID-19 and during the COVID-19 periods.

Aspect and period		n	Mean (SD)	Range	*P* value (Levene test)	*P* value (*t* test)	Increase in mean, %
Listeners					.54	<.001	247.31
	Before COVID-19	20	143.20 (80.93)	39-338			
	During COVID-19	20	497.35 (99.84)	270-672			
Engaged listeners					.48	<.001	247.84
	Before COVID-19	20	117.15 (68.17)	28-289			
	During COVID-19	20	407.50 (82.19)	218-555			
Number of total episode plays					.95	<.001	215.88
	Before COVID-19	20	2892.85 (2764.59)	412-11,879			
	During COVID-19	20	9137.80 (2315.19)	4547-14,296			
Website visits					.06	<.001	504.3
	Before COVID-19	20	573.20 (321.17)	178-1064			
	During COVID-19	20	3463.85 (689.29)	1879-4664			
Website bounce rate					.03	<.001	24.07
	Before COVID-19	20	0.54 (0.05)	46%-62%			
	During COVID-19	20	0.67 (0.07)	52%-75%			
Website unique visitors					.004	<.001	538.99
	Before COVID-19	20	366.55 (179.62)	114-620			
	During COVID-19	20	2342.20 (523.45)	1170-3222			
Website page views					.27	<.001	346.6
	Before COVID-19	20	1678.60 (1041.70)	443-3405			
	During COVID-19	20	7496.65 (1577.68)	5183-11,326			

Although periods of similar length (ie, 20 months each) were used for comparison, the number of postings available during these 2 periods differed considerably because as new postings were made available, the earlier postings were still available for review for visitors. To account for the differences in the number of postings, analytics were adjusted by computing outcome per posting available. For example, number of listeners per podcast was computed as follows:

- Before COVID-19: # of listeners/podcast = # listeners/68
- During COVID-19: # listeners/podcast = # listeners/132

Note that this accounts for all podcasts that were available to listeners. Before COVID-19 accounts for all podcasts put out during that time and during COVID-19 used all podcasts available, that is, those that were put out before COVID-19, in-between, and during COVID-19 periods. Number of engaged and total plays were adjusted similarly by number of podcasts. Number of visits, unique visitors, and page views were adjusted similarly using all postings (ie, podcasts plus handouts). Bounce rate was adjusted similarly using per 100 postings because rate of per posting resulted in very small numbers and this change from per posting to per 100 postings does not affect the outcome of statistical tests.

Resulting comparisons of outcomes are listed in Table 3, which shows a significant increase in mean rates for all analytics except mean bounce rate per 100 postings from before COVID-19 to during COVID-19. Bounce rate per 100 postings showed a significant decrease from before COVID-19 to during

XSL•FO
RenderX

COVID-19 (0.55 to 0.30 per 100 podcasts; *t* test, *P*<.001). Mean number of listeners per podcast increased significantly from 2.11 (SD 1.19) to 3.77 (SD 0.76; *t* test, *P*<.001), mean number engaged per podcast increased from 1.72 (SD 1.00) to 3.09 (SD 0.62; *t* test, *P*<.001), and mean number of plays per podcast increased from 42.54 (SD 40.66) to 69.23 (SD 17.54; *t* test, *P*=.0122). Similarly, mean number of visits per posting increased from 5.85 (SD 3.28) to 15.39 (SD 3.06; *t* test, *P*<.001), mean number of unique visitors per posting increased from 3.74 (SD 1.83) to 10.41 (SD 2.33; *t* test, *P*<.001); and mean number of page views per posting increased from 17.13 (SD 10.63) to 33.32 (SD 7.01; *t* test, *P*<.001). Even nonparametric comparisons using Mann-Whitney *U* test gave the same results.

Table . Comparison of podcast and website visit analytics rates per posting available to viewers before COVID-19 and during COVID-19 periods.

Aspect and period	n	Mean (SD)	Range	Median (IQR)	*P* value (*t* test)	*P* value (Mann Whitney *U* test)
Listeners per podcast					<.001	<.001
Before COVID-19	20	2.11 (1.19)	0.57-4.97	1.60 (1.26-2.86)		
During COVID-19	20	3.77 (0.76)	2.05-5.09	3.83 (3.34-4.27)		
Engaged per podcast					<.001	<.001
Before COVID-19	20	1.72 (1.00)	0.41-4.25	1.34 (1.06-2.30)		
During COVID-19	20	3.09 (0.62)	1.65-4.20	3.22 (2.62-3.47)		
Number of total episode plays per podcast					.0122	<.001
Before COVID-19	20	42.54 (40.66)	6.06-174.69	29.71 (18.36-56.81)		
During COVID-19	20	69.23 (17.54)	34.45-108.30	69.95 (56.13-81.99)		
Website visits per posting					<.001	<.001
Before COVID-19	20	5.85 (3.28)	1.82-10.86	5.07 (2.68-9.08)		
During COVID-19	20	15.39 (3.06)	8.35-20.73	15.64 (13.80-17.41)		
Website bounce rate per 100 postings					<.001	<.001
Before COVID-19	20	0.55 (0.05)	0.47-0.63	0.54 (0.51-0.59)		
During COVID-19	20	0.30 (0.03)	0.23-0.33	0.30 (0.27-0.32)		
Website unique visitors per posting					<.001	<.001
Before COVID-19	20	3.74 (1.83)	1.16-6.33	3.60 (2.03-5.53)		
During COVID-19	20	10.41 (2.33)	5.20-14.32	10.65 (9.14-11.84)		
Website page views per posting					<.001	<.001
Before COVID-19	20	17.13 (10.63)	4.520-34.745	13.98 (6.85-28.35)		
During COVID-19	20	33.32 (7.01)	23.036-50.338	32.49 (28.39-38.13)		

XSL•FO
RenderX

Discussion

Principal Findings

The results demonstrate that our online EM board review podcast and platform experienced significantly increased levels of engagement during the COVID-19 pandemic. Our learning platform included multiple media, such as PDF study guides, video and picture-based modules, and online question banks. The aim was for the podcast and handouts to be integrated into an asynchronous study plan, as the platform provided easy accessibility and use.

Implication of Findings

The COVID-19 pandemic disrupted medical education, forcing learners in both medical school and residency to navigate vast amounts of information, largely in isolation. This shift from interactive, in-person learning raised concerns about students overextending themselves, leading to only a surface-level understanding of the material. One study comparing first and second-year medical student education during the pandemic highlighted the importance of face-to-face learning, finding that the first-year medical students in isolation performed worse than the previous year's first-year medical students [20]. Another retrospective study performed at the University of Hawaii Burn School of Medicine demonstrated that fourth-year medical students who were enrolled during the pandemic displayed improved note-taking with a 9-point increase in exam scores, yet worse physical examinations in their standardized patient encounters with a 12-point average decrease in scores [21].

In response, many innovative educational tools have emerged to attempt to provide asynchronous learning. Online resources like the one in this study are unique. Diverse topics are integrated into a single, cost-effective, and efficient platform, with podcast episodes <20 minutes, as well as downloadable PDF handouts. This model is beneficial for both visual and auditory learners.

While other learning platforms were not analyzed during this study period, valuable information was collected from this study's podcast. EMBB offers a humanistic aspect to learning with the dual physician hosts, pertinent banter, and narrative medicine aspect, of which may anthropomorphize the learning despite pandemic isolation.

Comparison With the Literature

Podcasts have been welcomed by those looking for a nontraditional method of learning in recent years, most notably those practicing in EM, where it is the most represented specialty that regularly hosts podcasts [22-24]. A survey in 2014 showed EM residents devote more time to podcasts than journals, citing podcasts as "the most beneficial" for education [22]. In another large survey, 80% of EM residents had listened to medical podcasts at least once [25].

Traditional lectures continue to be replaced by various digital teaching methods and this was hastened by the arrival of COVID-19. Podcasts' major benefit is their customization to fit learner's educational goals as well as time constraints, allowing users to optimize their study goals while balancing work and private life.

Feasibility of Implementation

In terms of feasibility, the podcast required a dedicated amount of time and monetary investment. The cost of standard microphones, basic recording software, and a website to host the podcast required approximately US $300 to 400 annually. As discussed in the methods section, the hourly commitment was close to 5 - 10 hours weekly.

Next Steps

A review of "Learning Through Listening: A Scoping Review of Podcast Use in Medical Education" examines podcasts for learning across many specialties, most often referencing anesthesia, with some reference to EM [26]. The data cited an increase in retention of information pre- and posttest for medical students, who are not specialized in EM compared with the level of a resident or attending physician. The review briefly mentions a podcast that improved EM in-training exam scores and a podcast that reportedly worsened in-training exam scores. The data gleaned from this study are of interest, but due to varied in-training exam scores, a comparative study is needed that examines test performance matching which podcast was used most for learning. Another future area of study will be to observe if the effects of the COVID-19 pandemic on asynchronous web-based learning are long-term.

Limitations

Our study is limited in generalizability due to it only measuring one specific podcast and website platform. A restricted sample size is one limitation of this study. Spotify and Android (Google) do not publish podcast statistics nor track individual usage, and therefore user data from both these platforms could not be obtained. According to Reuters in a survey of 2012 listeners, 20% used Apple Podcasts as their app of choice from 2019 - 2020, which is the second largest market share [27]. Previous studies have used podcast episode downloads as a metric for engagement. Despite the appeal of using number of downloads as a measurement, accurate analytics are difficult to obtain and fraught with bias. Downloads are defined differently depending on the podcast host. In addition, there have been reports that these numbers can be unreliable due to bot traffic and there can be manipulation of download data by hosts [28,29].

Another limitation is association versus causation. Given the retrospective study design and nature of COVID-19, it is difficult to completely credit the pandemic for increased podcast engagement. Confounding variables could also be a limitation, such as increased usage of social media during quarantine resulting in better promotion of the podcast and website.

One potential confounding variable was the launch of a procedural module in May 2020. This web-based learning instruction was an airway module, with recorded intubation videos and a pre- and postassessment. However, when reviewing website analytics, this was not a frequently viewed page on the website, accounting for only 2.59% of total website page views. It cannot entirely account for the sudden increase in website

XSL•FO
RenderX

visitors and podcast listeners. Thus, in this study, we can only establish differences observed in analytics between 2 time periods.

No quantitative data were tracked regarding listener exam performance, in particular in-training or board examinations. The purpose of this study was to assess the level of engagement for an EM board review podcast and website platform, before and during the COVID-19 pandemic. Future research should be aimed at assessing whether this educational intervention is an effective form of test preparation.

Conclusion

During the COVID-19 pandemic, there was an accelerated level of engagement for our EM board review podcast and website platform over a long-term period. This educational platform is a feasible, low-cost asynchronous study tool. Medical educators should be aware of the increasing usage of web-based education tools, and that asynchronous learning is favorably viewed by learners.

Acknowledgments

Statistical analysis by MM reported in this publication was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under award UL1TR001417.

Conflicts of Interest

None declared.

References

1. Apple iTunes Preview webpage. URL: https://podcasts.apple.com/us/genre/podcasts-health-fitness-medicine/id1518 [accessed 2020-02-07]
2. Sandars J, Correia R, Dankbaar M, et al. Twelve tips for rapidly migrating to online learning during the COVID-19 pandemic. MedEdPublish (2016) 2020;9(1):82. [doi: 10.15694/mep.2020.000082.1] [Medline: 38058947]
3. Theoret C, Ming X. Our education, our concerns: the impact on medical student education of COVID-19. Med Educ 2020 Jul;54(7):591-592. [doi: 10.1111/medu.14181] [Medline: 32310318]
4. Rana T, Hackett C, Quezada T, et al. Medicine and surgery residents' perspectives on the impact of COVID-19 on graduate medical education. Med Educ Online 2020 Dec;25(1):1818439. [doi: 10.1080/10872981.2020.1818439] [Medline: 32924869]
5. Alla A, Kirkman MA. PodMedPlus: an online podcast resource for junior doctors. Med Educ 2014 Nov;48(11):1126-1127. [doi: 10.1111/medu.12556] [Medline: 25307665]
6. Bhatti I, Jones K, Richardson L, Foreman D, Lund J, Tierney G. E-learning vs lecture: which is the best approach to surgical teaching? Colorectal Dis 2011 Apr;13(4):459-462. [doi: 10.1111/j.1463-1318.2009.02173.x] [Medline: 20041922]
7. Wilcha RJ. Effectiveness of virtual medical teaching during the COVID-19 crisis: systematic review. JMIR Med Educ 2020 Nov 18;6(2):e20963. [doi: 10.2196/20963] [Medline: 33106227]
8. Gopalan C, Butts-Wilmsmeyer C, Moran V. Virtual flipped teaching during the COVID-19 pandemic. Adv Physiol Educ 2021 Dec 1;45(4):670-678. [doi: 10.1152/advan.00061.2021] [Medline: 34498940]
9. Boreskie PE, Chan TM, Novak C, et al. Medical education blog and podcast utilization during the COVID-19 pandemic. Cureus 2022 Mar;14(3):e23361. [doi: 10.7759/cureus.23361] [Medline: 35475051]
10. Tintinalli JE, Stapczynski J, Ma OJ, Yealy D, Meckler G, Cline D. Tintinalli's Emergency Medicine: A Comprehensive Study Guide, 8th edition: McGraw-Hill Education; 2016.
11. UpToDate. URL: https://www.wolterskluwer.com/en/solutions/uptodate [accessed 2022-02-13]
12. EB Medicine. URL: https://www.ebmedicine.net [accessed 2022-04-05]
13. Beeson MS, Ankel F, Bhat R, et al. The 2019 model of the clinical practice of emergency medicine. J Emerg Med 2020 Jul;59(1):96-120. [doi: 10.1016/j.jemermed.2020.03.018] [Medline: 32475725]
14. Apple podcasts for creators. Understanding your analytics. URL: https://podcasters.apple.com/support/2553-understanding-your-subscriptions-reports [accessed 2022-02-16]
15. Squarespace. Traffic analytics. URL: https://support.squarespace.com/hc/en-us/articles/217999797 [accessed 2022-02-16]
16. ACGME. Updated: coronavirus (COVID-19) and ACGME site visits, educational activities, and other meetings. URL: https://www.acgme.org/newsroom/2020/3/updated-coronavirus-covid-19-and-acgme-site-visits-educational-activities-and-other-meetings [accessed 2022-02-17]
17. Nasca TJ. ACGME's early adaptation to the COVID-19 pandemic: principles and lessons learned. J Grad Med Educ 2020 Jun;12(3):375-378. [doi: 10.4300/JGME-D-20-00302.1] [Medline: 32595876]
18. WELCH BL. The generalisation of student's problems when several different population variances are involved. Biometrika 1947;34(1-2):28-35. [doi: 10.1093/biomet/34.1-2.28] [Medline: 20287819]
19. Levene H. Robust tests for equality of variances. In: Olkin I, Hotelling H, editors. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling: Stanford University Press; 1960:278-292.

XSL•FO
RenderX

20. Ayoubieh H, Alkhalili E, Nino D, Coue M, Herber-Valdez C, Pfarr CM. Analysis of pre-clerkship medical students' perceptions and performance during the COVID-19 pandemic. Med Sci Educ 2023 Feb;33(1):147-156. [doi: 10.1007/s40670-022-01723-6] [Medline: 36688011]
21. Noel ER, Mizuo B, Yamamoto LG, Ishikawa KM, Chen JJ, Len KA. Effect of the COVID-19 pandemic on medical student performance and evaluation scores. SVOA Med Res 2024;2(1):10-18. [doi: 10.58624/SVOAMR.2024.02.011] [Medline: 39144736]
22. Riddell J, Swaminathan A, Lee M, Mohamed A, Rogers R, Rezaie SR. A survey of emergency medicine residents' use of educational podcasts. West J Emerg Med 2017 Feb;18(2):229-234. [doi: 10.5811/westjem.2016.12.32850] [Medline: 28210357]
23. Cadogan M, Thoma B, Chan TM, Lin M. Free Open Access Meducation (FOAM): the rise of emergency medicine and critical care blogs and podcasts (2002-2013). Emerg Med J 2014 Oct;31(e1):e76-e77. [doi: 10.1136/emermed-2013-203502] [Medline: 24554447]
24. Berk J, Trivedi SP, Watto M, Williams P, Centor R. Medical education podcasts: where we are and questions unanswered. J Gen Intern Med 2020 Jul;35(7):2176-2178. [doi: 10.1007/s11606-019-05606-2] [Medline: 31898131]
25. Gottlieb M, Riddell J, Cooney R, King A, Fung CC, Sherbino J. Maximizing the morning commute: a randomized trial assessing the effect of driving on podcast knowledge acquisition and retention. Ann Emerg Med 2021 Sep;78(3):416-424. [doi: 10.1016/j.annemergmed.2021.02.030] [Medline: 33931254]
26. Kelly JM, Perseghin A, Dow AW, Trivedi SP, Rodman A, Berk J. Learning through listening: a scoping review of podcast use in medical education. Acad Med 2022 Jul 1;97(7):1079-1085. [doi: 10.1097/ACM.0000000000004565] [Medline: 34935729]
27. eMarketer. Most commonly used apps for listening to podcasts among podcast listeners in the united states in 2019 and 2020. URL: https://www.statista.com/statistics/943537/podcast-listening-apps-us [accessed 2022-02-16]
28. RSS.com. What cause the spike in my podcast analytics? URL: https://rss.com/blog/what-caused-the-spike-in-my-podcasts-analytics [accessed 2024-08-28]
29. Tani, Max. Semafor.com. The bots have come for podcasts. URL: https://www.semafor.com/article/10/08/2023/the-bots-have-come-for-podcasts [accessed 2024-08-28]

Abbreviations

ABEM: American Board of Emergency Medicine
EM: emergency medicine
EMBB: Emergency Medicine Board Bombs

Edited by B Lesselroth; submitted 05.03.24; peer-reviewed by AP Johnson, A Hosny, E Vashishtha; revised version received 28.10.24; accepted 02.01.25; published 21.02.25.

Please cite as:
Briggs B, Mulekar M, Morales H, Husain I
Comparison of an Emergency Medicine Asynchronous Learning Platform Usage Before and During the COVID-19 Pandemic: Retrospective Analysis Study
JMIR Med Educ 2025;11:e58100
URL: https://mededu.jmir.org/2025/1/e58100
doi:10.2196/58100

© Blake Briggs, Madhuri Mulekar, Hannah Morales, Iltifat Husain. Originally published in JMIR Medical Education (https://mededu.jmir.org), 21.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring Social Media Use Among Medical Students Applying for Residency Training: Cross-Sectional Survey Study

Simi Jandu*, MD; Jennifer L Carey*, MD

Department of Emergency Medicine, University of Massachusetts T.H. Chan School of Medicine, Worcester, MA, United States

*all authors contributed equally

Corresponding Author:
Simi Jandu, MD
Department of Emergency Medicine, University of Massachusetts T.H. Chan School of Medicine, Worcester, MA, United States

Abstract

Background: Since the COVID-19 pandemic, residency candidates have moved from attending traditional in-person interviews to virtual interviews with residency training programs. This transition spurred increased social media engagement by residency candidates, in an effort to learn about prospective programs, and by residency programs, to improve recruitment efforts. There is a paucity of literature on the effectiveness of social media outreach and its impact on candidates' perceptions of residency programs.

Objective: We aimed to determine patterns of social media platform usage among prospective residency candidates and social media's influence on students' perceptions of residency programs.

Methods: A cross-sectional survey was administered anonymously to fourth-year medical students who successfully matched to a residency training program at a single institution in 2023. These data were analyzed using descriptive statistics, as well as thematic analysis for open-ended questions.

Results: Of the 148 eligible participants, 69 (46.6%) responded to the survey, of whom 45 (65.2%) used social media. Widely used social media platforms were Instagram (19/40, 47.5%) and Reddit (18/40, 45%). Social media influenced 47.6% (20/42) of respondents' opinions of programs and had a moderate or major effect on 26.2% (11/42) of respondents' decisions on program ranking. Resident-faculty relations and social events showcasing camaraderie and wellness were the most desired content.

Conclusions: Social media is used by the majority of residency candidates during the residency application process and influences residency program ranking. This highlights the importance of residency programs in leveraging social media usage to recruit applicants and provide information that allows the candidate to better understand the program.

(JMIR Med Educ 2025;11:e59417) doi:10.2196/59417

KEYWORDS

social media; residency recruitment; Instagram; Reddit; medical students; student; residency; residency training; social media engagement; training programs; social media usage; cross-sectional survey; survey; residency training program; thematic analysis

Introduction

Since the 2020‑2021 residency application cycle, the Association of American Medical Colleges and Liaison Committee on Medical Education have recommended that programs conduct virtual interviews exclusively for residency applicants [1,2]. This recommendation allows for a more equitable residency application process, as it offloads financial and time burdens from the applicant involved with traveling, the applicant pool, bias, and interview flexibility; however, having an exclusively virtual process also comes with a loss of applicants developing rapport with faculty and residents and appreciating resident camaraderie, program culture, and what resident daily life is like [1,3].

Students and residency programs have both turned toward social media to lessen this void. Residency applicants turn to social media to gather more information about residency programs, such as details about the work environment and facilities, and residency programs use social media to promote their programs and institutions and highlight their culture, personnel, and network [4-8]. This has been shown to increase the number of programs applicants can apply to [9]. Residency programs must embrace this digital shift to adapt to the postpandemic landscape and efforts to enhance diversity and equity in medical education. Thus, social media remains an important platform for residency applicants and programs alike.

Despite its widespread usage, there is a lack of information on the impact that applicants across different specialties derive from residency programs' social media accounts. There have been single-specialty studies that have shown that social media is used by prospective applicants during the residency recruitment process, but limited studies across specialties have

XSL•FO
RenderX

been performed [4,10-12]. In this study, we performed a survey across multiple specialties to elucidate the patterns of social media consumption and its influences on medical students' selection of a residency program.

Methods

Population and Setting

The study population consisted of fourth-year medical students who graduated in 2023 from an allopathic medical school in Massachusetts and who participated in the residency match program during the 2022 - 2023 cycle.

Ethical Considerations

The survey was approved by the institutional review board and deemed not human research (STUDY00001121). The survey contained a description containing the risks of participation in the study, and completion of the survey implied voluntary, informed consent. No personally identifiable information was collected, and no incentives were offered.

Survey Development and Distribution

We developed the survey based on guidelines by Artino et al [13]. After a literature review, a focus group was held to learn more about medical students' use and opinions of social media. This information was synthesized, and survey items were created using a combination of a Likert scale, yes or no, and open-ended questions; the survey explored demographic data, the use of social media and types of platforms, preferred social media content, and the impact of social media on residency programs. The survey was reviewed by faculty and fellows and assessed for acceptability, feasibility, and content validity of survey questions. We performed cognitive interviews for the questions and then piloted and revised the survey for clarity based on user feedback from medical students. The survey had a total of 5 pages with 6 or less questions per page, and answers could be changed. The survey was then distributed to all students at our institution who graduated in May 2023.

All fourth-year medical students who graduated in 2023 at UMass Chan Medical School were eligible for the survey and emailed a link to the anonymous electronic survey (Multimedia Appendix 1). Study data were collected and managed using the Qualtrics XM platform. The survey was distributed in May 2023 and was open for 28 days. In total, 5 reminders were sent to nonrespondents and nonfinishers at 3, 7, 14, 18, and 23 days. To avoid duplicates, each participant was sent an individual link via Qualtrics.

Outcomes Measured

The outcomes measured included demographic data; social media platform use (platforms that were used daily, platforms used for residency programs, and the influence of social media on stages of the residency application process); content posted on social media platforms (student content that was trusted, not trusted, desired, deterrents, and then helpful); nonsocial media resources used for learning about residency programs; and reasons why participants did not use social media.

Data Analysis

We performed simple descriptive statistics for survey questions. Nominal variables were reported as percentages and frequencies. Ordinal variables were presented as percentages. Data analysis was conducted using Prism GraphPad (version 9.5.1).

We performed a thematic analysis using an inductive constructivist approach on deidentified responses to open-ended questions of fully completed questionnaires [14,15]. Coders (SJ and JLC) independently reviewed responses via open coding, systematically generating a preliminary list of codes for each question. Using methods outlined by Nowell et al [15], these were merged into concepts, and themes were generated via constant comparison, returning to raw data, and iterative modification to develop a consensus on themes.

Results

There were 69 respondents out of 148 eligible students in our study. Of these, 5 were excluded from the analysis because of incomplete survey responses, with a completion rate of 92.8%. The median age of survey respondents was 27 years old (range 24 - 39; IQR 27 - 29 years). In this cohort, 42.9% identified as a man and 57.8% identified as a woman. Further, 100% of respondents matched in the 2022 - 2023 cycle, and 73.4% matched in the Northeast Region. The most popular specialties were internal medicine (25%), pediatrics (15.6%), and emergency medicine (9.4%) (Table 1).

XSL•FO
RenderX

Table . Demographic characteristics.

Characteristics	Participants, n (%)
Gender	
Women	37 (57.8)
Men	27 (42.9)
Transgender or nonbinary	0 (0)
Race or ethnicity	
American Indian or Alaska Native	0 (0)
Asian	13 (20.3)
Black or African American	0 (0)
Native Hawaiian or other Pacific Islander	0 (0)
Hispanic White	2 (3.1)
Non-Hispanic White	44 (68.8)
Other-Hispanic	2 (3.1)
Multiracial-Asian or White	3 (4.7)
Specialty	
Anesthesiology	3 (4.7)
Emergency medicine	6 (9.4)
Family medicine	4 (6.3)
Internal medicine	15 (25.0)
Internal medicine—pediatrics	1 (1.6)
Neurological surgery	1 (1.6)
Neurology	1 (1.6)
Obstetrics-gynecology	5 (7.8)
Ophthalmology	2 (3.1)
Orthopedic surgery	2 (3.1)
Otolaryngology	1 (1.6)
Pathology	1 (1.6)
Pediatrics	10 (15.6)
Psychiatry	4 (6.3)
Radiation oncology	1 (1.6)
Radiation—diagnostic	3 (4.7)
Surgery—general or preliminary	3 (4.7)
Region matched	
Northeast	47 (73.4)
Southeast	1 (1.6)
West	3 (4.7)
Southwest	7 (10.9)
Midwest	6 (9.4)
Social media platforms daily use	
Discord	3 (4.7)
Facebook	20 (31.8)
Instagram	46 (73.0)
Reddit	12 (19.1)

XSL•FO
RenderX

Characteristics	Participants, n (%)
Snapchat	18 (28.6)
Twitter (X)	10 (15.6)
TikTok	14 (22.2)

The primary social media platform used was Instagram, with 73% (46/63) reporting daily use, followed by Facebook (20/63, 31.8%) and Snapchat (18/63, 28.6%) (Table 1). Among the respondents, 65.2% (45/69) reported using social media to learn about prospective residency programs. The most frequently used platforms for this purpose were Instagram, Reddit, and YouTube (Figure 1). Facebook, Snapchat, and TikTok were rarely or never used to learn about residency programs.

Social media had a moderate or major effect on 47.6% (20/42) of respondents' opinions about programs, while it had a lesser effect on respondents' decision to apply (6/42, 14.3%) or interview (5/42, 11.9%) at a program. However, 26.2% (11/42) of respondents indicated that social media had a moderate or major effect on their decision to rank a program (Figure 2).

Figure 1. Frequency of social media platform usage when learning about residency programs.

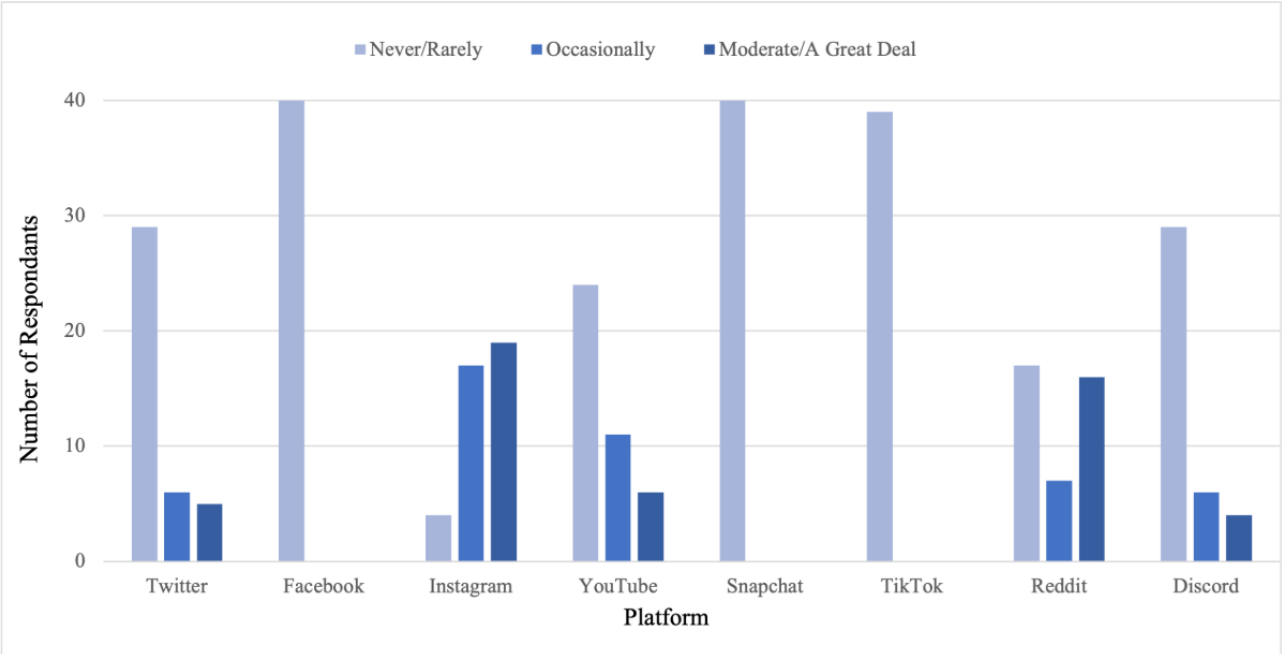
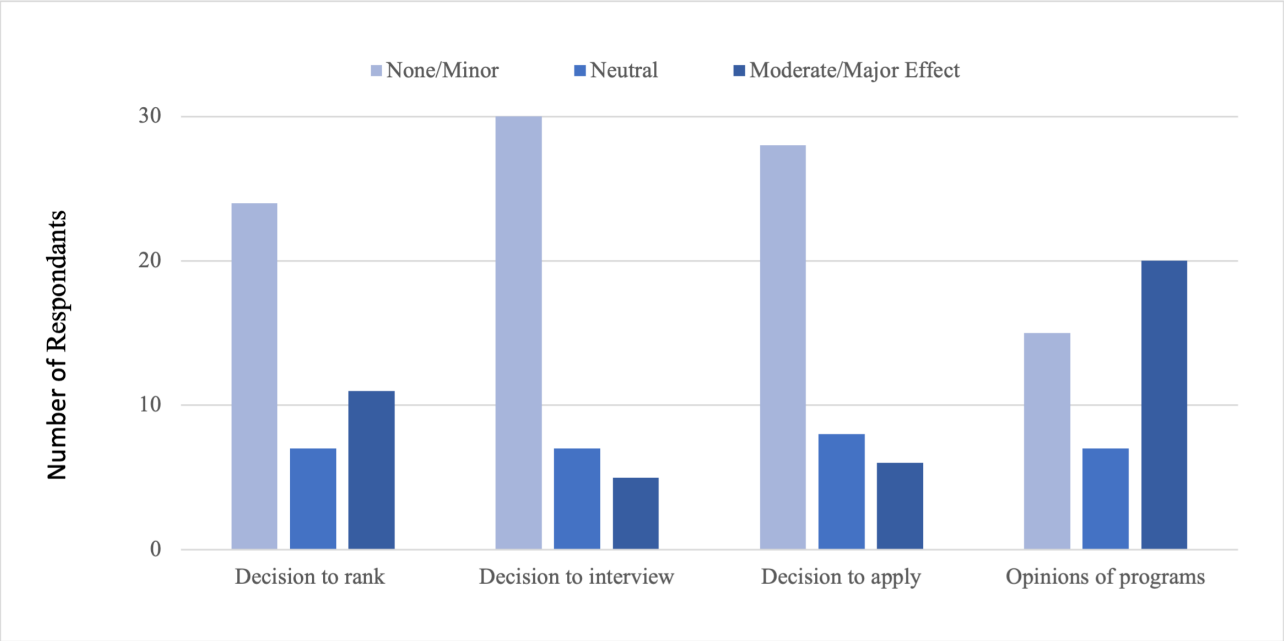


Figure 2. Social media influence on students' residency application process.



XSL•FO
RenderX

Among the 34.4% (22/64) of respondents who did not use social media, common reasons for abstaining included a perception that information could be easily distorted, social media can lead to mistrust and misrepresentation of the programs, lack of personal social media accounts, and limited usefulness or applicability of information resulting from the variable quality of social media posts. The sources trusted were the official program website, Fellowship and Residency Electronic Interactive Database Access (FREIDA), word of mouth, and personal contacts (Table 2).

Table . Resources used and reasons that students did not use social media.

	Themes
Resources used	• Program website • FREIDA[a] • Word of mouth • Current or past residents • Institutional contacts
Reasons that students did not use social media	• Mistrust of social media • Programs could be misrepresented • Irrelevant and not applicable postings that were unhelpful • Posts were not authentic • Respondents did not have or use social media personally and professionally • Variability in posts

[a]FREIDA: Fellowship and Residency Electronic Interactive Database Access.

The content deemed most valuable was resident and faculty relations (89.5%), followed by social events (86.8%) and education (60.5%) (Table 3). Thematic analysis revealed that students were attracted to posts emphasizing camaraderie, diversity, resident wellness, a genuine representation of the program's personality, and program and curriculum information. Participants particularly valued "Day in the Life" posts that visually depicted people, facilities, and location, as well as content that focused on personal experiences or resident wellness, highlighted unique attributes of the program, and provided information on the application process (Table 4).

Table . Desired social media content when researching residency programs.

Content	Participants, n (%)
Social events	33 (86.8)
Research	13 (34.2)
Didactics	12 (31.6)
Education	23 (60.5)
Resident and faculty relations	34 (89.5)
Other	2 (5.3)

XSL•FO
RenderX

Table . Themes and quotes of content that students deemed helpful, attractive, deterrents, trusted, and not trusted.

Themes		Quotes
Content that attracted students to a program		
	Camaraderie among residents and faculty; diversity of the program	"Seeing the residents spending time together and enjoying it, even if they were posting them working on the floor together"
	Resident wellness	"Camaraderie evident on posts"
	Content invoking a genuine feeling and showcasing its personality; informational content about the program and curricula	"Multiple social events with residents of all classes, photos with attendings and residents together"
Content that students deemed helpful in learning more about programs		
	Visualizing people, facility, and location; showcasing program's unique features; informational posts on the program; "Day in the Life"; resident wellness including what they do in their time off; personal experiences about a program; focus on advocacy	"When it was active, showed personality of programs/residents"
Content that deterred students		
	Minimal to no social media posts; lack of representation of multiple resident and faculty; negative personal anecdotes	"Less activity or presence on social media left an impression of overworked residents who didn't have time to post, or programs with less wellness/bonding activities to show off"
	Perception of ingenuine postings	"Too much of one person"
	Lack of types of content: photos and resident highlights	"No photos of residents and attendings"
Content trusted by students		
	Personal endorsements; resident-driven content; perspectives from residents and applicants; anonymous online platforms	"Reddit, student doctor network, because people post anonymously and be honest about the negative aspects of their program"
Content not trusted by students		
	Curated social media posts	"Social media posts are curated, and I don't trust that it's a true reflection of the day-to-day or vibe of the program"
	Program websites; promotional videos	"The promo videos for each program on the website because you can highlight very small parts."
	Nonanonymous content (subject to bias)	"Statements about the quality of the program, if they are happy, etc. I would not trust because the residents know the recordings will be posted. Hard to be honest when you can't be anonymous."
	Reddit, Discord, and chat/discussion boards (subject to bias)	"Chat/discussion boards due to potential for bias"
	Content from individuals outside the program; content from nonprogram accounts	"Message boards like Reddit I consider less reliable as anyone could share their experience with a good versus bad interview, I find these sites are very polarizing good or bad."

In a thematic analysis among all participants, anonymous digital platforms such as Reddit and Discord were considered trustworthy by some (n=3), although others perceived them as subject to bias and could be polarizing (n=15). Respondents also reported that nonanonymous content could also be subject to bias, noting that individuals posting may not want to be honest about negative aspects of a program when posting anonymously. Content from individuals outside the program and from nonprogram accounts was generally not trusted, as content could be posted and "filled with trolls," individuals who post intentionally provocative or inflammatory content. Finally, curated social media posts, program websites, and promotional videos were also listed among content that was not trusted (n=8), as it may only highlight certain aspects of the program (Table 4).

XSL•FO
RenderX

Discussion

Principal Findings

Our study revealed that Instagram was the most commonly used social media platform within our cohort. Instagram has experienced the greatest growth among new residency-specific social media accounts since March 2020, and its predominant demographic characteristics are similar to those of most prospective residents [12,16-18]. It has also been cited to be the most used platform, compared to Facebook and Twitter [19].

Interestingly, although the majority of students in this cohort reported using Facebook daily or weekly, it was almost never used to learn about residency programs by our respondents. Despite Facebook being commonly cited and compared to other platforms, it was only used more than Twitter by family medicine applicants. Otolaryngology, anesthesia, and plastic surgery applicants all used Facebook less than Twitter, with all specialties citing Instagram as the most used platform [4,7,19-21]. Facebook as a platform was shown to have the least growth, the least total number of accounts across specialties, and the least utilization among most specialties in comparison to Instagram and Twitter [17,19,20]. It is unclear why students did not use Facebook to learn about residency programs, despite their overall frequent use. One possible reason is the lack of Facebook posts by residency programs. To maximize the effectiveness of their social media presence, programs might consider focusing on Instagram rather than Facebook or linking the 2 platforms, thereby reaching 2 platforms with 1 post.

We found Reddit to be the second most popular platform. Reddit has been used by anesthesia and emergency medicine applicants as sources of information but was not this highly ranked by prior studies [20,22]. Its design facilitates ease of information exchange, and its built-in anonymity affords users the opportunity to post content without fear of repercussions. Students acknowledged that while anonymity introduces the potential for bias, anonymous online chat and discussion boards still have the potential to be trustworthy sources of information. Additionally, it is worthwhile for programs to note that negative anecdotes published on Reddit or similar platforms can deter students from programs and can be seen by those without social media accounts. The increasing popularity of Reddit suggests that it is a worthwhile avenue for social media outreach during the residency application season [22].

As social media influence the residency process, respondents are affected by their opinion and rank of a program. Social media can positively influence the opinions of programs, congruent with prior urology, otolaryngology, and plastic surgery studies, with a quarter of students' decisions affected by social media when creating their "rank list" [7,8,23]. However, as compared to Naaseh et al [9], who found 74% of respondents increased the number of programs they applied to due to social media, we found no significant effect when applying to programs found in our study despite the positive overall impression of the program in our study along with anesthesia, general surgery, and family medicine [4,9,11,20]. Based on studies, social media can influence opinion and rank of a program, which may ultimately change where a student matches for residency and whether a residency program is able to fill all its residency positions.

Regardless of surgical or nonsurgical specialty, posts that showcase resident and faculty relations, social events, and educational material are seen as the most desired content. As seen in our study, applicants desire a sense of camaraderie and resident wellness where "the residents are spending time together and enjoying it even if they were posting them working on the floor together" [19,21]. Consistent with prior studies, applicants are interested in the resident life in and outside of the hospital [19]. "Day in the Life" posts, where residents showcase a typical working day, can help students understand what their day-to-day life will be like at a particular program. They can also help to showcase aspects of the program that are difficult to show within the virtual interview setting, such as personnel interactions, diversity, and wellness [11,21]. Finally, they can be an adjunct to highlight specific program information, including curricula, electives, rotations, research, conferences, and even interview-specific information. These are all aspects sought by students in their evaluation of a program's social media presence and can be leveraged in the recruitment of residency candidates.

Importantly, social media can also have a negative influence on prospective applicants. Our study shows that social media accounts that do not consistently post or save content can leave the "impression of overworked residents who did not have time to post, or programs with less wellness or bonding activities to show off," consistent with prior investigations [24]. Negative anecdotes and comments left on anonymous platforms by single individuals, although possibly isolated, nonrepresentative experiences, can have a profound negative influence on an applicant's perception of a program, and it can be exceptionally difficult to correct these views. Programs must keep in mind that the amount and content a program posts and anonymous negative anecdotes can contribute to a negative opinion of a program, potentially affecting the application and rank process.

In our cohort, approximately one-third of the applicants did not use social media and reported using other resources. Thus, it is imperative to ensure that the official program website and Google are updated and accurate. Traditional resources include FREIDA, Doximity, word of mouth, current and past residents, and contacts within the institution [4,11,12]. These are all highly trusted content used by both social media users and nonusers.

With Instagram being the most popular social media platform, residency programs would likely benefit the most from using Instagram as their main social media platform [9,12,19,20]. Although an exact threshold is unknown, students associate less frequent Instagram posts with decreased resident wellness, and frequently posting information about residents, faculty, and program information is important for a program's image. High-impact posts might feature a particular resident for a "Day in the Life," social activities both inside and outside of the work environment, and highlights from resident wellness days and resident-faculty interactions. These can be categorized and saved, enabling prospective applicants to easily view them at later dates. Efforts should focus on creating authentic posts that showcase the people, diversity, and culture of the program in a

XSL•FO
RenderX

fun manner while taking care to avoid professional, ethical, and legal violations [25]. Students want a glimpse of what it is like to be a resident at a particular program, and posts containing pictures and videos can enable them to see and understand the program better in the current landscape of recruitment.

Limitations

As this is a survey-based study, the survey is subject to selection and response bias, with potential inaccuracies in the participants' recollection of their social media usage and influence in the residency application process. However, this is the first survey across specialties to delve into the social media usage throughout their residency application process; it has to be done after Match day. We attempted to limit response bias by ensuring anonymity and distributing this survey between match day and prior to graduation to all students in this class. We had no responses that indicated they did not match and did not ask whether any applicants went through the SOAP (Supplemental Offer and Acceptance Program) process. Thus, it is possible that those likely to respond may have been those who used social media throughout the application process and those who did not have to go through the SOAP process. Further studies could look at social media use in those that matched during the NRMP (National Resident Matching Program) or the SOAP process to see if there was a difference. Questions could also directly ask about positive and negative influence to gain more information on the drawbacks of social media while being a neutral question stem.

Although the survey was developed based on the guidelines of Artino et al [13], it needs to be further validated in the future. This was a single-center study from an allopathic medical school, limiting the generalizability of the findings, as social media patterns may vary among regions and medical schools. Thus, a multi-institutional study that examines applicants' use of social media throughout their application, interview season, and ranking process is needed to further elucidate information to be used by programs. Studies could delineate the widespread use of social media by specialties, as well as whether the applicants matched into their specialty of choice or not. Specific content that students are interested in could also be looked at for specialty (ie, procedural-based vs non–procedural-based specialties or adult vs pediatric specialties).

Conclusions

This study offers important insights into the effects of social media on residency recruitment from the student perspective. Students use social media platforms, specifically Instagram, to make informed decisions in their residency application process; therefore, programs can use these platforms to augment their recruitment. This information can help programs develop their social media platforms to cater to their target audience and mitigate the potential negative influence of social media. With the increasing popularity of social media among this generation of applicants, its use in the residency match process is expected to increase, with the current leading social media platforms being Instagram and Reddit.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Social media survey.
[DOCX File, 31 KB - mededu_v11i1e59417_app1.docx]

References

1. Copeland L, D'Antonio J, Dell M, et al. Interviews in GME: where do we go from here? Association of American Medical Colleges. 2023. URL: https://www.aamc.org/about-us/mission-areas/medical-education/interviews-gme-where-do-we-go-here [accessed 2023-09-06]
2. 2020-2021 recruitment cycle: issues for programs considering diversity and the COVID-19 pandemic. Accreditation Council for Graduate Medical Education. 2020. URL: https://www.acgme.org/globalassets/Diversity-and-COVID-19.pdf [accessed 2023-09-06]
3. Wang S, Denham Z, Ungerman EA, et al. Do lower costs for applicants come at the expense of program perception? A cross-sectional survey study of virtual residency interviews. J Grad Med Educ 2022 Dec;14(6):666-673. [doi: 10.4300/JGME-D-22-00332.1] [Medline: 36591433]
4. Oliver MG, Kelly K. Student perceptions and use of social media as residency program information. Fam Med 2022 May;54(5):380-383. [doi: 10.22454/FamMed.2022.968351] [Medline: 35536623]
5. Vallejo MC, Price SS, Vanek TW, et al. Virtual interviewing in the COVID-19 era: a survey of graduate program directors. J Dent Educ 2022 May;86(5):535-542. [doi: 10.1002/jdd.12848] [Medline: 35580990]
6. Lee E, Terhaar S, Shakhtour L, et al. Virtual residency interviews during the COVID-19 pandemic: the applicant's perspective. South Med J 2022 Sep;115(9):698-706. [doi: 10.14423/SMJ.0000000000001442] [Medline: 36055658]
7. Patro A, Landeen KC, Stevens MN, Cass ND, Haynes DS. The digital dilemma: perspectives from otolaryngology residency applicants on social media. Ann Otol Rhinol Laryngol 2022 Sep;131(9):954-961. [doi: 10.1177/00034894211050625] [Medline: 34617461]
8. Irwin TJ, Riesel JN, Amador RO, Helliwell LA, Lin SJ, Eberlin KR. The impact of social media on plastic surgery residency applicants. Ann Plast Surg 2021 Mar 1;86(3):335-339. [doi: 10.1097/SAP.0000000000002375] [Medline: 32349083]

XSL•FO
RenderX

9. Naaseh A, Thompson S, Tohmasi S, et al. Evaluating applicant perceptions of the impact of social media on the 2020-2021 residency application cycle occurring during the COVID-19 pandemic: survey study. JMIR Med Educ 2021 Oct 5;7(4):e29486. [doi: 10.2196/29486] [Medline: 34591779]
10. Fick L, Palmisano K, Solik M. Residency program social media accounts and recruitment - a qualitative quality improvement project [version 1]. MedEdPublish 2020;9(1):203. [doi: 10.15694/mep.2020.000203.1]
11. Fuller CC, Deckey DG, Brinkman JC, et al. General surgery residency applicants' perspective on social media as a recruiting tool. J Surg Educ 2022;79(6):1334-1341. [doi: 10.1016/j.jsurg.2022.06.003] [Medline: 35739022]
12. Li TM, Tepper DL, Burger AP, Weissman MA. Internal medicine recruitment in the age of social media: strategies to target residency applicants. Mayo Clinic Proc Digit Health 2023 Jun;1(2):55-59. [doi: 10.1016/j.mcpdig.2023.02.001]
13. Artino AR, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE Guide No. 87. Med Teach 2014 Jun;36(6):463-474. [doi: 10.3109/0142159X.2014.889814] [Medline: 24661014]
14. Braun V, Clarke V. Thematic analysis. In: APA Handbook of Research Methods in Psychology, Vol 2 Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological: American Psychological Association; 2012:57-71. [doi: 10.1037/13620-004]
15. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017;16(1). [doi: 10.1177/1609406917733847]
16. Checketts JX, Hunt T, Checketts BR, et al. Analysis of social media perceptions among orthopaedic surgery residency applicants and social media use by residency programs during the 2020 to 2021 cycle. JB JS Open Access 2021;6(4):e21.00083. [doi: 10.2106/JBJS.OA.21.00083] [Medline: 34957367]
17. Singh NP, DeAtkine AB, Hattaway RH, et al. Changes in United States residency program online presence following COVID-19. Teach Learn Med 2023;35(2):157-167. [doi: 10.1080/10401334.2022.2047050] [Medline: 35689361]
18. Dixon SJ. US Instagram users 2024, by age group. 2024 Nov 28. URL: https://www.statista.com/statistics/398166/us-instagram-user-age-distribution [accessed 2025-02-04]
19. Plack DL, Abcejo AS, Kraus MB, Renew JR, Long TR, Sharpe EE. Postgraduate-year-1 residents' perceptions of social media and virtual applicant recruitment: cross-sectional survey study. Interact J Med Res 2023;12(1):e42042. [doi: 10.2196/42042]
20. Dunn T, Patel S, Milam AJ, Brinkman J, Gorlin A, Harbell MW. Influence of social media on applicant perceptions of anesthesiology residency programs during the COVID-19 pandemic: quantitative survey. JMIR Med Educ 2023 Jun 29;9:e39831. [doi: 10.2196/39831] [Medline: 37205642]
21. Pflibsen LR, Deckey DG, Brinkman JC, Tummala SV, Casey WJ, Teven CM. The effects of website and social media presence of integrated plastic surgery residency programs on prospective applicants. Ann Plast Surg 2022;88(6):599-605. [doi: 10.1097/SAP.0000000000003064]
22. Mackey C, Feldman J, Peng C, Way DP, Messman A. How do emergency medicine applicants evaluate residency programs in the post-COVID-19 era? AEM Educ Train 2022 Dec;6(6):e10805. [doi: 10.1002/aet2.10805] [Medline: 36389651]
23. Ho P, Margolin E, Sebesta E, Small A, Badalato GM. #AUAMatch: the impact of COVID-19 on social media use in the urology residency match. Urology 2021 Aug;154:50-56. [doi: 10.1016/j.urology.2021.05.019] [Medline: 34033828]
24. Chandawarkar AA, Gould DJ, Stevens WG. Insta-grated plastic surgery residencies: the rise of social media use by trainees and responsible guidelines for use. Aesthet Surg J 2018 Sep 14;38(10):1145-1152. [doi: 10.1093/asj/sjy055] [Medline: 29474525]
25. Duque S, Riccelli V, Mulqueen S, Zhang AY. Global pandemic and plastic surgery residency match: can social media fill the void? Aesthet Surg J 2021 Oct 15;41(11):NP1747-NP1753. [doi: 10.1093/asj/sjab222] [Medline: 33970220]

Abbreviations

FREIDA: Fellowship and Residency Electronic Interactive Database Access
NRMP: National Resident Matching Program
SOAP: Supplemental Offer and Acceptance Program

Edited by B Lesselroth; submitted 11.04.24; peer-reviewed by JGZ Rodriguez, K Aguirre, MA Ream; revised version received 21.08.24; accepted 24.09.24; published 21.02.25.

Please cite as:
Jandu S, Carey JL
Exploring Social Media Use Among Medical Students Applying for Residency Training: Cross-Sectional Survey Study
JMIR Med Educ 2025;11:e59417
URL: https://mededu.jmir.org/2025/1/e59417
doi:10.2196/59417

XSL•FO
RenderX

© Simi Jandu, Jennifer L Carey. Originally published in JMIR Medical Education (https://mededu.jmir.org), 21.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Ethical Use of Social Media and Sharing of Patient Information by Medical Students at a University Hospital in Saudi Arabia: Cross-Sectional Survey

Sara Farsi, MD, MEd; Alaa Sabbahi, MD; Deyala Sait, MBBS; Raghad Kabli, MBBS; Ghaliah Abduljabar, MBBS

Department of Anesthesia and Critical Care, Faculty of Medicine, King AbdulAziz University, KAUH, Jeddah, Saudi Arabia

Corresponding Author:
Sara Farsi, MD, MEd
Department of Anesthesia and Critical Care, Faculty of Medicine, King AbdulAziz University, KAUH, Jeddah, Saudi Arabia

Abstract

Background: Social media (SM) has become an integral part of many medical students' lives, blurring the lines between their personal and professional identities as many aspects of their medical careers appear online. Physicians must understand how to responsibly navigate these sites.

Objective: This study aimed to identify how medical students use SM and their awareness and adherence to ethical guidelines of e-professionalism.

Methods: This is a cross-sectional study delivered as an online voluntary survey to senior medical students at King AbdulAziz University Hospital in Jeddah, Saudi Arabia. We investigated how many students used SM, their privacy settings, their possible breaches of ethical standards, and their portrayal of their training institute online.

Results: A total of 400/1546 (26%) senior medical students responded to our survey. Among the participants, 95/400 (24%) had public SM accounts, while 162/400 (41%) had both private and public accounts. As for breaches in e-professionalism, 11/400 (3%) participants posted a picture of a patient on SM without their permission, while 75/400 (20%) posted part of an excised organ or x-ray on SM without their permission, and 60/400 (16%) discussed a patient. With regards to sharing medical school information, 108/400 (29%) discussed an incident at their medical school, and 119/400 (31%) participants shared a lecture online without the presenter's permission. Approximately 66% of the participants reported that they were unaware if their institution had a professional code of conduct for SM use, and 259/371 (70%) did not receive training on the professional use of SM.

Conclusions: Medical students must be taught to recognize inappropriate online behavior, understand their role as representatives of their medical school, and know the potential repercussions of unprofessional conduct on SM. This could be accomplished by providing workshops, regular seminars on e-professionalism, and including principles of SM conduct in existing ethics courses.

(JMIR Med Educ 2025;11:e57812) doi:10.2196/57812

KEYWORDS

e-professionalism; professionalism; social media; medical education; curriculum development; privacy; confidentiality; ethics; patient confidentiality; cross-sectional; questionnaire

Introduction

Since its founding in 2006, the number of active users of Twitter (currently known as X) has increased to 237.8 million worldwide as of January 2023 [1]. Many medical students have grown up with online social media (SM) profiles. Studies conducted in Saudi Arabia have demonstrated that 75% - 87% of medical students use SM [2,3]. Owing to built-in camera-equipped smartphones, these students can now document their entire lives through pictures and videos and share them online like a public diary. Therefore, medical school is an integral part of their lives, and aspects of it are bound to find their way onto their SM profiles. However, do medical students understand the rules and implications of sharing that information online?

In the past decade, medical students have transitioned from discussing complex patient details with a few colleagues in the hospital's breakroom to doing so with hundreds of "followers" worldwide. During the COVID-19 crisis, SM played a major role by building bridges across health care communities, allowing physicians and patients to connect worldwide, exchange experiences, access the latest health recommendations, and provide and receive emotional support [4-6]. SM has even been used as an educational resource, with studies showing that most students use it to access or share learning material; 30% do not even use a textbook [7-9]. In addition, students may also share patient encounters, conflicts between staff, recordings of lectures, and other occurrences on these SM sites. These medical student posts eventually become a reflection of their profession and institution. The images they share are not always

XSL•FO
RenderX

complimentary. A cross-sectional study in the United States revealed 9 incidents of medical students posting negative information about their medical school online [10]. Furthermore, the same study revealed that 13% of those schools described a violation of patient confidentiality, and 4% of those incidents were reported by the patients or their families. Health care workers' online posts have also led to dismissals and lawsuits [4,11,12]. Moreover, several articles document unprofessional behavior by medical students online, including drinking and illicit drug use [4,10,12,13].

We hypothesize that many medical school curricula emphasize disease management and patient care, which are undeniably important. However, they have not fully evolved to address the complexities of the modern social and digital landscape, leaving students underprepared to navigate these challenges effectively. This gap can inadvertently contribute to lapses in judgment because students face situations for which they may not have been adequately equipped. Against this background, our study aimed to determine whether medical students shared unprofessional content related to patients or their medical school that could impact public perception of their institution or profession. Additionally, we sought to assess their awareness of and adherence to ethical standards of e-professionalism. A further objective was to compare our findings within the context of Saudi culture to those reported in previously published Western studies.

Methods

Study Design

This is a cross-sectional study that includes senior medical students and interns at King AbdulAziz University (KAU). Medical school in KAU lasts 6 years in addition to an internship year. We defined senior medical students as those in their 4th to 6th years of training. This group was selected because the earlier years of medical education focus primarily on lecture-based and laboratory-based basic sciences, with no direct patient exposure. We developed a 2-part, 19-item survey and included 3 demographic questions (age, gender, and year of training). The question content and design were based on our primary and secondary research goals. We developed our research questions through an extensive review of the literature, aiming to identify common challenges, breaches, and issues faced by medical students and medical schools in the context of SM use [4,10,12,14]. We identified common issues among medical students, including the sharing of confidential patient information—both textual and visual—on SM, as well as the dissemination of negative encounters experienced in their hospitals. Additionally, this study found that numerous lecturers faced consequences for remarks or actions during lectures that were unknowingly recorded by students and shared publicly [15-17]. This prompted us to investigate the frequency of teaching sessions being recorded without the lecturer's permission. Our survey questions were regarding sharing images of patients, parts of patients, colleagues, and lectures without permission. We also included questions on whether they discussed patients or incidents at their medical school online. To identify the effects students' online behavior may have on their professional image, we included questions that addressed students' profiles' privacy or anonymity (eg, Do you use your real name? Is your profile picture a clear image of yourself?), and link to their profession (eg, Do you mention the name of your institution? Do you identify your profession?). We revised the survey to ensure that the final questions were relevant, contained appropriate wording, and appeared in a logical order. A questionnaire was developed using the website Survey Monkey. The results could only be accessed by the principal researcher under a password-protected online account. We shared the survey with 10 medical students from the target group to ensure that all respondents would similarly interpret the questions as well as the usability and technical functionality of the survey platform. After piloting the survey, some questions were modified (in the question "what social media platform do you use regularly?" we added options such as Telegram, Discord, and Reddit). We also changed the wording of some questions to improve clarity. These 10 students were not included in this study's group. The final questionnaire consisted of 19 questions distributed over 4 pages (Multimedia Appendix 1). The questionnaire does not allow multiple responses for the whole duration of this study. If a student attempts to take the survey again using the same browser, they will see a message that they already took the survey. After final approval of the questionnaire and design, we invited senior medical students from years 4, 5, and 6 and the internship year to participate in the survey voluntarily through an open link. Members of the research team contacted the chief students of each academic year in person to explain the purpose and details of this study to share with all students in their year of training. Then, they sent the chief students a link to the survey via a WhatsApp (Meta Platforms, Inc) message to distribute to all students in their year individually. This message included the name and contact information of the principal researcher, the duration of the survey (3 min), and a link to the survey. The message also informed the students that their responses would be kept confidential, participation was completely voluntary, there was no incentive, and their evaluation and training would not be affected by their decision to participate in the study. We also included a QR code link on the last slide of anesthesia lectures given to the target group and invited the students to this study. The survey link was opened on August 10, 2022, and closed on June 16, 2023.

Descriptive statistics of variables were presented as counts and percentages to summarize the characteristics of the participants, including gender, age, and year of medical school. Chi-square tests assessed associations between categorical variables, and Fisher exact tests, as indicated. Univariate and multivariate logistic regression analyses were performed to identify predictors of cyberbullying exposure, with odds ratios (ORs) and 95% CI reported for each predictor. Variables included in the regression models were gender, age category, year of medical school, SM privacy status, time spent on SM, and training on the professional use of SM. Statistical analyses were performed using Stata (version 12.1 software, StataCorp LP). Cronbach α was used to measure internal consistency (0.75).

XSL•FO
RenderX

Ethical Considerations

We obtained institutional review board approval to conduct the study from KAU's Ethics Committee (reference #414- - 22). The online survey began with an informed consent statement that explained the purpose of the questionnaire and assured participants that all information would be kept confidential with no names or contact details recorded in the survey. Participation was entirely voluntary, with no reward for completing the survey and no penalty for choosing not to participate. The data were stored securely under password protection, and only the principal researcher had access.

Results

We distributed the survey to 1546 participants, of whom 400 responded, yielding a response rate of 26%. Survey completion rate was 86% and both incomplete and complete surveys were used in analysis. Approximately half of the participants were sixth-year students (194/400, 49%), and two-thirds were women (246/400, 62%), as illustrated in Table 1. Snapchat was the most used platform (287/400, 72%), followed by Twitter (275/400, 69%) and Instagram (256/400, 64%). Facebook was the least used platform (8/400, 2%), and only 8/400 (2%) of the participants reported not using any SM platform at least once a week.

Table . Characteristics of the participants (N=400).

Characteristic	Value
Gender, n (%)	
Male	154 (38.5)
Female	246 (61.5)
Age (years), n (%)	
18 - 20	3 (0.8)
21 - 25	378 (94.5)
26 - 30	18 (4.5)
>30	1 (0.2)
Year of medical school[a], n (%)	
Fourth	16 (4)
Fifth	142 (35.5)
Sixth	194 (48.5)
Intern	48 (12)
Platform used at least once a week, n (%)	
Facebook	8 (2)
TikTok	184 (46)
Snapchat	287 (71.8)
Twitter	275 (68.8)
Instagram	256 (64)
Reddit	32 (8)
Discord	30 (8)
Telegram	247 (61.8)
Own YouTube channel	25 (6.3)
None	8 (2)

[a]Has missing value for 1 participant.

Only 95/400 (24%) of the participants had public SM accounts, whereas 162/400 (41%) had a combination of private and public accounts. Most of the participants (307/400, 77%) used their real names on SM, and one-third used their own photos for their profile image (118/400, 30%). Approximately half of the participants used SM for more than 3 hours a day (180/400, 47%), whereas only 15/400 (4%) used it for less than 1 hour a day (Table 2). Most of the participants used SM for entertainment (340/400, 85%); some used it for networking with other professionals worldwide (91/400, 29%) and for staying in touch with family and friends (300/400, 75%).

XSL•FO
RenderX

Table . Description of social media use among the participants (N=400).

Variable	Participants
Privacy status of social media account, n (%)	
Public	95 (23.8)
Private	139 (34.8)
Some public, some private	162 (40.5)
Do not use social media	4 (1)
Privacy practices in social media use, n (%)	
Use of real name on social media	307 (76.8)
Use of a clear photo of self as a profile image	118 (29.5)
Identify as a King AbdulAziz University student	76 (19)
Identify as a medical student	127 (31.8)
None of the above	69 (17.3)
Time spent on social media, n (%)	
Less than 1 h/d	15 (3.9)
1 h/d	27 (7.1)
2 h/d	71 (18.6)
3 h/d	88 (23.1)
More than 3 h/d	180 (47.2)
Reason for social media use, n (%)	
Networking with other medical students or professionals worldwide	91 (22.8)
Keeping in touch with family or friends	300 (75)
Providing medical advice and advocacy	30 (7.5)
Entertainment	340 (85)
Medical education	172 (43)

Institution-related SM use practices are presented in Table 3. Regarding the professional use of SM, only 125/371 (34%) of the participants said they were aware that their institution had a professional code of conduct for SM use. Additionally, just 112/371 (30%) recalled having received training in the professional use of SM. Approximately one-third of the participants reported checking SM while rounding on patients (138/382, 36%), discussing an incident that occurred at their institution online (108/371, 29%), or uploading the content of a lecture or workshop online without the lecturer's permission (119/382, 31%). Only 11/380 (3%) posted pictures of patients on SM after obtaining the patient's permission, while 75/381 (20%) posted pictures of parts of a patient (x-ray, excised organ, etc) on SM without obtaining their permission.

XSL•FO
RenderX

Table . Number of participants who answered yes to questions regarding institution-related social media use practices, code of conduct, and training on social media use among the participants.

Survey question	Number of respondents[a], n	"Yes" response, n (%)
Does your institution have a professional code of conduct or protocol that addresses the use of social media?	371	125 (33.7)
Did you receive any training during medical school or residency on the rules and regulations for the professional use of social media?	371	112 (30.1)
Checked your social media account while rounding on patients	382	138 (36.1)
Posted a picture of a patient on social media without their permission	380	11 (2.9)
Posted an image of part of a patient (including excised tumors or organs) or a radiographic image of a patient without a patient's permission	381	75 (19.7)
Posted an image of a work colleague or senior without their permission	382	25 (6.5)
Uploaded a video or image of a lecture or workshop online without the lecturer's permission	382	119 (31.2)
Discussed an incident that happened in your institution online	371	108 (29)
Discussed a patient you saw at your institution online	372	60 (16.1)

[a]Some of the values do not add up to the total because of missing values.

Furthermore, many participants used apps to search for medical information (Table 4). The most common apps were YouTube (314/340, 92%; Google LLC) and AMBOSS (301/340, 75%; AMBOSS GmbH), followed by Osmosis (250/340, 74%; Elsevier) and UpToDate (235/340, 70%). Wikipedia (35/340, 10%; Wikimedia Foundation, Inc) and Medline (40/340, 12%; Medline Industries, LP) were the least commonly used sources.

Table . Applications used among the participants to look up medical information (N=340).

Application	Participants, n (%)
YouTube	314 (92.4)
Medline	40 (11.8)
UpToDate	235 (69.1)
Wikipedia	35 (10.3)
AMBOSS	301 (75.3)
Osmosis	250 (73.5)
Other:	40 (11.8)

Other sources were BMJ, Board and Beyond (McGraw Hill), Dr. Najeeb (DrNajeebLectures.com), MedED (PW MedEd), Kaplan, Google, ChatGPT (OpenAI), Lecturio, OnlineMedEd, Mayo Clinic (Mayo Foundation for Medical Education and Research [MFMER]), Medscape (WebMD LLC), Medicosis Perfectionalis, Radiopaedia, Healthline (Healthline Media LLC), NCBI StatPearls (National Library of Medicine), Orthobullet (Lineage Medical, Inc), WikEM, Telegram (Telegram FZ-LLC), and Twitter (X Corp).

The associations between SM use practices and gender are presented in Table 5. Women were more likely than men to have private SM accounts (96/248, 39% and 43/154, 28%, respectively; $P<.001$) and were less likely than men to use a clear photo of themselves for a profile image (45/248, 18% and 73/154, 47%, respectively; $P<.001$).

XSL•FO
RenderX

Table . The association between cyberbullying, social media privacy status, social media privacy practices, and gender among the respondents (N=400).

Survey question	Male	Female	*P* value
Experienced cyberbullying, n (%)			.13[a]
No	118 (80.3)	194 (86.2)	
Yes	29 (19.7)	31 (13.8)	
Social media privacy status, n (%)			.001[b]
Public	53 (34.4)	42 (17.1)	
Private	43 (27.9)	96 (39)	
Some public, some private	56 (36.4)	106 (43.1)	
Do not use social media	2 (1.3)	2 (0.8)	
Privacy practices in social media use, n (%)			
Use of real name in social media	117 (76)	190 (77.2)	.77[a]
Use of a clear photo of self as a profile image	73 (47.4)	45 (18.3)	<.001[a]
Identify as a King AbdulAziz University student	30 (19.5)	46 (18.7)	.85[a]
Identify as a medical student	46 (29.9)	81 (32.9)	.52[a]
None of the above	29 (18.8)	40 (16.3)	.51[a]

[a]Chi-square test.

[b]Fisher exact test.

Of all the participants, 60/400 (16%) reported experiencing cyberbullying. In univariate analyses, participants with private SM accounts were less likely to experience cyberbullying compared to those with public accounts (OR 0.40, 95% CI 0.2-0.9). Additionally, those spending more than 3 hours per day on SM had significantly higher odds (OR 3.36, 95% CI 1.0-11.5) of experiencing cyberbullying compared to those spending 1 hour or less per day. Same findings were found in multivariate analyses but became borderline significant (all had confidence intervals that narrowly include the null value).

Table 6 presents the association between patient privacy practices among the participants and the privacy status of the SM accounts. Participants who reported posting an image of part of a patient (including excised tumors or organs) or a radiograph were more likely to have a mix of public and private accounts (39/75, 52%) than public (21/75, 28%) or private accounts (15/75, 20%; *P*<.001). Among the participants who reported posting an image of a colleague without obtaining permission, 12/25 (48%) had a public account, whereas 8/25 (32%) and 5/25 (20%) had mixed and private accounts, respectively (*P*<.001). Moreover, participants who uploaded the content of a lecture online without the lecturer's permission were more likely to have a public account (37/119, 31%) than mixed (54/119, 45%) or private (28/119, 24%) accounts.

XSL•FO
RenderX

Table . Association between the privacy status of social media accounts and patient privacy practices.

Survey question	Privacy status of participants who answered "yes" to social media accounts			*P* value
	Public, n (%)	Private, n (%)	Mixed, n (%)	
Posted a picture of a patient on social media without their permission	5 (45.5)	2 (18.2)	4 (36.4)	.24
Posted an image of part of a patient (including excised tumors or organs) or a radiographic image of a patient without a patient's permission	21 (28)	15 (20)	39 (52)	.01
Posted an image of a work colleague or senior without their permission	12 (48)	5 (20)	8 (32)	.01
Uploaded a video or image of a lecture or workshop online without the lecturer's permission	37 (31.1)	28 (23.5)	54 (45.4)	.004
Discussed an incident that happened in your institution online	29 (26.9)	36 (33.3)	43 (39.8)	.68
Discussed a patient you saw at your institution online	20 (33.3)	20 (33.3)	20 (33.3)	.14

Discussion

Our study reveals that a substantial portion of students frequently share medical school-related content online, with notable instances of ethical breaches such as discussing patients and posting images without consent. While most published studies examine unprofessional online content posted by students, we investigate how often aspects of their medical school that may affect public perception appear on their profiles. These results underscore the urgent need for enhanced e-professionalism training. Of the students who responded to our survey, 246/371 (66%) were unaware of institutional guidelines addressing the use of SM, and 259/371 (70%) felt they had not received training on the professional use of SM. However, most students in our study (389/400, 97%) refrained from posting images of a patient online despite not having received e-professional training. Probably, they recognized this as a breach of the well-known Hippocratic oath.

This study did uncover some breaches of professionalism. Of the students who participated in our survey, 60/372 (16%) discussed patients online, and 75/381 (20%) posted pictures of a patient's excised organ or radiological image. Their intention was likely to share clinical experiences and demystify rare medical conditions, possibly unaware that they may be violating privacy regulations. Even if the information is deidentified using the Health Insurance Portability and Accountability Act's "safe harbor" technique, it may not be anonymous [18]. If the clinical scenario is unique enough, the patient might be recognized or even appear in the local news [19]. Furthermore, patients or their families may find the case description or the public's online comments hurtful or offensive. In response to several incidents, the Saudi Ministry of Health developed guidelines requiring physicians to obtain the patient's consent before sharing their images or health information online or submitting it to a journal [20,21]. Any breach of these guidelines carries a hefty penalty.

When a personal profile is linked to a profession or institution, it becomes part of its public image, brand, and professional identity. In our study, in the participants' SM profiles, 127/400 (31.8%) indicated that they were medical students, and 76/400 (19%) indicated the name of their university. Among them, 91/400 (22.8%) used their accounts to network with other professionals worldwide, making them representatives of their institutions and professions. Furthermore, students used YouTube (314/340, 92%) as a clinical reference more than websites with verified peer-reviewed content, such as UpToDate (235/340, 69%) and AMBOSS (301/340, 75%). Among our participants, 162/400 (41%) had both a public and private profile (one profile may have reflected a professional identity and the other a private one). Female students in our study are more likely than male students to have private profiles (96/248, 39% and 43/154, 28%, respectively; $P<.001$) and less likely to use a clear photo of themselves for their profile image (45/248, 18% and 73/154, 47%). This gender difference could stem from the conservative culture in Saudi Arabia or the universally higher vulnerability of women to online criticism and cyberbullying [22,23]. Regardless of privacy settings, medical students must be cautious when deciding what to post on their SM profiles since the content can be leaked.

Among the students, 119 (37 with a public profile) uploaded recordings of lectures or workshops without obtaining the presenter's permission. This behavior is concerning, as comments and expressions made by educators or attendees may be taken out of context by worldwide viewers. Educators often tailor teaching material to their intended audience. They also ensure the cultural appropriateness of their expressions and

XSL•FO
RenderX

comments while observing the audience's social norms. If educators are aware that their work will be shared with a wider online audience, they may decide to change their appearance, behavior, and lecture content. They may also choose to avoid comments that may cause controversy among other groups. These fears have led many UK universities to implement lecture capture policies to manage the recording and dissemination of lecture content [24]. The policies address concerns related to intellectual property rights, emphasizing the need for the lecturer's consent before recording and sharing materials. Furthermore, in our study, 108 students (29 with a public profile) discussed online incidents that had occurred at their institutions. These incidents may have been unintentionally misrepresented by these students. Studies have proven that eyewitness accounts are not always accurate [25]. Additionally, these incidents may have been exaggerated online for comedic or dramatic purposes or unintentionally reveal confidential patient information. Unfortunately, public criticism of these online posts will be directed at the students' profession and medical school [26,27].

Our findings contribute to the growing body of literature on medical students' SM use by highlighting specific behaviors and awareness levels in the context of the Kingdom of Saudi Arabia. While many of our results align with previous studies, notable differences were also observed. For instance, similar to a French study, most of our students used YouTube for medical education [28]. Although, our numbers (314/340, 92%) are much higher than those in France (504/762, 66%). However, only 172/400 (43%) of our students use SM for education compared to 42/63 (67%) of Canadian students in 2015 [29]. Additionally, 60% - 92% of medical schools in the United States have also experienced unprofessional online behavior by medical students [10,14]. Most students in both regions reported using restrictive privacy settings, with only 20% - 37% of US students failing to do so [30,31]. However, unlike our American counterparts, our students are less likely to use a clear profile photo, with female students being less likely than male students to do so. By contrast, an American study found that 57% of medical students had a clear profile photo, with females being more likely to display one than males [31]. While in India, 80% of students used a clear profile photo [32]. This discrepancy may reflect cultural differences in SM use. In Saudi Arabia, where our study was conducted, cultural norms and societal expectations may influence their online behaviors. These findings emphasize the importance of contextualizing SM behaviors within cultural and geographical frameworks to develop targeted interventions that address both universal and region-specific challenges.

This study's findings are consistent with the results of other studies suggesting that medical school curricula should be regularly updated and adapted to the constantly changing clinical environment, which now includes the internet [4]. Developing guidelines alone would not be sufficient, as evidenced by the fact that 51% of US medical schools that reported incidents already had policies in place addressing online content [10]. Based on our findings, medical schools must integrate e-professionalism training into their curriculum. This refers to attitudes and behaviors that reflect traditional professionalism paradigms but are manifested through digital media [33]. These guidelines should not be restricted to patient privacy but must also emphasize respect and consideration for their professors, colleagues, and medical school. We recommend that medical schools (1) develop comprehensive e-professionalism guidelines, (2) implement mandatory training sessions on SM use, (3) regularly update curricula to reflect the evolving digital landscape and its impact on professional practice, (4) introduce regular audits and feedback sessions where students' SM activities are reviewed and constructive feedback is provided, and (5) develop an anonymous reporting system for unprofessional behavior, ensuring students can report concerns without fear of retribution.

The limitations of our study include the use of a voluntary questionnaire that depended on self-reporting. Additionally, the generalizability of the findings may be limited due to the single institution sample and cultural context. The potential impact of self-reporting bias must be acknowledged, as participants might underreport unprofessional behavior. Moreover, this study did not account for other possible confounding variables such as the influence of peers or external SM trends. Two of the researchers are associate professors and 3 of them are students at the institution which may have influenced their study design and interpretation of results. Furthermore, we did not examine the specific content of medical students' posts. We are, therefore, unaware if shared patient information followed Health Insurance Portability and Accountability Act guidelines and if posts positively or negatively depicted their school. Future studies should include content analysis of SM posts as that could provide deeper insights into the types of information shared and help identify specific areas for intervention. This analysis involves categorizing posts into themes such as educational content, patient confidentiality breaches, and professional interactions.

In conclusion, this study reveals significant gaps in medical students' online behavior that can affect their medical schools' image, patient care, and reputation. To foster students' understanding of these issues, e-professionalism must be included in training curricula and assessments. This curriculum should include workshops, regular seminars on e-professionalism, and integration of SM conduct into existing ethics courses. Now, more than ever, medical schools should ensure that students develop a sense of belonging and pride in their institution and care about how it is represented worldwide. Information-sharing guidelines should strive to strike a balance between clinical knowledge sharing, protecting patients' privacy, and reflecting an institution's values and public image.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
This is a copy of the survey.
[PDF File, 29 KB - mededu_v11i1e57812_app1.pdf]

References

1. Aslam S. Twitter by the numbers: stats, demographics & fun facts. Omnicore. 2024. URL: https://www.omnicoreagency.com/twitter-statistics/ [accessed 2025-03-07]
2. Asiri AK, Almetrek MA, Alsamghan AS, Mustafa O, Alshehri SF. Impact of Twitter and WhatsApp on sleep quality among medical students in King Khalid University, Saudi Arabia. Sleep Hypn 2018 Jan 26;20(4):247-252. [doi: 10.5350/Sleep.Hypn.2018.20.0158] [Medline: 28090183]
3. Alsuraihi AK, Almaqati AS, Abughanim SA, Jastaniah NA. Use of social media in education among medical students in Saudi Arabia. Korean J Med Educ 2016 Dec;28(4):343-354. [doi: 10.3946/kjme.2016.40] [Medline: 27907981]
4. Guckian J, Utukuri M, Asif A, et al. Social media in undergraduate medical education: a systematic review. Med Educ 2021 Nov;55(11):1227-1241. [doi: 10.1111/medu.14567] [Medline: 33988867]
5. Avcı K, Çelikden SG, Eren S, Aydenizöz D. Assessment of medical students' attitudes on social media use in medicine: a cross-sectional study. BMC Med Educ 2015 Feb 15;15(1):18. [doi: 10.1186/s12909-015-0300-y] [Medline: 25890252]
6. Rosen AO, Holmes AL, Balluerka N, et al. Is social media a new type of social support? Social media use in Spain during the COVID-19 pandemic: a mixed methods study. Int J Environ Res Public Health 2022 Mar 26;19(7):3952. [doi: 10.3390/ijerph19073952] [Medline: 35409634]
7. Judd T, Elliott K. Selection and use of online learning resources by first-year medical students: cross-sectional study. JMIR Med Educ 2017 Oct 2;3(2):e17. [doi: 10.2196/mededu.7382] [Medline: 28970187]
8. Jaffar AA. YouTube: an emerging tool in anatomy education. Anat Sci Educ 2012;5(3):158-164. [doi: 10.1002/ase.1268] [Medline: 22383096]
9. Scott K, Morris A, Marais B. Medical student use of digital learning resources. Clin Teach 2018 Feb;15(1):29-33. [doi: 10.1111/tct.12630] [Medline: 28300343]
10. Chretien KC, Greysen SR, Chretien JP, Kind T. Online posting of unprofessional content by medical students. JAMA 2009 Sep 23;302(12):1309-1315. [doi: 10.1001/jama.2009.1387] [Medline: 19773566]
11. Gibson M. Nursing students expelled for posting photo of a placenta on Facebook. TIME Magazine. 2011. URL: https://newsfeed.time.com/2011/01/04/nursing-students-expelled-for-posting-photo-of-a-placenta-on-facebook/ [accessed 2025-03-07]
12. Greysen SR, Chretien KC, Kind T, Young A, Gross CP. Physician violations of online professionalism and disciplinary actions: a national survey of state medical boards. JAMA 2012 Mar 21;307(11):1141-1142. [doi: 10.1001/jama.2012.330] [Medline: 22436951]
13. Barlow CJ, Morrison S, Stephens HON, Jenkins E, Bailey MJ, Pilcher D. Unprofessional behaviour on social media by medical students. Med J Aust 2015 Dec 14;203(11):439. [doi: 10.5694/mja15.00272] [Medline: 26654611]
14. Greysen SR, Kind T, Chretien KC. Online professionalism and the mirror of social media. J Gen Intern Med 2010 Nov;25(11):1227-1229. [doi: 10.1007/s11606-010-1447-1] [Medline: 20632121]
15. Cardiff university apology after students called "idiots. BBC News. 2021. URL: https://www.bbc.com/news/uk-wales-55633371 [accessed 2025-03-07]
16. Joseph‐Richard P, Jessop T, Okafor G, Almpanis T, Price D. Big brother or harbinger of best practice: can lecture capture actually improve teaching? British Educational Res J 2018 Jun;44(3):377-392 [FREE Full text] [doi: 10.1002/berj.3336]
17. MacKay JRD. Show and 'tool': how lecture recording transforms staff and student perspectives on lectures in higher education. Comput Educ 2019 Oct;140:103593. [doi: 10.1016/j.compedu.2019.05.019]
18. Guidance regarding methods for de-identification of protected health information in accordance with the health insurance portability and accountability act (HIPAA) privacy rule. US Department of Health and Human Services. 2022. URL: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html [accessed 2025-03-07]
19. Child swallows spongebob squarepants. CBS News. 2015. URL: https://www.cbsnews.com/news/child-swallows-spongebob-squarepants/ [accessed 2025-03-07]
20. Health ministry warns against negative practices in the media and social media. Ministry of Health Saudi Arabia. 2017. URL: https://www.moh.gov.sa/Ministry/MediaCenter/News/Pages/News-2017-07-09-001.aspx [accessed 2025-03-07]
21. Saudi guidelines for informed consent. Ministry of Health Saudi Arabia. 2019. URL: https://www.moh.gov.sa/en/Ministry/MediaCenter/Publications/Pages/Saudi-Guidelines-for-Informed-Consent.pdf [accessed 2025-03-07]
22. Li Q. Cyberbullying in schools: a research of gender differences. Sch Psychol Int 2006;27(2):157-170. [doi: 10.1177/0143034306064547]
23. Wang J, Iannotti RJ, Nansel TR. School bullying among adolescents in the United States: physical, verbal, relational, and cyber. J Adolesc Health 2009 Oct;45(4):368-375. [doi: 10.1016/j.jadohealth.2009.03.021] [Medline: 19766941]
24. Ibrahim Y, Howarth A, Stone I. Lecture capture policies: a survey of British universities. Postdigit Sci Educ 2021 Jan;3(1):144-161. [doi: 10.1007/s42438-020-00102-x]

XSL•FO
RenderX

25. Albright TD. Why eyewitnesses fail. Proc Natl Acad Sci U S A 2017 Jul 25;114(30):7758-7764. [doi: 10.1073/pnas.1706891114] [Medline: 28739937]
26. Reimann N. Report: 17 florida medical residents have coronavirus after throwing house party. Forbes. 2020 Jul 27. URL: https://www.forbes.com/sites/nicholasreimann/2020/07/27/report-17-florida-medical-residents-have-coronavirus-after-throwing-house-party/ [accessed 2025-03-07]
27. Frehse R, Compinoti MS. Ohio plastic surgeon who livestreamed patient operations on TikTok has state medical license revoked permanently. CNN Ohio. 2023 Jul 14. URL: https://edition.cnn.com/2023/07/13/us/ohio-doctor-tiktok-license-revoked/index.html [accessed 2025-03-07]
28. Clavier T, Chevalier E, Demailly Z, Veber B, Messaadi IA, Popoff B. Social media usage for medical education and smartphone addiction among medical students: national web-based survey. JMIR Med Educ 2024 Oct 22;10:e55149. [doi: 10.2196/55149] [Medline: 39437450]
29. El Bialy S, Jalali A. Go where the students are: a comparison of the use of social networking sites between medical students and medical educators. JMIR Med Educ 2015 Sep 8;1(2):e7. [doi: 10.2196/mededu.4908] [Medline: 27731847]
30. MacDonald J, Sohn S, Ellis P. Privacy, professionalism and Facebook: a dilemma for young doctors. Med Educ 2010 Aug;44(8):805-813. [doi: 10.1111/j.1365-2923.2010.03720.x] [Medline: 20633220]
31. Walton JM, White J, Ross S. What's on YOUR Facebook profile? Evaluation of an educational intervention to promote appropriate use of privacy settings by medical students on social networking sites. Med Educ Online 2015;20(1):26198434. [doi: 10.3402/meo.v20.28708] [Medline: 26198434]
32. Gupta S, Singh S, Dhaliwal U. Visible Facebook profiles and e-professionalism in undergraduate medical students in India. J Educ Eval Health Prof 2015;12:50. [doi: 10.3352/jeehp.2015.12.50] [Medline: 26582630]
33. Cain J, Romanelli F. E-professionalism: a new paradigm for a digital age. Curr Pharm Teach Learn 2009 Dec;1(2):66-70. [doi: 10.1016/j.cptl.2009.10.001]

Abbreviation

KAU: King AbdulAziz University
MFMER: Mayo Foundation for Medical Education and Research
OR: odds ratio
SM: social media

Edited by B Lesselroth; submitted 29.02.24; peer-reviewed by AM Pawar, JH Ye; revised version received 20.12.24; accepted 25.02.25; published 24.03.25.

Please cite as:
Farsi S, Sabbahi A, Sait D, Kabli R, Abduljabar G
Ethical Use of Social Media and Sharing of Patient Information by Medical Students at a University Hospital in Saudi Arabia: Cross-Sectional Survey
JMIR Med Educ 2025;11:e57812
URL: https://mededu.jmir.org/2025/1/e57812
doi:10.2196/57812

© Sara Farsi, Alaa Sabbahi, Deyala Sait, Raghad Kabli, Ghaliah Abduljabar. Originally published in JMIR Medical Education (https://mededu.jmir.org), 24.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Instagram as a Tool to Improve Human Histology Learning in Medical Education: Descriptive Study

Alejandro Escamilla-Sanchez[1,2*], MD, PhD; Juan Antonio López-Villodres[1,2*], MD, PhD; Carmen Alba-Tercedor[1], PhD; María Victoria Ortega-Jiménez[1,2,3], MD, PhD; Francisca Rius-Díaz[4], MD, PhD; Raquel Sanchez-Varo[1,2,5], PhD; Diego Bermúdez[1], MD, PhD

[1]Department of Human Physiology, Human Histology, Anatomical Pathology and Physical and Sports Education, Faculty of Medicine, University of Malaga, Malaga, Spain

[2]IBIMA Bionand Platform Biomedical Research Institute, University of Malaga, Malaga, Spain

[3]Unit of Anatomical Pathology, University Hospital Virgen de la Victoria, Malaga, Spain

[4]Department of Public Health and Psychiatry, Faculty of Medicine, University of Malaga, Malaga, Spain

[5]Centre for Networked Biomedical Research in Neurodegenerative Diseases, Madrid, Spain

[*]these authors contributed equally

Corresponding Author:
Raquel Sanchez-Varo, PhD
Department of Human Physiology, Human Histology, Anatomical Pathology and Physical and Sports Education
Faculty of Medicine
University of Malaga
Bl Luis Pasteur 32, 1st floor
Malaga, 29071
Spain
Phone: 34 952131585
Email: raquelsv@uma.es

Abstract

Background: Student development is currently taking place in an environment governed by new technologies and social media. Some platforms, such as Instagram or X (previously known as “Twitter”), have been incorporated as additional tools for teaching and learning processes in higher education, especially in the framework of image-based applied disciplines, including radiology and pathology. Nevertheless, the role of social media in the teaching of core subjects such as histology has hardly been studied, and there are very few reports on this issue.

Objective: The aim of this work was to investigate the impact of implementing social media on the ability to learn human histology. For this purpose, a set of voluntary e-learning activities was shared on Instagram as a complement to traditional face-to-face teaching.

Methods: The proposal included questionnaires based on multiple-choice questions, descriptions of histological images, and schematic diagrams about the subject content. These activities were posted on an Instagram account only accessible by second-year medical students from the University of Malaga. In addition, students could share their own images taken during the laboratory practice and interact with their peers.

Results: Of the students enrolled in Human Histology 2, 85.6% (143/167) agreed to participate in the platform. Most of the students valued the initiative positively and considered it an adequate instrument to improve their final marks. Specifically, 68.5% (98/143) of the student body regarded the multiple-choice questions and image-based questions as the most useful activities. Interestingly, there were statistically significant differences between the marks on the final exam (without considering other evaluation activities) for students who participated in the activity compared with those who did not or barely participated in the activity ($P<.001$). There were no significant differences by degree of participation between the more active groups.

Conclusions: These results provide evidence that incorporating social media may be considered a useful, easy, and accessible tool to improve the learning of human histology in the context of medical degrees.

(JMIR Med Educ 2025;11:e55861) doi:10.2196/55861

XSL•FO
RenderX

KEYWORDS

medical education; medical students; histology; pathology; e-learning; computer-based; social media; Instagram; Meta; community-oriented; usability; utility; accessibility

Introduction

Social media platforms are web-based technologies particularly suited to facilitate the exchange of ideas through collaboration, interaction, and discussion. The accessibility and low cost of internet access, together with the high number of users of these platforms, make social media one of the easiest and most effective ways to disseminate information. In fact, 4.65 billion people, equivalent to 58.4% of the world's population, use social media [1]. In addition, most current medical students are far more knowledgeable and experienced with emerging technologies than preceding generations. Unlike traditional media (journals or television), social media emphasizes interactivity, motivation through social connections, and immediacy [2]. In this sense, the "social constructivism theory" states that interaction and socialization may help students learn and construct their knowledge and personal learning processes, supporting the use of social media for educational activities as a different tool for teaching and learning [3]. For all these reasons, social media platforms have progressively been incorporated into health care and medical education [4].

Technological advances have enabled rapid dissemination of medical updates through social networks such as Facebook, X (previously known as Twitter), or Instagram. Thus, students often have access to significant amounts of information, including content taught during traditional classes. Nevertheless, this information has not always been rigorously verified or is outdated, representing a formative disadvantage for medical students. Moreover, there is a lack of engagement and even dropout from classes because traditional education methodology is considered by the student body to be boring, unnecessary, or repetitive. Therefore, the faculty must adapt to meet their specific needs, changing traditional teaching styles and implementing new e-learning technologies [5].

Recently, the COVID-19 pandemic forced teaching staff to move further into a virtual education environment and highlighted the importance of communication between educators and learners through social media platforms. The rapid and efficient dissemination of information during the pandemic illustrated the significant influence of social media in the dissemination of medical literature and knowledge, not only among health care professionals but also among the student body [6,7]. For instance, Chan et al [8] demonstrated the benefits of using tools such as infographics posted on social media platforms (such as X and WeChat) to educate frontline health care workers about respiratory tract management and infection control in the setting of COVID-19. Thus, the pandemic prompted a paradigm shift in learning for students and medical residents by using different platforms (eg, YouTube, Zoom, Microsoft Teams) as e-learning tools under the new circumstances. Although face-to-face teaching is possible and desirable today, the use of social networks as educational instruments must continue with apps such as Instagram and aim to share image-based educational content to complement the classes.

Histology has long been an integral part of the medical curriculum [9] and continues to provide key information about biological tissues, physiology, and disease; it is therefore highly valued in clinical medicine and research. Furthermore, histopathology is a fundamental tool for diagnosis and prognosis. In addition, a thorough knowledge of histology is necessary for the surgical field and in general practice. However, histology and its nomenclature can be complex to understand for novice medical students, and consequently, it is often perceived as a secondary subject without clinical relevance [10]. From a pedagogical point of view, one of the main goals of histology courses is to ensure that students acquire the competencies necessary to understand histophysiology. For example, histology requires students to develop pattern recognition skills. Specifically, they must be able to identify what they are observing based on specific histologic features. Consequently, histology courses commonly include laboratory practices for students to train and develop these abilities. In this context, the study of histology through digital imaging might be a relevant alternative for the development of their curricula.

Instagram is a social networking service owned by Meta Platforms Inc that was launched in October 2010. Instagram allows photo and video sharing accompanied by text. The information can be shared either publicly or privately. Followers can archive shared posts, and the account's owner can track the number of people reached and give feedback to their followers. The literature shows that this social network is being used for educational purposes in medical schools, predominantly in imaging-related subjects such as radiology [11], ophthalmology [12], dermatology [13], anatomy [14], fertility [15], pathology [16], plastic surgery [17], dentistry [18], and (with very few proposals) in histology [19]. Understanding how students interact with these novel social media–based teaching environments and their approaches during e-learning processes is a matter of high relevance [20]. On the other hand, there is a lack of evidence on how the use of social networks impacts the learning and follow-up of Spanish medical students in the first years of their formation in the field of histology. In the first courses, the curricula of a Spanish medical degree include a basic thematic area with fundamental core subjects to obtain the essential knowledge for the subsequent study of pathological alterations. Among these subjects, some necessarily require the use of images, such as anatomy, cytology, histology, and microbiology. For that purpose, an educational experience was carried out using the social network Instagram to make the subject more attractive to the students of the official degree of Medicine at the University of Malaga in Spain during the 2022-2023 academic year. Our main objective was to test whether the use of Instagram might facilitate knowledge acquisition and increase engagement with histology, leading to a positive impact on students' qualifications. Additionally, we aimed to elucidate which type of visual material was more useful

XSL•FO
RenderX

for medical students. Finally, we determined students' perceptions of the integration of this tool in medical education.

Methods

Content of Histology in the Degree of Medicine at the University of Malaga

According to the syllabus for the degree in Medicine at the University of Malaga, histology is divided into 2 subjects (Human Histology 1 and 2) that are taught during the first and second years, respectively. The different didactic content is distributed sequentially, progressively increasing the theoretical difficulty (Table 1). First-year students learn general histology along with some special histology topics (eg, the immune system). The remaining systems and organs are studied during the second school year in the subject Human Histology 2. Some of the potential skills to be developed during these courses are knowledge about the architecture, morphology, and function of the different tissues or systems; recognizing the morphology and structure of tissues by microscopy and imaging techniques; and how to handle basic laboratory equipment and methodology. In addition, our curricula include the acquisition of some transversal competencies such as the capacity for analysis and synthesis, problem-solving or critical reasoning, and analysis, together with other abilities and skills (autonomous work, information management, and oral or written communication skills).

Table 1. Curricula content of the human histology subjects in the degree in Medicine at Malaga University.

Content	Issues dedicated to each topic, n
Human Histology 1	
Tissues (epithelial, muscle, osseous, connective, nervous)	17
Stem cells	1
Blood and hematopoiesis	3
Circulatory system	1
Immunity and lymphoid tissues	5
Human Histology 2	
Digestive system	6
Respiratory system	2
Urinary system	2
Genital apparatus	6
Tegumentary system	1
Endocrine system	6
Nervous system and neurosensorial organs	12

Sample Size

This study was carried out with 167 students enrolled in the subject Human Histology 2 in the degree in Medicine at the University of Malaga during the 2022-2023 academic year. The final examination was performed by most of the students (153 students), of which 143 participated until the end of the Instagram experience. Thus, only 10 (6.5%) of the 153 involved students did not follow our account.

Design of the Instagram Profile

After downloading the free app on a smartphone, a private Instagram account (username: @histologiauma) was created for the subject Human Histology 2 at the University of Malaga. The Instagram profile was linked to an institutional email address created to receive questions and comments from the student body of this subject. Students were notified of the availability of this account and were informed about the procedure to participate. For instance, they had to register by giving their real first name and last name. Once we checked they belong to the subject, the students were accepted as followers of @histologiauma.

Virtual Microscope Images

The images published in @histologiauma belong to the image bank of the Histology Unit of the Department of Human Physiology, Human Histology, Anatomical Pathology and Physical-Sports Education of the Medical School at the University of Malaga. During the COVID-19 pandemic, we introduced a highly interactive, web-based digital microscope system to view histological images during online practical lessons, either from classroom or personal computers. This virtual microscope is currently based on the digitalization of 66 slides, providing the element of real-time dynamic microscopy and offering students a truly innovative experience at exceptionally high resolution. Interestingly, this virtual microscope offers the possibility to capture specific tissue areas and use these pictures to formulate specific questions.

Account Content Feed

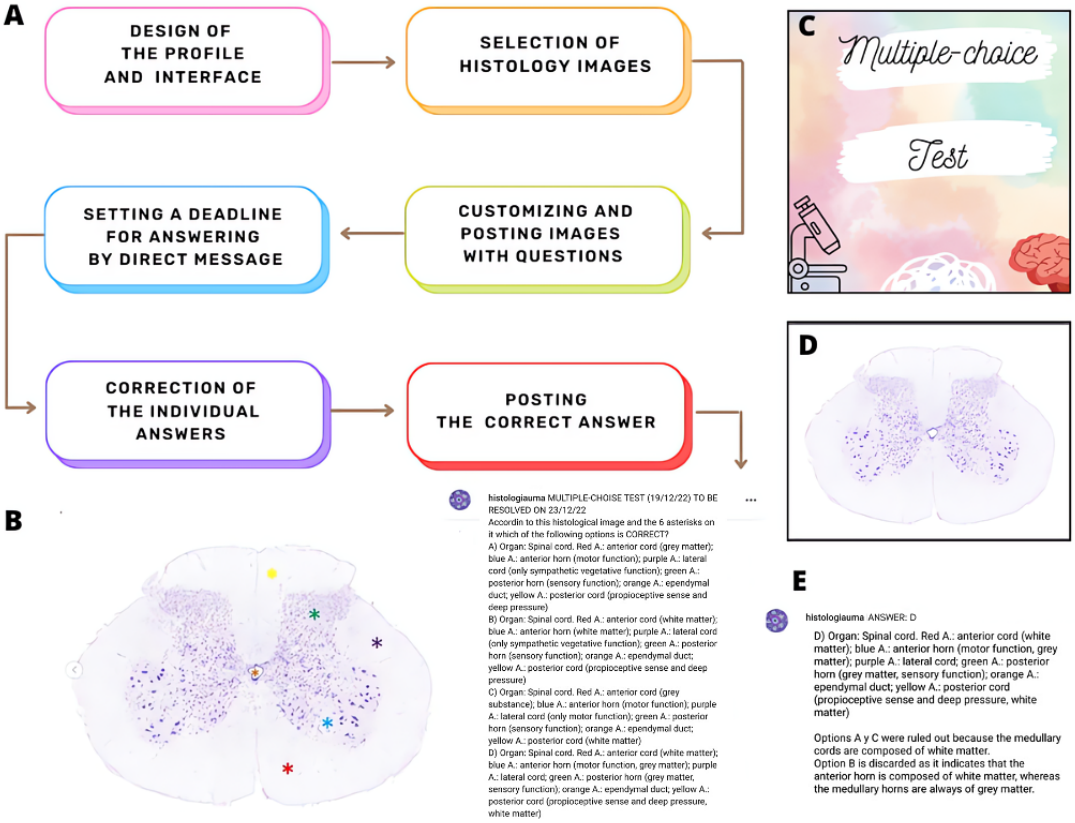
Two software applications were used to design the images: Canva and Microsoft Office PowerPoint. The free basic mode of Canva offers access to thousands of templates and 1 million free photos. Both applications allow drag-and-drop operations familiar to both average users and design professionals and

XSL•FO
RenderX

feature templates, photo filters, images, icons, and shapes useful for customizing histological images (eg, including different shapes to highlight structures or cells within a tissue, adding numbers or letters).

During a 3-month trial period, 35 posts were published, of which 5 were announcements about the account rules. The general process for uploading new content to @histologiauma account is summarized in Figure 1.

Figure 1. (A) Workflow for uploading a post in @histologiauma account, (B) screenshot of a post from @histologiauma account, (C) multiple-choice test interface, (D) image of a spinal cord transversal section with Klüver-Barrera staining, and (E) a student's direct message including a reasoned correct answer.



Ethical Considerations

The repository of digital images is composed of scanned slides with anonymized tissue remnants from Virgen de la Victoria University Hospital, whose patients provided signed informed consent for educational purposes.

The account @histologiauma was created as a private profile to be exclusively accessed by those second-year students of the degree in Medicine at the University of Malaga who voluntarily requested to participate. Images displaying captures from @histologiauma have been edited to make students' profiles unidentifiable. Moreover, all the surveys were anonymously filled out by students.

The manuscript is a retrospective case report that does not require ethics committee approval at our institution since no demographic nor clinical data from patients were used.

Type of Questions Posted on @histologiauma

The following sections were included in the Instagram platform for human histology education.

Image-Based Multiple-Choice Questions

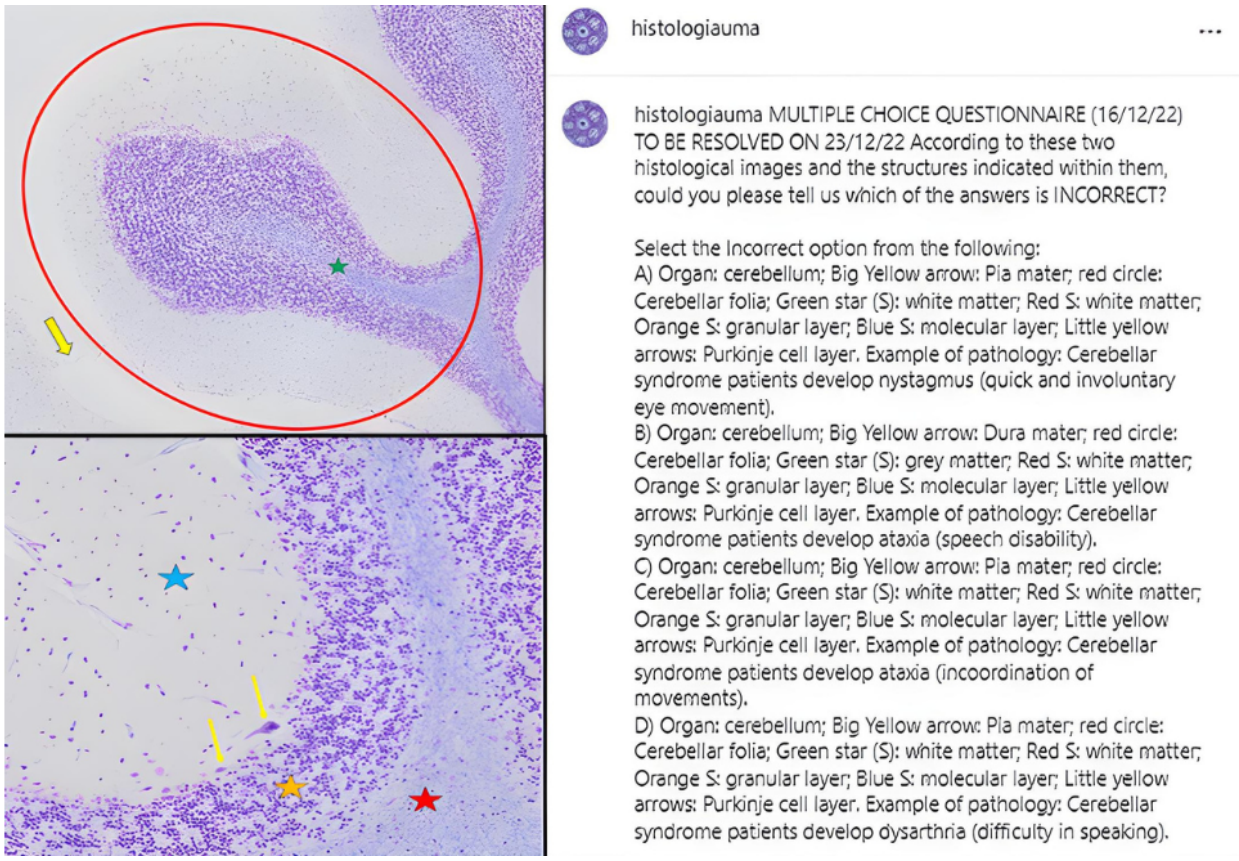
There were 13 posts with image-based multiple-choice questions (Figure 2). Histological images of several organs studied during the academic year were posted. Different structures or cells were highlighted with arrows or other shapes (eg, stars, circles, asterisks). Multiple-choice questions with a single correct answer were posted, and 4 options were marked with labels A, B, C, and D in each question. For example, Figure 2 shows a post that asked students to select the incorrect option from the following answer choices: "A) Organ: cerebellum; Large yellow arrow: Pia mater; Red circle: Cerebellar folia; Green star: white matter; Red star: white matter; Orange star: granular layer; Blue star: molecular layer; Small yellow arrows: Purkinje cell layer. Example of pathology: Cerebellar syndrome, nystagmus as a quick and involuntary eye movement is included. B) Organ: cerebellum; Large yellow arrow: Dura mater; red circle: Cerebellar folia; Green star: gray matter; Red star: white matter; Orange star: granular layer; Blue star: molecular layer; Small yellow arrows: Purkinje cell layer. Example of pathology: Cerebellar syndrome, ataxia as a problem to speak is included. C) Organ: cerebellum; Large yellow arrow: Pia mater; red circle: Cerebellar folia; Green star: white matter; Red star: white matter; Orange star: granular layer; Blue Star: molecular layer; Small yellow arrows: Purkinje cell layer. Example of pathology: Cerebellar syndrome, ataxia or incoordination of movements is included. 4) Organ: cerebellum; Large yellow arrow: Pia mater; red circle: Cerebellar folia; Green star: white matter; Red star: white matter; Orange Star: granular layer; Blue star:

XSL•FO
RenderX

molecular layer; Small yellow arrows: Purkinje cell layer. Example of pathology: Cerebellar syndrome patients develop dysarthria (difficulty in speaking).” The correct answer is B.

No negative scores were given in the case of wrong answers. All questions had a single correct answer. Each student provided their answer, including a brief justification in the form of a private message on Instagram. They were then notified about their success or encouraged to try again in case of failure. The answers were made public 5 days later in a comment, accompanied by a summary of the most common mistakes.

Figure 2. Screenshot of an image-based multiple-choice question about the cerebellum from the account @histologiauma.

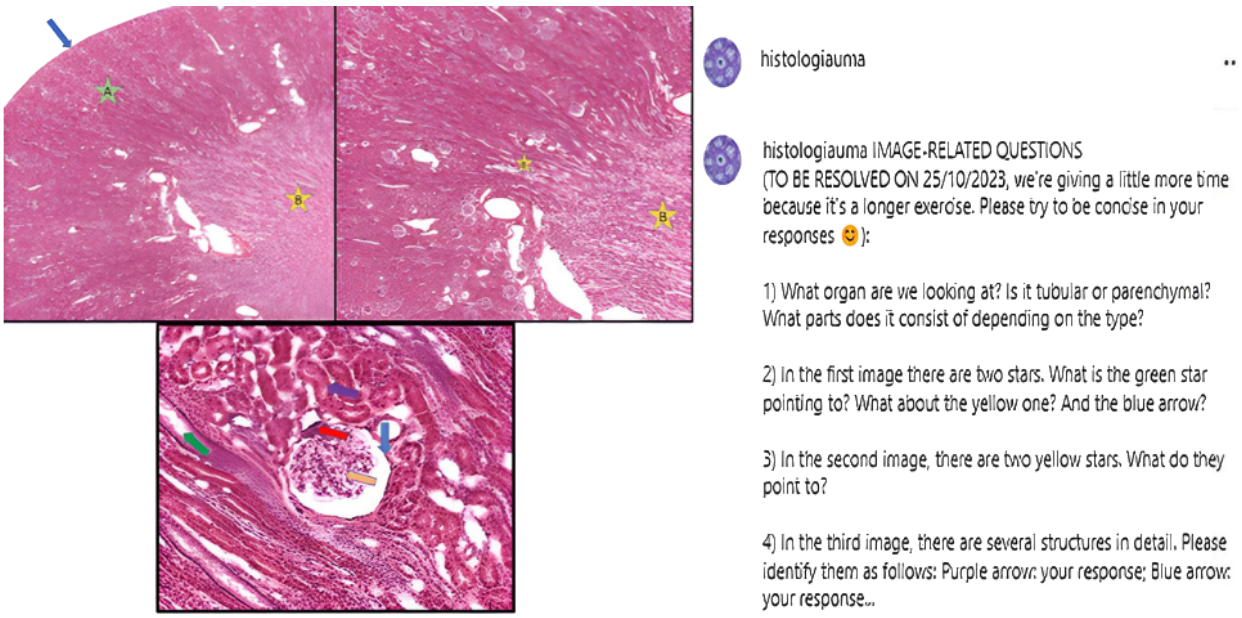


Descriptions or Questions Associated With Histological Images

This section, consisting of 10 posts, showed histological images pointing out different structures and components to be identified by the students. For example, Figure 3 shows a post that included the following questions: “1) Identify the organ shown in the image. Is it a tubular or a parenchymatous organ? 2) What is the green star (A) pointing at? And the yellow star (B)? And the blue arrow?” Occasionally, comparisons between pathological and healthy tissues were posted, along with an introduction to clinical medicine. This section was conceived in accordance with the curricular competency entitled “From Histology to Medicine,” which aims to highlight the clinical aspects of human histology. The correct answer was published 5 days later as a comment on the post, and feedback was given to the students, as explained for the multiple-choice questions.

XSL•FO
RenderX

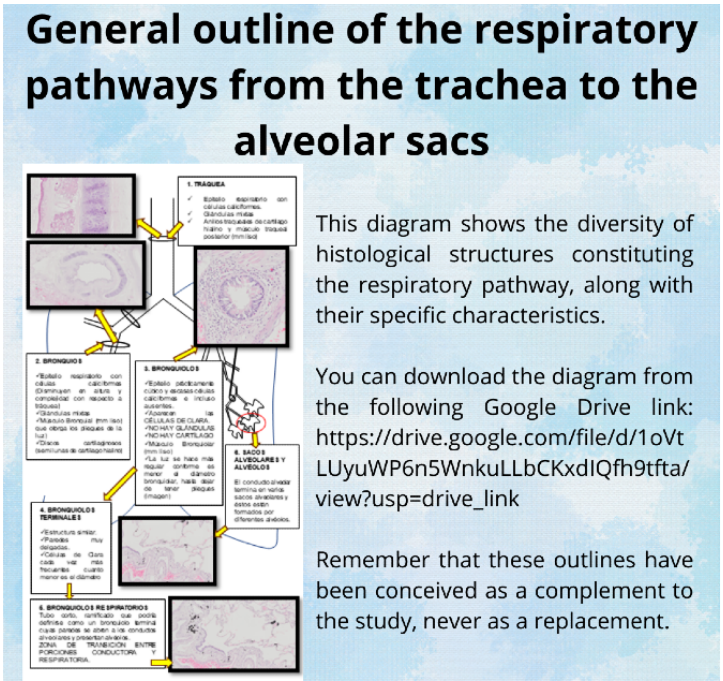
Figure 3. Screenshot of a histological image section (kidney) with the accompanying questions.



Didactic Schemes

This section included 7 posts based on student requests for visual or explanatory diagrams of the content they found most difficult. Teachers then prepared specific outlines based on these requests, avoiding the inclusion of new content. An example is shown in Figure 4. Diagrams were created using free-design and educational software, such as PowerPoint or Canva and stored in a shared Google Drive folder. The link to access the content was posted on the Instagram account and made available for 1 week. The content of these diagrams was derived from the theoretical material already provided to the students, as they were conceived as a complementary tool to the study. Students were also encouraged to make their own schemes to learn how to summarize concepts.

Figure 4. Screenshot of a scheme requested by students about the respiratory tract.

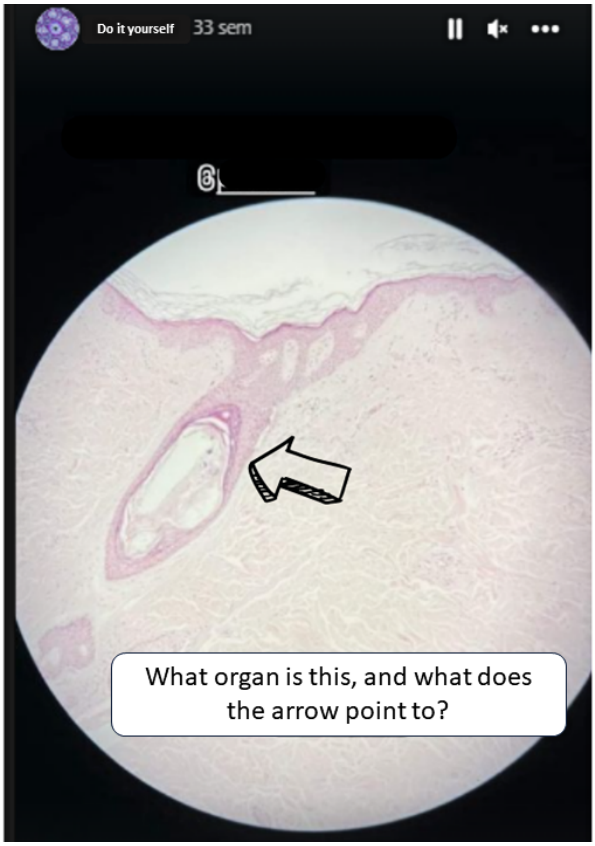


"Do It Yourself" Section

During the practical classes, students were encouraged to take images of histological slides through the eyepiece of the microscopes with their own smartphones. Later, they posted the images for 24 hours in the form of a story on @histologiauma accompanied by a specific question to be solved by their classmates. In total, 25 images were shared as stories, and an example is shown in Figure 5.

XSL•FO
RenderX

Figure 5. Screenshot of a "Do it yourself" section depicting a histological section of skin that was taken and elaborated on by one of the students.



Teaching Commitment

The teaching staff played a crucial role in providing feedback to the students, notifying them about their successes or mistakes through private messages and designing diagrams. Overall, 1 hour of work was required daily during the experience.

Rating

The activity was conceived as a voluntary pilot study. Students who actively participated could add a maximum of up to 0.50 points to their final mark, regardless of whether their interventions were successful. Thus, students earned points in proportion to their level of participation. A score of 10 was assigned to students who answered 100% of the questions. The student body was organized in 4 groups according to the rewards received on @histologiauma (group 1=0-4.4 points; group 2=4.5-6.5 points; group 3=6.5-8.5 points; group 4=8.5-10 points). Therefore, these points served as an indicator of participation.

Evaluation of the Activity as an Educational Innovation

To explore the influence of this innovative learning tool, data on student engagement and perceptions were collected during the lectures and at the end of the course through a final evaluation. The results were gathered throughout the academic semester with 3 anonymous surveys using a 5-point Likert scale. Each topic in the surveys covered a gradient of agreement with the statement presented (1=strongly disagree, 5=strongly agree). All questions were designed in Spanish by members of the teaching staff (AES, RSV, and DB) and later translated into English for publication.

At the beginning of this educational experience, the opinions of students about the inclusion of new technologies and the implementation of social media in our medical school were assessed through a first survey (called the pre-experience survey; Textbox 1).

This first survey also included specific questions regarding students' perceptions of the use of Instagram in the histology course (Textbox 2).

The second survey (middle experience; Textbox 3) was conducted 2 months after the start of the project. This questionnaire focused on the general operation of the account and their early perceptions of the experience. The third and final survey contained the same questions (Textbox 3) and was carried out during the last week of theory classes.

XSL•FO
RenderX

Textbox 1. General questions in the initial survey (pre-experience).

1. The university's educational systems are up to date and adapted to the times.
2. Most teachers use social networks as an educational resource.
3. Currently, the use of alternative educational tools is essential.
4. Bringing the basic subjects of a medical degree closer to the reality of professional practice is essential.
5. In general, teachers are concerned about updating educational tools.

Textbox 2. Specific questions for the initial survey (pre-experience).

1. Instagram facilitates access to educational content on histology.
2. Accessing Instagram allows me to consult the content anywhere at any time (eg, bus, train).
3. Test questions are a useful tool to review theoretical or practical content.
4. Downloading the subject outlines helps me to complete histology concepts.
5. I expect to improve my academic grades thanks to this academic experience.

Textbox 3. Specific questions for the 2nd (mid-experience) and 3rd (end of the experience) surveys.

1. I followed the @histologiauma updates daily.
2. I answered test questions.
3. I answered questions shared in the @histologiauma stories.
4. I shared some photographs in the “Do it yourself” section.
5. I answered the image-associated questions.
6. I used @histologiauma when traveling by public transport.
7. I used @histologiauma during class exchanges.
8. The test questions were adequate.
9. The schemes were useful.
10. We received highly personalized attention.
11. Circle the most useful sections of @histologiauma: “Do it yourself,” image-associated questions,” “multiple-choice questions,” “schemes.”
12. I will use @histologiuma to prepare for my final exams.

Statistical Analysis

Raw data from the survey responses were collected and analyzed using SPSS v.24 (IBM Corp). Descriptive statistics were used to characterize the data, with a frequency study carried out for each of the variables evaluated in the different surveys. Homoscedasticity (equality of variances) and normal distribution of the data were checked. Data corresponding to students’ marks (score range 0-10) are expressed as mean (SD). Mann-Whitney tests were conducted to compare overall grades from different cohorts (groups from different academic years or current cohort with the extra points versus the cohort without the extra points). The comparisons between the marks of the different groups (previously categorized into groups 1-4 according to the level of participation) and the degree of acquired knowledge evidenced by the final exam (global marks) were conducted using 1-way ANOVA followed by a post hoc Bonferroni test. Pearson correlation coefficients were calculated using the individual scores according to the students’ participation in @histologiauma and their final grades. An $r>1$ indicated a positive linear correlation between the 2 variables. The significance was set at 95% of confidence.

Results

Participant Demographics

From the very first post, 73 of 167 students enrolled in Human Histology 2 followed the account. At the end of the learning experience, there were 143 followers (143/167, 85.6%), of which 106 students had actively participated during the entire period. Analytics from these 143 students showed that 76.2% (109/143) were women and 23.8% (34/143) were men. Most (133/143, 93%) of them were 18 years to 22 years old and in their second year of the medical degree (139/143, 97.2%). There were only 4 repeaters of this subject, who were concomitantly in the third year of the degree. Full-time students represented 97.9% (140/143) of the respondents. The full demographic profile of the students is shown in Table 2.

XSL•FO
RenderX

Table 2. Student demographic data (N=143).

Characteristics	Results, n (%)
Gender	
Female	109 (76.2)
Male	34 (23.8)
Age (years)	
18-22	133 (93)
23-26	7 (4.9)
27-30	2 (1.4)
31-40	1 (0.7)
Academic year	
2	139 (97.2)
3	4 (2.8)
Enrollment	
Complete	140 (97.9)
Partial	3 (2.1)

Pre-Experience Test About Students' Perceptions of the Use of Social Media and New Technologies in Medical School Curricula

Of the participants, 83.9% (120/143) considered that the current educational system requires a significant update. Thus, 97.9% (139/143) of them strongly believed that the use of social networks should be significantly improved. Of the students, 99% (141/143) considered that using alternative educational tools is relevant, and 77.1% (110/143) of the students agreed that the use of social media such as Instagram facilitates access to the didactic content. Thereby, 95% (133/143) of the respondents found the subject more accessible thanks to @histologiauma, and 99.5% (142/143) believed they might improve their academic marks thanks to this experience.

Survey About the Students' Perceptions During the Experience

Multiple-choice questions (65/143, 45.7%) and image-based questions (33/143, 22.8%) were the students' favorite sections, with 96.5% (138/143) and 99.3% (142/143) of the students, respectively, considering them highly useful for learning the subject. In contrast, the schemes and "Do it yourself" sections were the favorite sections for 12.6% (18/143) and 0.8% (1/143), respectively, of the students. The remaining 18.1% (26/143) reported a stated preference for combinations of the different sections (multiple-choice questions + image-based questions: 17/143, 11.8%; multiple-choice questions + schemes: 5/143, 3.9%; image-based questions + schemes: 3/143, 2.4%). No other activities were included in the experience. Additionally, 96.7% (138/143) of the students felt well-supported and guided by the teaching staff throughout the experience. In addition, students showed no preference between public transport or the exchange of classes for visualizing the didactic content available on Instagram.

Impact of the Experience on Final Marks

The average grade obtained by the students from the Histology course during the academic year prior to the implementation of the experience (2021-2022: n=158 students) was 6.49 (SD 1.87) out of 10, whereas marks from the 2022-2023 cohort were significantly higher (mean 7.13, SD 1.68; $P<.002$; n=153), regardless of the extra points for participating in the experience. Once the earned points were included, the final outcome was not significantly different (2022/2023 without extra points: mean 7.29, SD 1.70; 2022/2023 with extra points: mean 7.13, SD 1.68; $P=.03$). Furthermore, the mean final grade from the 4 previous academic courses showed homogeneity in terms of having lower results (mean 6.12, SD 0.27; n=628 students) in comparison to our cohort. Overall, our data support that the use of social media produced a positive impact on students' performance, even without considering the points for participating in @histologiauma.

Interestingly, a positive linear correlation between individual participation scores and final marks (not including the extra reward points) was found ($r=0.439$, $P<.001$). Moreover, the ANOVA showed significant differences between students' marks according to their degree of participation ($P<.001$; Table 3). There was a trend of higher ratings according to the level of participation. The Bonferroni test showed that group 1 (the least engaged group with 0-4.4 points) achieved significantly lower global mean scores than the other 3 groups (all $P<.001$). Finally, there were no significant differences among groups 2 to 4 (all $P=.99$).

XSL•FO
RenderX

Table 3. Students' global marks (without the extra points) and rating obtained according to the degree of participation in @histologiauma.

Group	Global score, mean (SD)
Group 1 (0-4.4 points)	5.43 (2.65)
Group 2 (4.5-6.5 points)	7.44 (0.93)
Group 3 (6.5-8.5 points)	8.04 (0.51)
Group 4 (8.5-10 points)	8.27 (1.32)

Discussion

Background

Histology is one of the first morphological disciplines faced by medical students. Since it is necessary to integrate basic knowledge from other fields (eg, anatomy, cytology, biology, biochemistry) with spatial awareness, histology is perceived as a difficult subject by most learners. Moreover, students consider histology as irrelevant on their board examination (ie, "Spanish Specialized Health Training examination") and even for their future clinical practice [21,22]. Most current medical students use social networks daily and demand considerable effort from educators to make the subjects more attractive and dynamic. Creating a social media account is a free educational option that enables access to information and allows users to easily connect with others [23]. Thus, in this work, we analyzed the impact of using a specific Instagram account (@histologiauma) as a teaching resource during a histology course (2022-2023).

Principal Findings and Implications

Overall, our data demonstrate that medical students who followed and interacted with @histologiauma improved their exam scores compared with those who did not. In fact, a complete lack or a low level of participation generated significant differences in comparison with students who actively engaged with the activity. Most importantly, the enhancement of final grades compared with previous cohorts was not a direct consequence of the extra points awarded to the participants. Thus, improved test performance may serve as indirect and tangible evidence of better long-term knowledge acquisition [24]. These results are supported by the previous opinion of the majority of our students about the positive impact of this experience on academic results. In the first instance, the pre-experiment survey already showed that most of our students believed that social media is rarely used in educational contexts and considered that it may be relevant to include social media platforms as teaching tools, not only to increase accessibility to the content but also to improve their marks. In fact, the results demonstrated a positive disposition toward this innovative approach, since 99.5% of the participants believed they could improve their academic grades thanks to this experience even before participating in it.

Research on the strength or quality of motivation as a predictor of academic success has yielded both definitive and inconclusive findings. In this work, higher engagement and interaction with the content through the proposed interactive activities may have helped in the learning process, which was later reflected in the scores. Indeed, motivation is a determining factor not only for medical students but also for all students to develop sophisticated and successful learning strategies. A study on small group learning found that increased knowledge and understanding of subject matter increase students' motivation for studying and interest in the course content [24]. The "social constructivist theory" states that socialization can also help students during their personal learning processes [3]. In this sense, social media facilitates active interaction and collaboration by enabling instant communication and motivation [24,25].

Furthermore, our Instagram activities also served as additional virtual tests. Testing is no longer considered as only a tool for evaluation but also for learning [26,27]. Thus, using Instagram for educational purposes incorporates not only these phenomena in the process (as it could have been done through a virtual platform like Moodle) but also other factors that are particularly relevant for current young students: direct interaction with their classmates and immediacy, in addition to their own behavior and daily routine with smartphones and social media. We believe that all these factors increased the motivation and engagement of students with histology, leading to greater retention of the content that was finally reflected as higher scores.

Unfortunately, the information available on social media platforms might not be updated or subjected to peer review; thus, it may be invalid, incorrect, or even false. Conversely, @histologiauma is a platform controlled by our group of specialized teachers, prepared to guide learners toward appropriate knowledge according to the content of the subject. Therefore, the creation of a platform adapted by the teaching staff to the curricular content is ideal not only to boost interest but also to prevent students from accessing unreliable information [28].

Additionally, it is essential to comprehend the preferences of learners in order to create a quality digital learning environment [29,30]. During the experience, the image-based questions, multiple-choice questions, and histological descriptions were considered very useful by the students. Ultimately, this knowledge may help teachers to understand the strengths and weaknesses of the subject matter as well as its impact on adherence.

Comparison With the Literature

Numerous social media accounts disseminate information about many different types of pathologies to the general public. Although our work focuses on a course within the medical degree program, it is evident that Instagram serves as an optimal and cost-effective platform for capturing attention through passive learning in the field of histology and pathology [31]. In this sense, Nguyen et al [14] reported that 92.5% of students visit Instagram for educational purposes. Accordingly, 97.9%

XSL•FO
RenderX

of our respondents strongly believed that social networks should be implemented in higher education.

As far as we know, many accounts share educational content about pathology [16,32], but very few are specifically targeted at histology and assessing the impact of sharing this information on social media on medical students. Another novelty of our approach is that we used Instagram as an educational tool specifically tailored to our students, offering personalized content directly aligned with the course curriculum. Although many other studies have examined the use of social media in education, few have focused on how a targeted, image-based platform like Instagram can enhance engagement and learning outcomes in medical education, particularly in a subject highly reliant on visual materials.

For instance, Essig et al [19] from the School of Medicine at the University of North Carolina created an experience with the Instagram profile @InstaHisto in 2020, which is the most similar to our work in the existing literature. However, one of the main differences between these profiles could be summarized by the word "personalization." Our private account was created solely and exclusively for second-year students in the medicine degree program at the University of Malaga, in contrast with their public profile. Moreover, they examined the impact of the posts based on the number of views, not focusing on student interaction but rather on the general public. Instead, our work aimed to potentiate students' knowledge acquisition and to increase engagement with the subject. We also intended to understand students' perceptions about this educational tool. Similarly, the work by Essig et al [19] focused on the National Board of Medical Examiners final exams, reporting that 77% of their students found the histology content from @InstaHisto useful for passing the test. In this line, our survey data reflected a high degree of satisfaction with the utility in the educational environment of these virtual activities (96.5% and 99.3% with multiple-choice questions and image-based questions). More recently, Prabhu and Munawar [11] evaluated 49 Instagram profiles dedicated to the dissemination and teaching of radiology, concluding that it is indeed a better application of this image-based social media platform due to its easy accessibility and appeal to students. In this line, our data reflect that 95% of students believe that using Instagram would enhance their perception of the course and its appeal.

Limitations

In this work, rating grades increased after use of @histologiauma, even before adding the points awarded for students' participation in Instagram. It is noticeable that scarce involvement led to no or low improvement, and although there was no significant difference between the most active groups (G2 to G4), we did find a trend toward higher ratings according to the level of participation. In this sense, we cannot completely rule out the influence of other factors masking the impact of this experience, including other curricular or extracurricular activities performed during the school year or personal preferences regarding social media. Nevertheless, as far as we are concerned, standard students from the same academic course share identical academic schedules. All other activities performed during the histology course (such as problem-based learning or the section "From Histology to Medicine") were developed during theoretical classes or practice and were mandatory. However, we are aware that every academic group is different and repeating the experience during additional academic years would yield more reliable data. Relevant to this, the algorithm used by social media platforms like Instagram tends to favor posts from accounts with which users interact more frequently or with related content, creating information bias for customers [33]. Given this scenario, it is reasonable to interpret that students who interacted with @histologiauma above a threshold were later shown our account or similar profiles in their private feed more often than those who barely engaged or did not participate in the voluntary activity. This interaction likely results in students passively reviewing content each time they visit this platform, which finally positively impacts their acquisition and assimilation of knowledge and therefore their final results. However, this may also minimize the differences between the most active groups.

Another specific limitation of this work is that participating in these activities was not mandatory, which may have led to potential selection bias. We cannot rule out that students following the @histologiauma account were more eager to participate in additional activities than the general medical students. However, the high participation rate (85.6% of students enrolled) notably reduced the impact of this possibility. Even so, eliminating the optional nature of this activity would have yielded clearer data. In addition, it is possible that some of our students do not have an Instagram account because they do not find social networks attractive. Nevertheless, this was rarely the case, since we detected very few exceptions of students who performed the final exam without participating in Instagram (10 of 153 students). For future editions, we intend to propose @histologiauma as an educational instrument in a public mode, encouraging users to create their own hashtags and check the transcendence of the posts.

On the other hand, the use of social media platforms during the educational process of histology should be recommended only as a complement for regular teaching. Evidence that social media is not a panacea was provided in a separate analysis of 131 students who were using the microblog X during class with the aim of fostering student-faculty interaction on two campuses. Although it facilitated discussion, 71% of students found it distracting [34]. For this reason, it is important to find a balance between the usual lecture-based methodology and the inclusion of social media in higher education, not only to meet the curricular needs of students but also to ensure their engagement with their studies.

For future research, the sample size may be broadened to increase the validity and reliability of our findings and include cohorts from other courses or health science degrees such as podiatry, physiotherapy, or nursing. Moreover, specific tests about the content shown on the Instagram account could be implemented. The inclusion of a longitudinal study to track students' performance and engagement over multiple semesters would allow better understanding of the long-term impact of Instagram-based learning. Although, in general, we detected typical mistakes of pattern recognition of histological structures, a range of accuracy-based rewards could be incorporated into

XSL•FO
RenderX

the activity to avoid participation without true commitment. Finally, future experiences could explore the impact of different types of Instagram content (eg, video, live question-and-answer sessions, different quizzes).

Conclusions

Medical students consider there is inadequate use of social networks for teaching purposes, probably due to a lack of updated methodological approaches in the context of university subjects. Compared with the conventional educational system, social media platforms have a considerable impact on both teachers and students as they offer the possibility to easily connect and collaborate. In fact, one of the main objectives of medical education is to capitalize on the engaging nature of social media tools as part of an overall strategy to use a learner-centered approach. In addition, to increase student engagement during the first year of the degree in Medicine, it is desirable to use attractive didactic methods for learning histology. In this regard, the visual nature of histology is particularly appropriate for the introduction of new image-based tools. Thus, the aim of this study was to investigate an innovative online educational approach for histology based on an Instagram account specifically designed for medical students. In this work, we showed that the use of Instagram has great potential to improve not only the knowledge but also the scores of students of human histology. Our results provide evidence that this teaching strategy boosts students' learning motivation. In the near future, the classical practical lessons based on the physical microscope might not be enough to meet the needs of medical students. Therefore, Instagram may be considered as a relevant tool for current students to achieve their curricular objectives in a more dynamic, friendly, and enjoyable way under the supervision of the faculty.

Acknowledgments

Article processing charges were funded by Plan Propio de IBIMA-Plataforma Bionand 2024 and Project Oficina de Transferencia de Resultados de Investigacion (Office for Transference of Research Results) of the University of Malaga (reference: 806/86.6531 to AES). We thank the faculty of the Histology Unit from the University of Malaga for the histological images.

Conflicts of Interest

None declared.

References

1. Kemp S. Digital 2020: Global Overview Report. DATAREPORTAL. 2022 Jan 26. URL: https://datareportal.com/reports/digital-2022-global-overview-report [accessed 2025-01-20]
2. Wutoh R, Boren SA, Balas EA. eLearning: a review of Internet-based continuing medical education. J Contin Educ Health Prof 2004;24(1):20-30. [doi: 10.1002/chp.1340240105] [Medline: 15069909]
3. Kalasi R. The impact of social networking on new age teaching and learning: an overview. Journal of Education & Social Policy 2014;1(1):23-28 [FREE Full text]
4. Guckian J, Utukuri M, Asif A, Burton O, Adeyoju J, Oumeziane A, et al. Social media in undergraduate medical education: a systematic review. Med Educ 2021 Nov;55(11):1227-1241 [FREE Full text] [doi: 10.1111/medu.14567] [Medline: 33988867]
5. Kennedy G, Gray K, Tse J. 'Net Generation' medical students: technological experiences of pre-clinical and clinical students. Med Teach 2008 Feb;30(1):10-16 [FREE Full text] [doi: 10.1080/01421590701798737] [Medline: 18278643]
6. Katz M, Nandi N. Social media and medical education in the context of the COVID-19 pandemic: scoping review. JMIR Med Educ 2021 Apr 12;7(2):e25892 [FREE Full text] [doi: 10.2196/25892] [Medline: 33755578]
7. Villagaray-Pacheco N, Villacorta-Landeo P, Tejada-Llacsa PJ. [Social media and medical education in COVID-19 pandemic]. Rev Med Chil 2020 Aug;148(8):1220-1221 [FREE Full text] [doi: 10.4067/S0034-98872020000801220] [Medline: 33399791]
8. Chan AKM, Nickson CP, Rudolph JW, Lee A, Joynt GM. Social media for rapid knowledge dissemination: early experience from the COVID-19 pandemic. Anaesthesia 2020 Dec;75(12):1579-1582 [FREE Full text] [doi: 10.1111/anae.15057] [Medline: 32227594]
9. Chapman JA, Lee LMJ, Swailes NT. From scope to screen: the evolution of histology education. Adv Exp Med Biol 2020;1260:75-107. [doi: 10.1007/978-3-030-47483-6_5] [Medline: 33211308]
10. Tauber Z, Lacey H, Lichnovska R, Erdosova B, Zizka R, Sedy J, et al. Students´ preparedness, learning habits and the greatest difficulties in studying histology in the digital era: A comparison between students of general and dental schools. Eur J Dent Educ 2021 May 19;25(2):371-376. [doi: 10.1111/eje.12613] [Medline: 33012128]
11. Prabhu V, Munawar K. Radiology on Instagram: analysis of public accounts and identified areas for content creation. Acad Radiol 2022 Jan;29(1):77-83. [doi: 10.1016/j.acra.2020.08.024] [Medline: 32980242]
12. Huang AS, Abdullah AAN, Chen K, Zhu D. Ophthalmology and social media: an in-depth investigation of ophthalmologic content on Instagram. Clin Ophthalmol 2022;16:685-694 [FREE Full text] [doi: 10.2147/OPTH.S353417] [Medline: 35300033]

XSL•FO
RenderX

13. Chen JY, Gardner JM, Chen SC, McMichael JR. Instagram for dermatology education. J Am Acad Dermatol 2020 Oct;83(4):1175-1176. [doi: 10.1016/j.jaad.2020.02.001] [Medline: 32035941]
14. Nguyen VH, Lyden ER, Yoachim SD. Using Instagram as a tool to enhance anatomy learning at two US dental schools. J Dent Educ 2021 Sep 29;85(9):1525-1535 [FREE Full text] [doi: 10.1002/jdd.12631] [Medline: 33913160]
15. Peyser A, Goldstein L, Mullin C, Goldman RH. Fertility education: what's trending on Instagram. Fertil Res Pract 2021 Jan 18;7(1):3 [FREE Full text] [doi: 10.1186/s40738-021-00095-6] [Medline: 33461628]
16. Cutshall H, Hattaway R, Singh NP, Rais-Bahrami S, McCleskey B. The #Path2Path virtual landscape during the COVID-19 pandemic: preparing for the 2020 pathology residency recruitment season. Acad Pathol 2021;8:23742895211002783 [FREE Full text] [doi: 10.1177/23742895211002783] [Medline: 34192133]
17. Dorfman RG, Vaca EE, Mahmood E, Fine NA, Schierle CF. Plastic surgery-related hashtag utilization on Instagram: implications for education and marketing. Aesthet Surg J 2018 Feb 15;38(3):332-338. [doi: 10.1093/asj/sjx120] [Medline: 29040378]
18. Douglas NKM, Scholz M, Myers MA, Rae SM, Elmansouri A, Hall S, et al. Reviewing the role of Instagram in education: can a photo sharing application deliver benefits to medical and dental anatomy education? Med Sci Educ 2019 Dec;29(4):1117-1128 [FREE Full text] [doi: 10.1007/s40670-019-00767-5] [Medline: 34457591]
19. Essig J, Watts M, Beck Dallaghan GL, Gilliland KO. InstaHisto: utilizing Instagram as a medium for disseminating visual educational resources. Med Sci Educ 2020 Sep 17;30(3):1035-1042 [FREE Full text] [doi: 10.1007/s40670-020-01010-2] [Medline: 34457765]
20. Kebritchi M, Lipschuetz A, Santiague L. Issues and challenges for teaching successful online courses in higher education. Journal of Educational Technology Systems 2017 Aug 08;46(1):4-29 [FREE Full text] [doi: 10.1177/0047239516661713]
21. Johnson S, Purkiss J, Holaday L, Selvig D, Hortsch M. Learning histology - dental and medical students' study strategies. Eur J Dent Educ 2015 May;19(2):65-73 [FREE Full text] [doi: 10.1111/eje.12104] [Medline: 24809952]
22. Dennis JF. The HistoHustle: supplemental histology sessions to enrich student learning and self-efficacy. Med Sci Educ 2020 Dec;30(4):1725-1726 [FREE Full text] [doi: 10.1007/s40670-020-01060-6] [Medline: 34457834]
23. Prudencio J, Wongwiwatthananukit S, Lozano A, Xu Y. Instagram as a tool to enhance pharmacy student learning of ambulatory care pharmacy. Curr Pharm Teach Learn 2021 Feb;13(2):134-138. [doi: 10.1016/j.cptl.2020.09.007] [Medline: 33454069]
24. Kusurkar RA, Ten Cate TJ, van Asperen M, Croiset G. Motivation as an independent and a dependent variable in medical education: a review of the literature. Med Teach 2011;33(5):e242-e262. [doi: 10.3109/0142159X.2011.558539] [Medline: 21517676]
25. Hennessy C, Smith C. Digital and social media in anatomy education. Adv Exp Med Biol 2020;1260:109-122. [doi: 10.1007/978-3-030-47483-6_6] [Medline: 33211309]
26. Yang BW, Razo J, Persky AM. Using testing as a learning tool. Am J Pharm Educ 2019 Nov;83(9):7324 [FREE Full text] [doi: 10.5688/ajpe7324] [Medline: 31871352]
27. Murphy DH, Little JL, Bjork EL. The value of using tests in education as tools for learning—not just for assessment. Educ Psychol Rev 2023 Sep 08;35(3):1. [doi: 10.1007/s10648-023-09808-3]
28. Popoola-Samuel HAO, Bhuchakra HP, Tango T, Pandya ND, Narayan KL. Instagram and seizure: knowledge, access, and perception of circulating information on the internet. Cureus 2023 Jul;15(7):e41664 [FREE Full text] [doi: 10.7759/cureus.41664] [Medline: 37575724]
29. Clauson KA, Singh-Franco D, Sircar-Ramsewak F, Joseph S, Sandars J. Social media use and educational preferences among first-year pharmacy students. Teach Learn Med 2013 Apr;25(2):122-128 [FREE Full text] [doi: 10.1080/10401334.2013.770742] [Medline: 23530673]
30. Timothy PG, Jeffrey B, Kaitlyn L, Margarita VD. Delivery of educational content via Instagram. Med Educ 2016 Apr 13;50(5):575-576 [FREE Full text] [doi: 10.1111/MEDU.13009] [Medline: 27072461]
31. Gomaa B, Houghton RF, Crocker N, Walsh-Buhi ER. Skin cancer narratives on Instagram: content analysis. JMIR Infodemiology 2022;2(1):e34940 [FREE Full text] [doi: 10.2196/34940] [Medline: 37113805]
32. Schukow CP, Kilpatrick SE. Highlighting bone and soft tissue pathology on Instagram. Adv Anat Pathol 2023 Mar 06:1. [doi: 10.1097/PAP.0000000000000396] [Medline: 36882880]
33. Gurler D, Buyukceran I. Assessment of the medical reliability of videos on social media: detailed analysis of the quality and usability of four social media platforms (Facebook, Instagram, Twitter, and YouTube). Healthcare (Basel) 2022 Sep 22;10(10):1 [FREE Full text] [doi: 10.3390/healthcare10101836] [Medline: 36292284]
34. Fox BI, Varadarajan R. Use of Twitter to encourage interaction in a multi-campus pharmacy management course. Am J Pharm Educ 2011 Jun 10;75(5):88 [FREE Full text] [doi: 10.5688/ajpe75588] [Medline: 21829262]

XSL•FO
RenderX

Edited by B Lesselroth; submitted 04.01.24; peer-reviewed by F Tume, H Zou, J Amaro, M Garrosa, H Alshawaf; comments to author 14.08.24; revised version received 10.09.24; accepted 17.12.24; published 19.02.25.

Please cite as:
Escamilla-Sanchez A, López-Villodres JA, Alba-Tercedor C, Ortega-Jiménez MV, Rius-Díaz F, Sanchez-Varo R, Bermúdez D
Instagram as a Tool to Improve Human Histology Learning in Medical Education: Descriptive Study
JMIR Med Educ 2025;11:e55861
URL: https://mededu.jmir.org/2025/1/e55861
doi:10.2196/55861
PMID:

©Alejandro Escamilla-Sanchez, Juan Antonio López-Villodres, Carmen Alba-Tercedor, María Victoria Ortega-Jiménez, Francisca Rius-Díaz, Raquel Sanchez-Varo, Diego Bermúdez. Originally published in JMIR Medical Education (https://mededu.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring the Role of Immersive Virtual Reality Simulation in Health Professions Education: Thematic Analysis

Jordan Talan, MHPE, MD; Molly Forster, MD; Leian Joseph, MD; Deepak Pradhan, MHPE, MD

Division of Pulmonary, Critical Care, & Sleep Medicine, Department of Medicine, NYU Grossman School of Medicine, 550 First Avenue, 15th Floor, Medical ICU, New York, NY, United States

Corresponding Author:
Jordan Talan, MHPE, MD
Division of Pulmonary, Critical Care, & Sleep Medicine, Department of Medicine, NYU Grossman School of Medicine, 550 First Avenue, 15th Floor, Medical ICU, New York, NY, United States

Abstract

Background: Although technology is rapidly advancing in immersive virtual reality (VR) simulation, there is a paucity of literature to guide its implementation into health professions education, and there are no described best practices for the development of this evolving technology.

Objective: We conducted a qualitative study using semistructured interviews with early adopters of immersive VR simulation technology to investigate use and motivations behind using this technology in educational practice, and to identify the educational needs that this technology can address.

Methods: We conducted 16 interviews with VR early adopters. Data were analyzed via directed content analysis through the lens of the Unified Theory of Acceptance and Use of Technology.

Results: The main themes that emerged included focus on cognitive skills, access to education, resource investment, and balancing immersion. These findings help to clarify the intended role of VR simulation in health professions education. Based on our data, we synthesized a set of research questions that may help define best practices for future VR development and implementation.

Conclusions: Immersive VR simulation technology primarily serves to teach cognitive skills, expand access to educational experiences, act as a collaborative repository of widely relevant and diverse simulation scenarios, and foster learning through deep immersion. By applying the Unified Theory of Acceptance and Use of Technology theoretical framework to the context of VR simulation, we not only collected validation evidence for this established theory, but also proposed several modifications to better explain use behavior in this specific setting.

(JMIR Med Educ 2025;11:e62803) doi:10.2196/62803

KEYWORDS

virtual reality; medical education; virtual reality simulation; extended reality; simulation; VR; health professions education; health education; thematic analysis; evolving technology; qualitative study; qualitative; semistructured interviews; educational experiences; theoretical framework

Introduction

Background

As technology rapidly advances in immersive virtual reality (VR) simulation, there is a growing interest among educators to develop VR simulation curricula for health professions education. However, there is a paucity of literature to guide these efforts, and there are no accepted best practices for the development or implementation of this technology. While experts anticipate the potential for VR to transform medical education [1], without a better understanding of the role VR will play in our training programs, these statements may amount to nothing more than vague future promises. Therefore, characterization of the early use of VR is imperative to clarify its evolving role and gain insights that will allow us to implement this technology to its fullest potential.

VR Simulation Technology

Immersive VR creates a simulated environment, allowing users to “step inside” a computer-generated world and engage authentically with their surroundings [1]. VR offers several potential benefits for health professions education, including facilitating distance learning and providing training that is difficult to deliver via traditional simulation [2]. In addition, VR shows comparable educational outcomes to high-fidelity mannequin simulation with more cost-effectiveness [3-7]. Many institutions are enthusiastic about VR simulation and are already piloting or studying VR curricula [1,8] However, there is still

XSL•FO
RenderX

much to learn in order to best guide the development and implementation of these curricula.

While prior research has concentrated on individual VR usage-scenarios or software evaluations [9,10], effective educational interventions require a broader understanding of the context of our learners [11]. Therefore, we must study VR user needs across a wider spectrum to guide development that aligns with the context of health professions training. By analyzing current VR educational practices, we can better identify the gaps that this technology can bridge, and move toward a consensus about how best to use VR simulation in the future. Without a better understanding of these gaps, we risk pouring resources into technology for technology's sake—a solution looking for a problem [12].

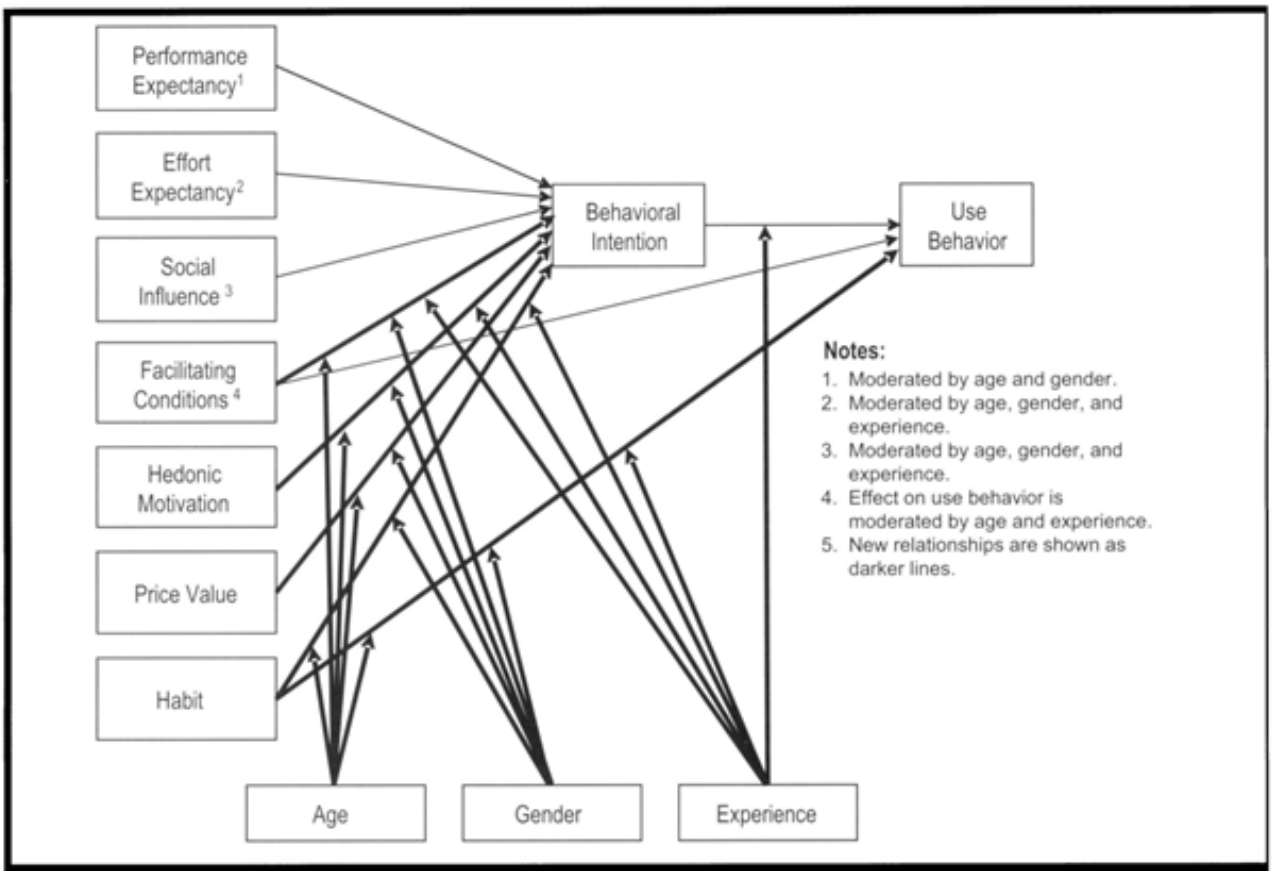
Study of Early Adopters

The technology adoption life cycle categorizes users into 5 groups based on their likelihood to adopt new technology: innovators, early adopters, early majority, late majority, and laggards [13]. Our study focuses on early adopters, as they represent educational stakeholders pioneering the implementation of VR within authentic educational environments and collaborating with VR innovators to adapt the technology to their needs. They therefore have expertise evaluating VR technology, but unlike the innovators, their experience is more practical than theoretical.

Unified Theory of Acceptance and Use of Technology

The Unified Theory of Acceptance and Use of Technology (UTAUT) explains factors that affect the adoption of new technologies and predicts future technology use [14]. UTAUT provides a robust theoretical framework for understanding the drivers incentivizing early adopters to embrace VR as an educational strategy. The original theory described 4 constructs as direct determinants of technology usage behavior (Figure 1): performance expectancy (user expectation that the technology improves performance), effort expectancy (ease associated with using the technology), social influence (user perception that others believe they should be using the technology), and facilitating conditions (organizational and technological infrastructure for technology implementation). These determinants are modified to varying degrees by user gender, age, or experience [14]. Extensively applied across multiple fields for assessing new technologies [15], the UTAUT was expanded to the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) with 3 additional constructs: hedonic motivation (pleasure derived from using the technology), price value (perceived cost of the new technology), and habit (degree of automatic use of the technology) [16].

Figure 1. The UTAUT2 from Venkatesh et al [16], used with permission. UTAUT2: Unified Theory of Acceptance and Use of Technology 2.



With validity evidence across multiple fields since 2003, UTAUT has become one of the most developed and intensive models to test new technologies [15]. In addition to being well-validated in multiple settings, UTAUT is an ideal theoretical framework to explore and better understand thoughts and behaviors associated with the use of VR simulation technology. While other theoretical frameworks have been used to approach VR simulation research, few will allow the isolation of factors associated with VR specifically rather than those that apply to simulation in general. Even fewer might facilitate the

XSL•FO
RenderX

prediction of its use in the future. For example, constructivist learning theory has been applied to VR simulation because learners can manipulate a problem and construct learning from active participation in an engaging experience [17]. Experiential learning has also been used to contextualize VR simulation because it provides a safe and forgiving training environment that facilitates learning by doing [3]. However, these theoretical frameworks serve better to characterize simulation in general rather than to focus on the specific experience provided by VR technology. The use of UTAUT2 on the other hand provides a structured framework by which we can distinguish the features of VR technology from other modes of simulation, and by which we can attempt to predict its future use.

Prior VR research using UTAUT2 focuses primarily on understanding learner experience and learner acceptance [18,19]. While these concepts are critical for the successful adoption of evolving technological innovations [20], we must progress further by investigating how VR simulation can address specific educational needs and gathering validation evidence for its most effective future role in the evolving landscape of health professions education.

Study Aims

To fill this gap in our understanding, this study interviews early adopters of VR simulation in health professions education, with the following aims: (1) characterize how early adopters are adapting VR to meet their educational needs, (2) define the educational problems or gaps that early adopters are trying to address with VR, and (3) explore factors influencing the ability of early adopters to meet their needs with VR.

Methods

Ethical Considerations

This study was approved by the New York University Langone Institutional Review Board (22 - 01346). Informed consent was obtained from all study participants, and participants had the ability to opt out or withdraw from the study at any time. Interview transcripts were deidentified for confidentiality. Participants were not compensated.

Study Design

This is a qualitative study using thematic analysis of semistructured interviews. The research methodology is directed content analysis, starting with a limited code book of 7 a priori codes defined through the lens of the UTAUT2 theoretical framework, followed by an exploratory coding phase [21,22]. The research paradigm is postpositivist. Reporting was completed following the Standards for Reporting Qualitative Research guidelines [23].

Semistructured Interview Guide

We iteratively developed a semistructured interview guide based on our research questions and grounded in the UTAUT2 theoretical framework [24,25] (Multimedia Appendix 1). The interview guide was piloted with local stakeholders to ensure capture of meaningful data within the 45-minute interview timeframe.

Recruitment and Sampling

We recruited educational stakeholders who were identified as "early adopters" of immersive VR simulation technology. Inclusion criteria were experience educating, implementing, or researching with VR. Exclusion criteria included technology developers without educational practice experience, and participants with experience limited to 360° video, augmented reality, or nonsimulation immersive learning.

The first 3 participants were recruited as a convenience sample, as they were known to our research team based on their work with the American College of Chest Physicians to develop and pilot an immersive VR simulation program teaching endotracheal intubation. These participants were recruited as an entry into the community of VR early adopters, with subsequent recruitment by snowball sampling. We sought to map the terrain of VR use-cases by recruiting for maximal diversity. We asked if participants could identify additional early adopters who had different experiences (ie, worked with a different company, in a different learner setting, at a different institution, or who had differing perspectives on VR technology). We estimated a sample size of 12 - 18 interviews. Data were iteratively analyzed for thematic saturation, and recruitment was terminated upon achieving saturation of meaning [26].

Interviews and Data Analysis

Each participant completed a 45-minute semistructured interview via Zoom videoconferencing. Interviews were audio-recorded and transcribed verbatim into a written document via Speechmatics software with manual verification. Transcripts were imported to ATLAS.ti (ATLAS.ti Scientific Software Development GmbH) web, which was used for iterative qualitative data coding and analysis. First-round coding was performed via an a priori coding template corresponding to the UTAUT2 domains. Any additional codes used process coding and descriptive coding. All codes were approved by 2 independent reviewers (JT and DP) with deliberation over any discrepancies. Second-round coding then checked all codes against the initial coding template, collapsing as necessary to capture any new domains not described by the UTAUT2 framework. Field notes and memos were maintained by both reviewers. Themes were identified and their interrelationships characterized [27]. Themes were then shared with study participants via member checking to ensure the accuracy of our analyses.

Reflexivity

JT and DP are Pulmonary/Critical Care Medicine physicians. JT has worked with technology companies and educational technologists researching immersive VR simulation, but is relatively suspicious of new technology unless it fulfills a specific need. DP is also an early adopter, who is a self-described "gamer" and owns a VR headset for recreational use. JT, DP, and MF are simulation educators at New York University. All authors kept memos to practice reflexivity throughout this study's period.

XSL•FO
RenderX

Results

Overview

We completed 16 semistructured interviews. Coding saturation occurred after 11 interviews and thematic saturation after 12 interviews. Four additional interviews were completed to ensure saturation of meaning [26]. Participant demographics are described in Table 1. Our study population included early adopters from diverse health professions whose educational interventions targeted the following groups of learners: physician trainees (premedical students, medical students, residents, and fellows), advanced practice providers (nurse practitioners and physician assistants), nurses and nursing students, respiratory therapists, pharmacists, and emergency service members (emergency medical technician students and paramedical students).

XSL•FO
RenderX

Table . Demographics of interview participants (N=16).

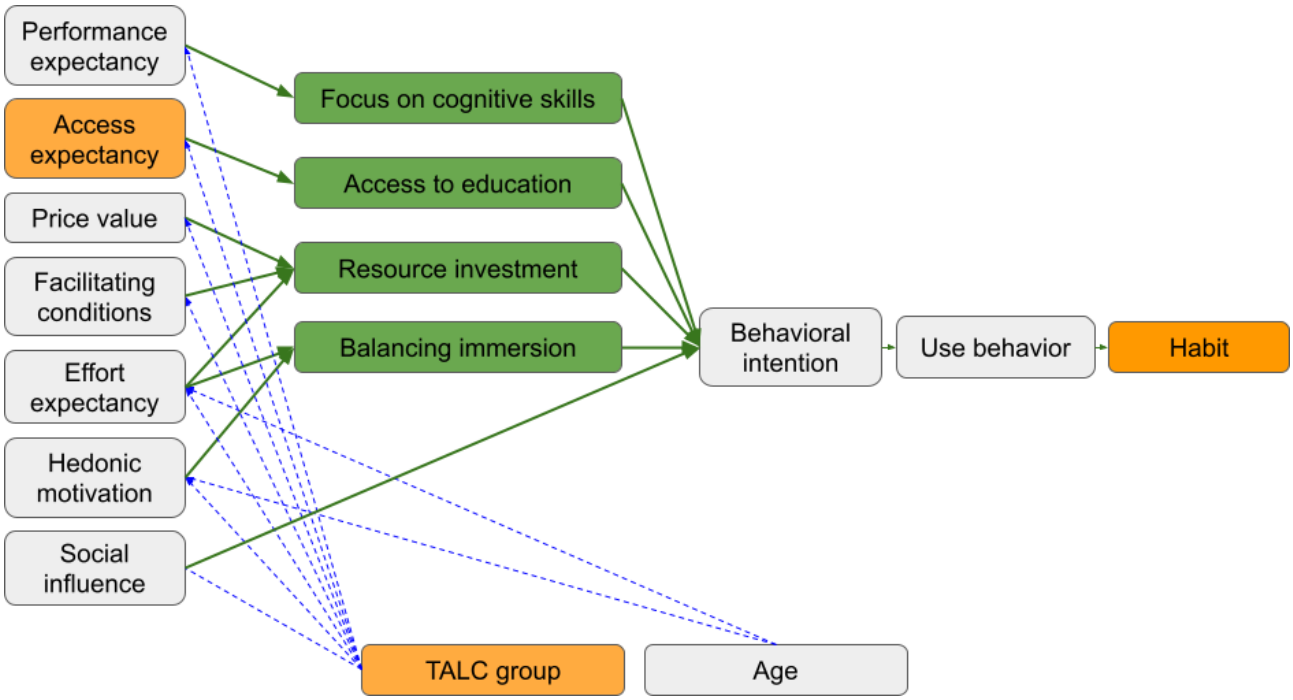
Demographics of interview participants	Values (n)
Gender	
Male	10
Female	6
Age (years)	
31 - 50	8
51 - 65	3
>65	3
Unknown	2
Technology adoption life cycle group	
Innovator	6
Early adopter	6
Early majority	3
Unknown	1
Geography	
Northeast (United States)	3
Midwest (United States)	5
South (United States)	4
West (United States)	3
Canada	1
Setting	
Urban	12
Suburban	4
Rural	0
Health profession	
Advanced practice provider (nurse practitioner or physician assistant)	1
Emergency medical service (paramedics or emergency medical technicians)	1
Health care education technologist	1
Nurse	3
Physician	8
Anesthesiology	2
Cardiology (pediatrics)	1
Emergency medicine (adult)	2
Armed forces	1
Emergency medicine (pediatrics)	1
Internal medicine	1
Pulmonary and critical care medicine	2
Respiratory therapist	1

Coding and Themes

First-round coding generated 38 unique codes: 7 from the a priori coding template corresponding to UTAUT domains, and 31 new codes via process and descriptive coding. Four themes were identified and examined for their interrelationships. The resulting synthesis and validation evidence for the UTAUT2 framework are depicted in a thematic map (Figure 2).

XSL•FO
RenderX

Figure 2. Thematic map of results. The thematic map illustrates our results and changes we have made to the theoretical framework. Themes (green boxes) are superimposed between each UTAUT construct and the resulting behavioral intention. The new addition of access expectancy is highlighted in orange. Green arrows illustrate which theme most strongly relates to which UTAUT construct. Blue arrows represent effect modifying relationships. TALC: technology adoption lifecycle; UTAUT: Unified Theory of Acceptance and Use of Technology.



Theme 1: Focus on Cognitive Skills

Study participants focused on VR simulation for the development of cognitive skills, including communication, teamwork, clinical reasoning, situational awareness, and interdisciplinary skills, occasionally referred to as "soft skills."

> *It's really just about talking to each other, right? And sharing that mental model…I think that's where you can really benefit from VR because it's not really about the tasks you're doing. It's how are you communicating…I think if you look at a lot of the sentinel events or the near misses that happen, it's based on communication.* [Participant #4]

For procedural skills, the role of VR simulation was limited to building procedural knowledge or situational awareness.

> *It does help you remember the different steps. You know, don't forget the suction at the head of the bed or, you know, [we will] have the patient vomit…and maybe they won't forget it now.* [Participant #4]

However, participants found the teaching of fine psychomotor skills, such as laryngoscopy or peripheral intravenous placement, to be limited in VR.

> *We knew that you cannot teach the fine motor skills of intubating a patient in virtual reality. It is very difficult to do with the kind of tools that are available right now. And so it was more about the thinking…I truly believe that the mental process of approaching an airway is just as important, if not more important, than the fine tuning of technical skills.* [Participant #1]

The most common barrier to teaching psychomotor skills in the VR environment was haptic technology, noted unanimously by every interview participant.

> *What it doesn't do well? One is teaching people how to do things that require them to use their hands and fine motor skills, even just like how to use tools. It's really challenging to teach somebody how to hold a laryngoscope, how to hold the endotracheal tube…I think the big limiting factor is the fact that you have to use controllers because the controllers only work a certain way. You must hold it this way. These are the few buttons that you have. You're not using your hands the way you would in real life.* [Participant #1]

In addition to limitations in simulating authentic tools, participants noted limitations in simulating the weight or feel of human anatomy necessary to learn fine psychomotor skills.

> *You put the [laryngoscope] blade in their mouth, the vocal cords show up on the screen and you just drop the tube and it just clicks right in. But right now, we don't have…that feedback where you feel your scope in your hand or you feel the weight of the jaw when you're going to lift up.* [Participant #8]

Many participants discussed the development of haptic gloves which can allow users to simulate touch and contact experiences. However, most found current solutions either cost-prohibitive or inadequate.

> *I don't see that type of fine motor feedback, you know, where I know how to put the needle into an arm for an IV, for example. That's not going to happen for - I would say that's decades away. Easily. I don't think there are good solutions right now in even the most*

XSL•FO
RenderX

> *expensive labs with experimental haptics for that.* [Participant #16]

Therefore, to teach psychomotor and procedural skills, participants turned to other forms of simulation technology such as mixed reality or mannequin simulation. Several participants offered learners a blended experience, using VR to create an immersive scenario, followed by a task trainer to simulate any necessary fine motor tasks.

Theme 2: Access to Education

The ability of VR simulation to facilitate distance learning was seen as a significant driver of use behavior for most participants.

> *We wanted to break down the barriers of requiring learners to physically come to a place to get this type of education. We want this education to be deliverable over long distances to people in other parts of the world.* [Participant #1]

In areas without the resources for a high-fidelity mannequin-based simulation laboratory, VR was seen as expanding access to high-quality simulation learning from expert educators.

> *Places that don't have simulation labs and all of those resources…available at academic medical centers, at [professional society] headquarters…But outside of large centers or hospitals that have access to a sim lab - and I think probably the majority of hospitals in the country do not - those hospitals don't have access to that type of education.* [Participant #1]

VR simulation was also used as a solution to reduce the cost, inconvenience, and sometimes danger associated with traveling to simulation centers.

> *The ability to do remote simulation at a much lower cost than requiring travel, that's a huge benefit…If you've got employees spread across the country, even across the state - and I'll use Wyoming, for example… everything's 8 hours away. It's icy half the year. So if…you've got students all over the state that are part of your paramedic program, and you have these guys driving throughout the winter to come to your simulation center. Like, what are the chances of something bad happening there? Pretty high to be honest.* [Participant #8]

Even in centers with existing high-fidelity simulation laboratories, participants found a role for VR in facilitating collaboration and standard-setting at a national or international level.

> *As people adopt these headsets…somebody from [University] could do the same ACLS training as somebody at [University] and it would be the same across institutions. And there's crosstalk - so different learning points and different perspectives and shared information and shared values in terms of education.* [Participant #6]

Finally, VR simulation also expanded access to education during the COVID-19 pandemic, responding to the need for social distancing, and significantly accelerating VR adoption.

> *If heaven forbid there was another pandemic, now we are set up that. If our students were at home on lockdown, as long as they had a headset, their learning would not be interrupted.* [Participant #13]

Overall, VR afforded a distance learning advantage, providing equitable access to high-quality simulation education to centers in diverse settings and learners in adverse scenarios.

Theme 3: Resource Investment

Implementation of a VR simulation curriculum required extensive resources, particularly upfront costs (funding and time commitments).

> *There is a capital purchase that has to be made just for the equipment itself. But…how do you develop that program in a manner that somebody is not spending tons and tons of time to bring one little educational module to fruition.* [Participant #3]

These upfront costs also related to the process of cocreation with technology companies.

> *There was generally some frustration during the build process because we're all clinicians and we're like, 'yeah, this thing needs to be this way'. And you're trying to communicate that with someone who has no medical experience and is a software programmer… We speak one language and they speak a different language, and there was some inability to communicate that effectively.* [Participant #3]

Once the programming was complete, participants also described an ongoing cost to maintain the software through updates or licensing.

> *Keeping these things alive is really…the cost to maintain software…for servers and engineers and updates and things like that. So without some sort of continued funding from somewhere, it will become a useless pile of code as soon as the next [operating system] update hits.* [Participant #9]

The investment required to develop novel VR programming was frequently more resource-intensive than anticipated, and the risk of failed investment was wasted time and money.

> *I've seen many cases of projects that are developed and they're just abandoned…I would walk into my office every morning and I had a stack of 16 boxes of headsets we didn't use.* [Participant #10]

Therefore, participants wanted more opportunities for creative collaboration and sharing of software programs. However, some felt limited by the current state of technology.

> *There's no great way to share content yet. So a lot of stuff is just getting reinvented over and over again, which is a really expensive way to do things.* [Participant #12]

Others felt restricted by the current incentives within the VR marketplace, concerned by compatibility between different software or hardware companies.

> *There needs to be an ability for me to use multiple vendors within my one headset without having to pay*

XSL•FO
RenderX

> *millions of dollars to do so…I don't know about everybody's budget. On my budget, I cannot afford to pay four different guys for completely different programs.* [Participant #8]

Generally, participants desired to use pre-existing software that was universally relevant for multiple institutions and multiple users, and compatible with a variety of hardware.

Theme 4: Balancing Immersion

The immersiveness of VR was a powerful experience associated with learner enjoyment.

> *Being in virtual reality is an immersive experience, and it's just hard to describe in words until you try it. But when people try it, it's like seeing a new color.* [Participant #1]

At its best, immersion increased learner presence, stimulated intellectual curiosity, and accelerated learning.

> *When you go in and you see an environment in 3D that looks exactly like your cardiac ICU…you immediately have a 'wow' thing. And what I love about that is immediately when I start this scenario, I never really hear like, 'Wait, what do you want me to do? Are we starting now? Is the patient supposed to have pulses?'…It's so immersive that people immediately feel like they're in a football game and it's kickoff.* [Participant #7]

Participants also valued VR immersion for minimizing distractions more than other simulation technologies.

> *The thing that's nice about virtual reality is you put the headset on and that's what you're doing, right? So you're not looking at your phone or checking your email while someone's trying to teach you.* [Participant #3]

However, immersion could also create extraneous cognitive load, detracting from learning. Participants described unnecessary environmental elements that distracted from learning objectives, along with some tasks that were frustrating to simulate in VR.

> *The picking up of items in the ICU was difficult always…With the limited controller toggles, it was not always intuitive how to pick something up. And even when they told you what to do, it still sometimes fell on the floor and stuff like that.* [Participant #2]

Sometimes immersive scenarios became more about navigating the VR environment than mastering intended learning objectives.

> *You're never going to drop some instrument on the tray 6 times…Like is the goal to learn to pick the scope up, or to [learn the procedure]?…I think a lot of people try and make the virtual world exactly like the real world…but I think you have to simplify the haptics…If it's just so frustrating because you're an intensivist and you can't pick up the needle drivers, then forget it. There shouldn't be a five minute learning curve on how to pick up needle drivers, right?* [Participant #12]

Participants found an ideal immersive balance when the virtual world accomplished the intended learning objectives, but was not overly complex to create frustration in navigating the environment. In this way, there was constructive alignment between the intended learning outcomes and the virtual learning activities.

Validation Evidence for the UTAUT2 Framework

Codes were confirmed for each previously described construct within the UTAUT2 framework [16]. Performance expectancy was the most frequently coded driver for use intention with VR technology. We also found age to modify the effect of certain constructs, with the younger generation more easily adapting to VR technology (effort expectancy) and demonstrating greater VR learning enjoyment (hedonic motivation).

> *I think the current generation of learners is…becoming more and more comfortable with virtual reality. So I think the buy-in of our new generation of learners is going to be really quick…And so I think they're going to help drive the need for this type of education.* [Participant #5]

The UTAUT2 theoretical framework was able to explain patterns in use behavior and intention related to VR simulation, and provided conclusions relevant to educational practice. However, we modified the UTAUT2 model, most notably adding "access expectancy" as an independent driver of behavioral intention. The importance of distance learning, expanded access, and equity in educational experience was sufficient to qualify as an independent construct. To illustrate its relative importance, there were more instances of coding for access expectancy than hedonic motivation, habit, or social influence. This may reflect that the UTAUT was initially described in the individual consumer marketplace while "access expectancy" applies more to the context of the educational technology marketplace. Further research would be necessary to explore this hypothesis.

The other notable change was seen in the UTAUT modifier "gender." There were no instances of coding applicable to gender by either independent reviewer, and themes identified did not differ by participant gender. We found no signal for gender as a modifier of any UTAUT2 construct. We suggest that this is a reflection of both time and context. The initial publication of the UTAUT was in 2003, wherein it was discussed that effort expectancies may be more salient to women than men, and that women may be more sensitive to others' opinions than men [16]. We believe this contextualization of gender roles and social norms to be antiquated and due for revision. Furthermore, in this cohort of career medical educators, we found no differences in motivating factors between men and women related to their intention or use behavior with simulation technology. Therefore, we eliminated "gender" from our thematic map.

Discussion

Principal Findings

Participating early adopters have adapted their use of VR to meet specific educational needs. Whether it be the need for

XSL•FO
RenderX

distance learning during the pandemic, the need to bridge geographical or institutional divides, or the need for wide dissemination of teaching to address gaps in knowledge or skills, early adopters are implementing VR as a method to expand access to high-yield educational interventions. In terms of the role that VR served among the studied population, it was used primarily to teach cognitive skills as opposed to psychomotor or procedural skills. The most common factors that affected how successful any given implementation of VR would be is related to how educators managed their resources (funding, time, and design effort) and the degree to which they were able to foster learning through deep immersion.

The results of this study establish drivers of use behavior, providing practical insights into the educational gaps that VR might address in the future. Furthermore, this study contributes validity evidence for the UTAUT framework in studying the evolving role of immersive VR simulation in health professions education. This study represents an advancement to the literature in this field as it encompasses a wider variety of VR use cases than prior work, and it uses a well-validated theoretical framework to reflect on the perspectives of a diverse population within the health professions education community.

Implications for Future Research

Based on our findings, we synthesize a set of research questions that may help define best practices for future VR development and implementation (Table 2). We also list example study ideas corresponding to each research question in order to provide additional context and encourage reflection. These examples are not meant to be comprehensive or prescriptive, but rather to demonstrate how researchers might approach these questions with a variety of different methodologies and paradigms that could advance the literature in this field.

Table . Suggested research questions for immersive VR[a] simulation technology.

Themes from this study and suggested research questions		Example future research study
Focus on cognitive skills		
	How can we best implement immersive VR simulation given its strength in teaching cognitive skills (eg, communication, teamwork, clinical reasoning, situational awareness, and interdisciplinary skills)?	Multi-institutional study comparing 2 different VR implementation methods and using a validated assessment for cognitive skills
	What innovations can improve the teaching of fine psychomotor tasks in the VR environment?	Validation study using novel haptic gloves for VR simulation and assessing learning outcomes
Access to education		
	How can VR be most effectively leveraged to provide distance learning?	Mixed methods (quantitative or qualitative) needs assessment for distance simulation learning in post-COVID health professions education
	How can VR be used as a tool to create equity of educational experience?	Comparative study of learner outcomes at highly resourced centers versus resource-limited training programs for VR simulation
	How can VR facilitate collaboration on a larger scale (eg, national or international)?	Descriptive study demonstrating feasibility of an international VR curriculum offered by a professional society
Resource investment		
	What are the upfront investments and preparation processes necessary to start a new VR simulation program?	Focus group study of early adopters with concentration on preparation and upfront costs for establishing a VR simulation program
	How can we increase availability and decrease barriers for using pre-existing VR software programs?	Thematic analysis of focus groups after piloting a VR multi-case library targeting undergraduate medical education learners
	What processes facilitate the creation of novel VR software that is relevant to external users, institutions, and learner groups?	Systematic review and subsequent guideline development project to describe best practices in creation of VR curricula
Balancing immersion		
	How can we achieve sufficient immersion to accomplish intended learning objectives without creating extraneous cognitive load and frustrating part-tasks?	Comparative study of learning outcomes in a high-fidelity versus low-fidelity VR environment

[a]VR: virtual reality.

Much ongoing VR simulation research focuses on demonstrating that VR is equally or more effective than traditional simulation modalities [3-6,28]. While this is an important question, it risks overshadowing other questions that are raised by early adopters

XSL•FO
RenderX

in this study: how might we improve the ability of VR technology to teach psychomotor skills? How can we use VR simulation to create equity between learner populations? What solutions exist for shared and collaborative creation of VR software? How can we leverage the incentives of the marketplace for VR technology companies? These questions could significantly impact the future use of this technology in health professions education.

Study Limitations

This study has several limitations. First, early adopters tend to be optimistic about the advantages of new technologies, sometimes underemphasizing associated challenges. To account for this bias, we designed our interview guide with prompts targeted equally toward the advantages and challenges of VR technology, and we practiced reflexivity among this study's team to appreciate the effects of any personal biases. Future studies should examine perspectives from early majority, late majority, and laggards, but these groups are not yet readily identifiable.

Second, this study used nonprobability sampling, which harbors potential for bias toward participants with similar experiences and perspectives. However, the small size of the VR educator community limits the feasibility of random sampling. We therefore attempted to compensate by seeking participants with diverse experiences, working with different software applications, different VR companies, or different learner populations. To further assure accuracy and freedom from bias, future studies should attempt triangulation of this data, for example via data source triangulation using focus groups or via theory triangulation, analyzing this data through a different theoretical lens [29].

Third, while we targeted diversity, all health professions were not represented. Our sample included only individuals from the United States and Canada, and participants from rural workplace settings were underrepresented. These considerations may be important, particularly if geography is found to independently affect use behavior.

Finally, regarding the UTAUT modifier "experience," our sample size was inadequate to analyze its role as a modifier of use behavior. Therefore, we did not include experience in our thematic map and further research will be necessary to explore how experience may affect use behavior with VR simulation.

Conclusion

We used the UTAUT2 framework in a directed content analysis using semistructured interviews to investigate the role of immersive VR simulation in health professions education. We identified 4 key themes elucidating use behavior related to VR simulation, suggesting its optimal applications include teaching cognitive skills, expanding access to educational experiences, offering a collaborative repository of relevant simulation scenarios, and enhancing immersion for intended learning objectives. These themes may help to inform best practices for the future development and implementation of immersive VR simulation programs.

As immersive VR simulation technology continues to evolve in health professions education, the VR educator community will continue to grow alongside the rapid technological advancements. Therefore, defining best practices for integrating this technology into training programs is critical. Future research should focus on leveraging VR simulation's unique capabilities as compared to traditional simulation modalities.

Acknowledgments

The publication of this study was supported by a Fellowship Grant from the Stony Wold-Herbert Fund, Inc.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Semistructured interview guide.
[DOCX File, 22 KB - mededu_v11i1e62803_app1.docx]

References

1. Pottle J. Virtual reality and the transformation of medical education. Future Healthc J 2019 Oct;6(3):181-185. [doi: 10.7861/fhj.2019-0036] [Medline: 31660522]
2. Haowen J, Vimalesvaran S, Kyaw BM, Car LT. Virtual reality in medical students' education: a scoping review protocol. BMJ Open 2021 May 26;11(5):e046986. [doi: 10.1136/bmjopen-2020-046986] [Medline: 34039577]
3. Abulfaraj MM, Jeffers JM, Tackett S, Chang T. Virtual reality vs. high-fidelity mannequin-based simulation: a pilot randomized trial evaluating learner performance. Cureus 2021 Aug;13(8):e17091. [doi: 10.7759/cureus.17091] [Medline: 34527478]
4. Nassar AK, Al-Manaseer F, Knowlton LM, Tuma F. Virtual reality (VR) as a simulation modality for technical skills acquisition. Ann Med Surg (Lond) 2021 Nov;71:102945. [doi: 10.1016/j.amsu.2021.102945] [Medline: 34840738]
5. Khan R, Plahouras J, Johnston BC, Scaffidi MA, Grover SC, Walsh CM. Virtual reality simulation training in endoscopy: a Cochrane review and meta-analysis. Endoscopy 2019 Jul;51(7):653-664. [doi: 10.1055/a-0894-4400] [Medline: 31071757]

XSL•FO
RenderX

6. Berg H, Steinsbekk A. The effect of self-practicing systematic clinical observations in a multiplayer, immersive, interactive virtual reality application versus physical equipment: a randomized controlled trial. Adv Health Sci Educ Theory Pract 2021 May;26(2):667-682. [doi: 10.1007/s10459-020-10019-6] [Medline: 33511505]
7. Katz D, Shah R, Kim E, et al. Utilization of a voice-based virtual reality advanced cardiac life support team leader refresher: prospective observational study. J Med Internet Res 2020 Mar 12;22(3):e17425. [doi: 10.2196/17425] [Medline: 32163038]
8. Kyaw BM, Saxena N, Posadzki P, et al. Virtual reality for health professions education: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959. [doi: 10.2196/12959] [Medline: 30668519]
9. Saab MM, Hegarty J, Murphy D, Landers M. Incorporating virtual reality in nurse education: a qualitative study of nursing students' perspectives. Nurse Educ Today 2021 Oct;105:105045. [doi: 10.1016/j.nedt.2021.105045] [Medline: 34245956]
10. Wong M, Chue S, Jong M, Benny HWK, Zary N. Clinical instructors' perceptions of virtual reality in health professionals' cardiopulmonary resuscitation education. SAGE Open Med 2018;6:2050312118799602. [doi: 10.1177/2050312118799602] [Medline: 30245815]
11. Snell L, Son D, Onishi H. Instructional design. In: Swanwick T, Forrest K, O'Brien BC, editors. Understanding Medical Education: Evidence, Theory, and Practice: The Association for the Study of Medical Education (ASME); 2018:89-100. [doi: 10.1002/9781119373780]
12. Scalea JR. Technology for technology's sake no longer. Ann Surg 2019 Feb;269(2):e24. [doi: 10.1097/SLA.0000000000003132] [Medline: 30614843]
13. Lau KHV, Greer DM. Using technology adoption theories to maximize the uptake of e-learning in medical education. Med Sci Educ 2022 Apr;32(2):545-552. [doi: 10.1007/s40670-022-01528-7] [Medline: 35261814]
14. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
15. Momani A. The unified theory of acceptance and use of technology: a new approach in technology acceptance. Int J Soc Knowled Develop 2020 Jul;12:79-98. [doi: 10.4018/IJSKD.2020070105]
16. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157. [doi: 10.2307/41410412]
17. Chen CJ. Theoretical bases for using virtual reality in education. Theme Sci Techn Educ 2009;2(2):71-90 [FREE Full text]
18. Lange AK, Koch J, Beck A, et al. Learning with virtual reality in nursing education: qualitative interview study among nursing students using the unified theory of acceptance and use of technology model. JMIR Nurs 2020;3(1):e20249. [doi: 10.2196/20249] [Medline: 34345791]
19. Bracq MS, Michinov E, Arnaldi B, et al. Learning procedural skills with a virtual reality simulator: an acceptability study. Nurse Educ Today 2019 Aug;79:153-160. [doi: 10.1016/j.nedt.2019.05.026] [Medline: 31132727]
20. Baniasadi T, Ayyoubzadeh SM, Mohammadzadeh N. Challenges and practical considerations in applying virtual reality in medical education and treatment. Oman Med J 2020 May;35(3):e125. [doi: 10.5001/omj.2020.43] [Medline: 32489677]
21. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
22. Brooks J, McCluskey S, Turley E, King N. The utility of template analysis in qualitative psychology research. Qual Res Psychol 2015 Apr 3;12(2):202-222. [doi: 10.1080/14780887.2014.955224] [Medline: 27499705]
23. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251. [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
24. DeJonckheere M, Vaughn LM. Semistructured interviewing in primary care research: a balance of relationship and rigour. Fam Med Community Health 2019;7(2):e000057. [doi: 10.1136/fmch-2018-000057] [Medline: 32148704]
25. Kallio H, Pietilä AM, Johnson M, Kangasniemi M. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. J Adv Nurs 2016 Dec;72(12):2954-2965. [doi: 10.1111/jan.13031] [Medline: 27221824]
26. Hennink MM, Kaiser BN, Marconi VC. Code saturation versus meaning saturation: how many interviews are enough? Qual Health Res 2017 Mar;27(4):591-608. [doi: 10.1177/1049732316665344] [Medline: 27670770]
27. Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE Guide No. 131. Med Teach 2020 Aug;42(8):846-854. [doi: 10.1080/0142159X.2020.1755030] [Medline: 32356468]
28. Regehr G. It's not rocket science: rethinking our metaphors for research in health professions education. Med Educ 2010 Jan;44(1):31-39. [doi: 10.1111/j.1365-2923.2009.03418.x] [Medline: 20078754]
29. Carter N, Bryant-Lukosius D, DiCenso A, Blythe J, Neville AJ. The use of triangulation in qualitative research. Oncol Nurs Forum 2014 Sep;41(5):545-547. [doi: 10.1188/14.ONF.545-547] [Medline: 25158659]

Abbreviations:

UTAUT: Unified Theory of Acceptance and Use of Technology
UTAUT2: Unified Theory of Acceptance and Use of Technology 2
VR: virtual reality

XSL•FO
RenderX

Edited by LT Car; submitted 07.06.24; peer-reviewed by K Doo, SH Stige; revised version received 16.01.25; accepted 24.01.25; published 12.03.25.

Please cite as:
Talan J, Forster M, Joseph L, Pradhan D
Exploring the Role of Immersive Virtual Reality Simulation in Health Professions Education: Thematic Analysis
JMIR Med Educ 2025;11:e62803
URL: https://mededu.jmir.org/2025/1/e62803
doi:10.2196/62803

© Jordan Talan, Molly Forster, Leian Joseph, Deepak Pradhan. Originally published in JMIR Medical Education (https://mededu.jmir.org), 12.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Knowledge Mapping and Global Trends in Simulation in Medical Education: Bibliometric and Visual Analysis

Hongjun Ba[1], MD; Lili Zhang[2], BA; Xiufang He[2], BA; Shujuan Li[2], MD

[1]Department of Pediatrics, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

[2]Department of Pediatric Cardiology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

Corresponding Author:
Hongjun Ba, MD
Department of Pediatrics
The First Affiliated Hospital
Sun Yat-sen University
58 Zhongshan Road 2
Guangzhou, 510080
China
Phone: 86 15920109625
Email: bahj3@mail.sysu.edu.cn

Abstract

Background: With the increasing recognition of the importance of simulation-based teaching in medical education, research in this field has developed rapidly. To comprehensively understand the research dynamics and trends in this area, we conducted an analysis of knowledge mapping and global trends.

Objective: This study aims to reveal the research hotspots and development trends in the field of simulation-based teaching in medical education from 2004 to 2024 through bibliometric and visualization analyses.

Methods: Using CiteSpace and VOSviewer, we conducted bibliometric and visualization analyses of 6743 articles related to simulation-based teaching in medical education, published in core journals from 2004 to 2024. The analysis included publication trends, contributions by countries and institutions, author contributions, keyword co-occurrence and clustering, and keyword bursts.

Results: From 2004 to 2008, the number of articles published annually did not exceed 100. However, starting from 2009, the number increased year by year, reaching a peak of 850 articles in 2024, indicating rapid development in this research field. The United States, Canada, the United Kingdom, Australia, and China published the most articles. Harvard University emerged as a research hub with 1799 collaborative links, although the overall collaboration density was low. Among the 6743 core journal articles, a total of 858 authors were involved, with Lars Konge and Adam Dubrowski being the most prolific. However, collaboration density was low, and the collaboration network was relatively dispersed. A total of 812 common keywords were identified, forming 4189 links. The keywords “medical education,” “education,” and “simulation” had the highest frequency of occurrence. Cluster analysis indicated that “cardiopulmonary resuscitation” and “surgical education” were major research hotspots. From 2004 to 2024, a total of 20 burst keywords were identified, among which “patient simulation,” “randomized controlled trial,” “clinical competence,” and “deliberate practice” had high burst strength. In recent years, “application of simulation in medical education,” “3D printing,” “augmented reality,” and “simulation training” have become research frontiers.

Conclusions: Research on the application of simulation-based teaching in medical education has become a hotspot, with expanding research areas and hotspots. Future research should strengthen interinstitutional collaboration and focus on the application of emerging technologies in simulation-based teaching.

(JMIR Med Educ 2025;11:e71844) doi:10.2196/71844

KEYWORDS

medical education; simulation-based teaching; bibliometrics; visualization analysis; knowledge mapping

XSL•FO
RenderX

Introduction

In the rapidly evolving landscape of medical education, the integration of simulation-based training has emerged as a pivotal innovation. Simulation in medical education encompasses a broad spectrum of methodologies, including high-fidelity mannequins, virtual reality, standardized patients, and computer-based simulations [1,2]. These techniques aim to enhance clinical skills, decision-making, and teamwork among medical professionals without the direct involvement of real patients.

The adoption of simulation in medical training addresses several critical challenges [3,4]. First, it provides a safe and controlled environment where learners can practice and refine their skills. This is particularly crucial in high-stakes scenarios such as emergency medicine, surgery, and critical care, where errors can have severe consequences [5,6]. In addition, simulation allows for repetitive practice and immediate feedback, facilitating a deeper understanding of complex procedures and concepts.

Over the past few decades, there has been a significant increase in research focused on the effectiveness and impact of simulation-based education in the medical field [7,8]. This growing body of literature reflects the widespread recognition of simulation as a valuable educational tool. However, the rapid expansion of this field necessitates a comprehensive review and analysis to understand its development, trends, and future directions.

Several bibliometric analyses have been conducted on simulation in medical education [9,10], highlighting its growing importance and impact. However, these studies often focus on specific aspects of simulation, such as surgical training or virtual reality. Our study complements this body of research by providing a comprehensive overview of the entire field, including emerging technologies like 3D printing and augmented reality (AR), and by analyzing collaborative networks and thematic trends over a 20-year period.

A bibliometric analysis provides an ideal approach to systematically evaluate the literature on simulation in medical education. By using quantitative methods to analyze publication patterns, citation networks, and research themes, bibliometric studies can offer valuable insights into the evolution of this field. Such an analysis can identify key contributors, influential publications, and emerging trends, thereby guiding future research and practice.

This study aims to conduct a bibliometric analysis of the literature on simulation in medical education. By examining the scope, growth, and impact of research in this area, we seek to elucidate the current state of the field and identify potential gaps and opportunities for further investigation. Specifically, this analysis will focus on the following objectives:

1. To map the overall publication trends and growth in simulation-based medical education research.
2. To identify the most influential journals, articles, and authors contributing to this field.
3. To explore the thematic evolution and emerging trends within the literature.
4. To assess the collaborative networks and geographical distribution of research activities.

Through this comprehensive bibliometric analysis, we hope to provide a clearer understanding of the trajectory and impact of simulation in medical education, ultimately contributing to the enhancement of educational practices and outcomes in the medical field.

Methods

Data Acquisition and Search Strategy

The search was conducted in the Web of Science Core Collection (WoSCC) database, which is widely recognized for its comprehensive coverage of high-quality, peer-reviewed literature [11,12]. While we acknowledge that including additional databases such as PubMed or Scopus could provide a more comprehensive dataset, the WoSCC was chosen for its superior bibliographic accuracy and extensive coverage of medical education research. Therefore, we opted to perform our search within this database. We conducted a search in the Web of Science for all relevant papers published between January 1, 2004, and December 31, 2024. The time frame from January 1, 2004, to December 31, 2024, was selected because it marks the period when simulation-based medical education began to gain significant traction in the literature, reflecting the growing recognition of its importance in medical training. In medical education, we define “simulation” as a teaching and training method that encompasses high-fidelity mannequins, virtual reality, standardized patients, and computer-based simulations.

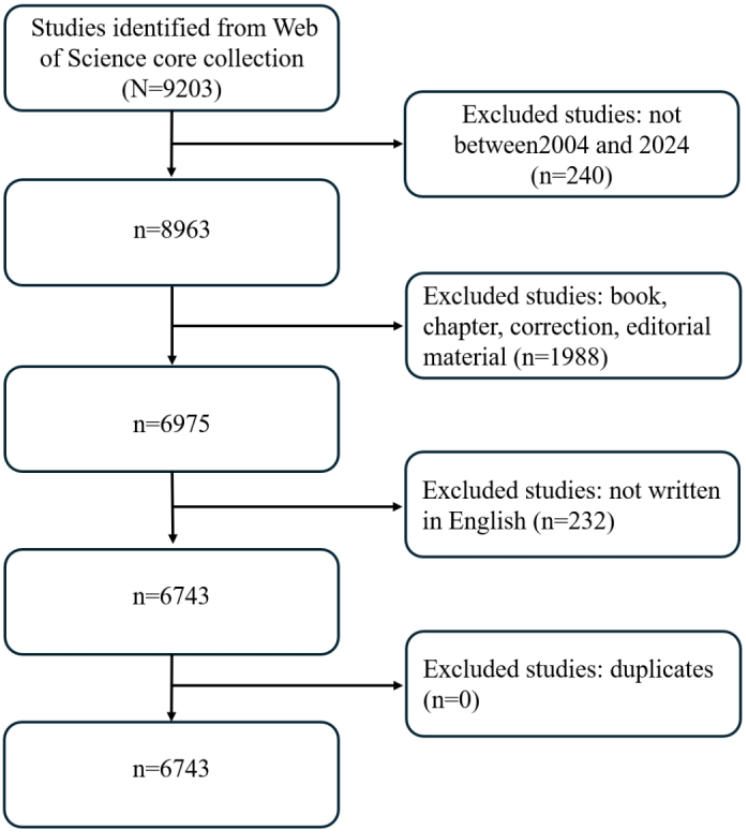
The search formula “TS=(Medical education) AND TS=(Simulation)” was used. The inclusion criteria were as follows: (1) full-text publications related to simulation in medical education, including original research articles and review articles; (2) articles written in English; and (3) papers published between January 1, 2004, and December 31, 2024. We excluded conference abstracts, theses, dissertations, and nonpeer-reviewed articles to ensure the quality and relevance of the data. The exclusion criteria were (1) topics not related to simulation in medical education and (2) papers in the form of conference abstracts, theses, dissertations, and non–peer-reviewed articles to ensure the quality and relevance of the data. A plain text version of the papers was exported.

General Data

Between January 1, 2004, and December 31, 2024, the WoSCC database recorded a total of 6743 publications concerning simulation in medical education. This body of literature included contributions from 121 countries and regions, 510 institutions, and 858 authors. Figure 1 shows the process of literature searching and bibliometric analysis.

XSL•FO
RenderX

Figure 1. The workflow of data collection and bibliometric analysis.



Data Analysis

We used GraphPad Prism (version 8.0.2; Dotmatics) to illustrate annual publication trends. The methodological approach was validated through the use of CiteSpace and VOSviewer, both of which are widely recognized and extensively used in bibliometric research [13,14]. These tools have been shown to provide robust and reliable analyses of large-scale bibliometric data.

VOSviewer, a Java-based software developed by van Eck and Waltman in 2009, facilitates the construction of various types of network maps, such as bibliographic coupling, cocitation, and coauthorship networks. CiteSpace, developed by Professor Chaomei Chen, provides a dynamic platform for identifying and visualizing patterns and trends in scientific literature, enabling the exploration of knowledge domains and predictive analysis of research trajectories [14]. Our methodological approach involved setting specific parameters for network density (eg, keyword co-occurrence density of 0.0127), node inclusion thresholds (eg, minimum occurrence frequency of keywords), and time-slicing techniques to analyze temporal changes. The references corresponding to the software applications were verified against our citation list to ensure accuracy [13,14]. When using VOSviewer and CiteSpace for bibliometric analysis, we established standards for defining international collaboration. This was done by examining the authorship of papers, specifically the first and corresponding authors, to ensure a comprehensive capture of collaborative efforts from researchers across different countries.

Burst detection in CiteSpace is based on the Kleinberg algorithm, which models document streams using infinite-state automata to extract meaningful structures [15]. These analyses can reveal rapidly growing topics over extended periods as well as short-term themes.

The rationale for selecting these techniques lies in their widespread application and effectiveness in bibliometric research. They provide robust and complementary insights into productivity, impact, and collaboration patterns within research fields.

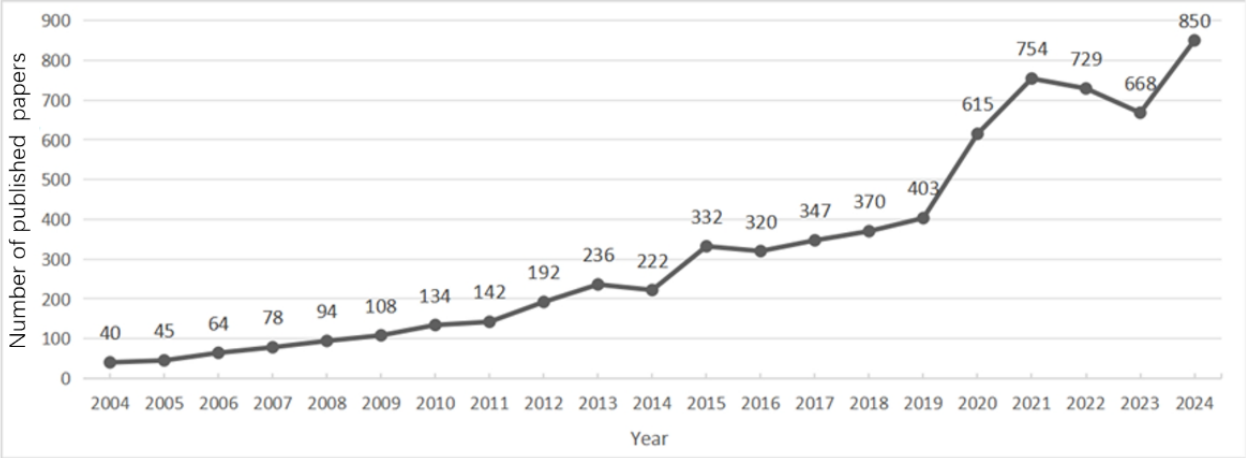
Results

Publication Trend

Figure 2 shows that from 2004 to 2008, the annual number of publications on simulation teaching in medical education did not exceed 100 articles, indicating that research in this field was still in its nascent stage. Since 2009, the number of publications in this field has steadily increased, showing a trend of fluctuating growth. Specifically, the number of publications in 2015 surpassed 300 for the first time, and by 2020, this number had exceeded 500. This significant increase marks the growing attention and interest of scholars and researchers in the field of simulation teaching in medical education. Since 2020, the annual number of publications in this field has consistently remained above 500, reaching a peak of 850 articles in 2024. This further highlights the vigorous development and extensive influence of research in the field of simulation teaching in medical education.

XSL•FO
RenderX

Figure 2. Trend chart of publications in the past 20 years.

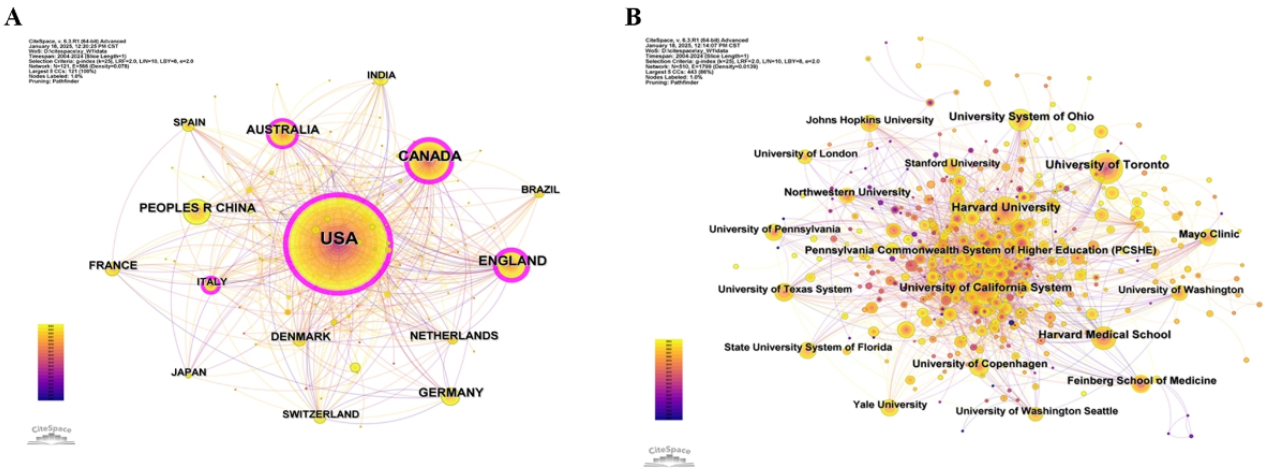


Country or Region and Institution Contributions

According to Figure 3A, the connections between circular nodes representing different countries to some extent reflect the existence of relationships and collaborations between these countries. Furthermore, the density of these connections in the network can serve as an important indicator of the closeness of collaborative relationships between countries. Among them, the countries with the highest number of publications are the United States (3083 articles), Canada (776 articles), England (510 articles), Australia (381 articles), and China (375 articles) (Table 1). In addition, countries such as Italy, the Netherlands, Sweden, and Belgium have numerous connections, indicating a complex network of relationships, which suggests that these countries have relatively close research collaborations with other regions.

Using the CiteSpace software, an institutional collaboration network diagram was obtained, as shown in Figure 3B. Upon statistical analysis, it was found that there are a total of 510 research institutions forming 1799 connections, with Harvard University being the central hub. The diagram reveals that the network density is 0.0139, indicating relatively weak collaborative relationships between research institutions, with a significant portion of them operating in a relatively independent research state. In terms of research output, the top-10 institutions by the number of publications are Harvard University, the University of Toronto, the University of California System, the University System of Ohio, Harvard Medical School, Mayo Clinic, Northwestern University, Feinberg School of Medicine, the University of Copenhagen, and Pennsylvania Commonwealth System of Higher Education (Table 2).

Figure 3. Network graph of national and institutional collaborations. (A) Network graph of national collaborations. (B) Network graph of institutional collaborations. The bubble size represents the number of publications.



XSL•FO
RenderX

Table 1. Top-10 most productive countries or regions.

Rank	Country or region	Articles, n	Centrality	Percentage	Half-life
1	United States	3083	0.51	45.71%	14.5
2	Canada	776	0.22	11.52%	13.5
3	England	510	0.25	7.57%	15.5
4	Australia	381	0.12	5.64%	15.5
5	China	375	0.02	5.56%	15.5
6	Germany	340	0.05	5.04%	16.5
7	France	189	0.03	2.8%	15.5
8	Denmark	184	0.05	2.73%	15.5
9	The Netherlands	164	0.1	2.43%	15.5
10	Switzerland	142	0.05	2.1%	16.5

Table 2. Top-10 most productive institutions.

Rank	Institution	Country	Studies, n	Centrality	Half-life
1	Harvard University	United States	167	0.06	13.5
2	University of Toronto	Canada	153	0.04	11.5
3	University of California System	United States	125	0.01	12.5
4	University System of Ohio	United States	118	0.04	11.5
5	Harvard Medical School	United States	82	0.09	13.5
6	Mayo Clinic	United States	68	0.03	12.5
7	Northwestern University	United States	66	0.01	10.5
8	Feinberg School of Medicine	United States	65	0.01	10.5
9	University of Copenhagen	Denmark	61	0.01	14.5
10	Pennsylvania Commonwealth System of Higher Education	United States	60	0.02	13.5

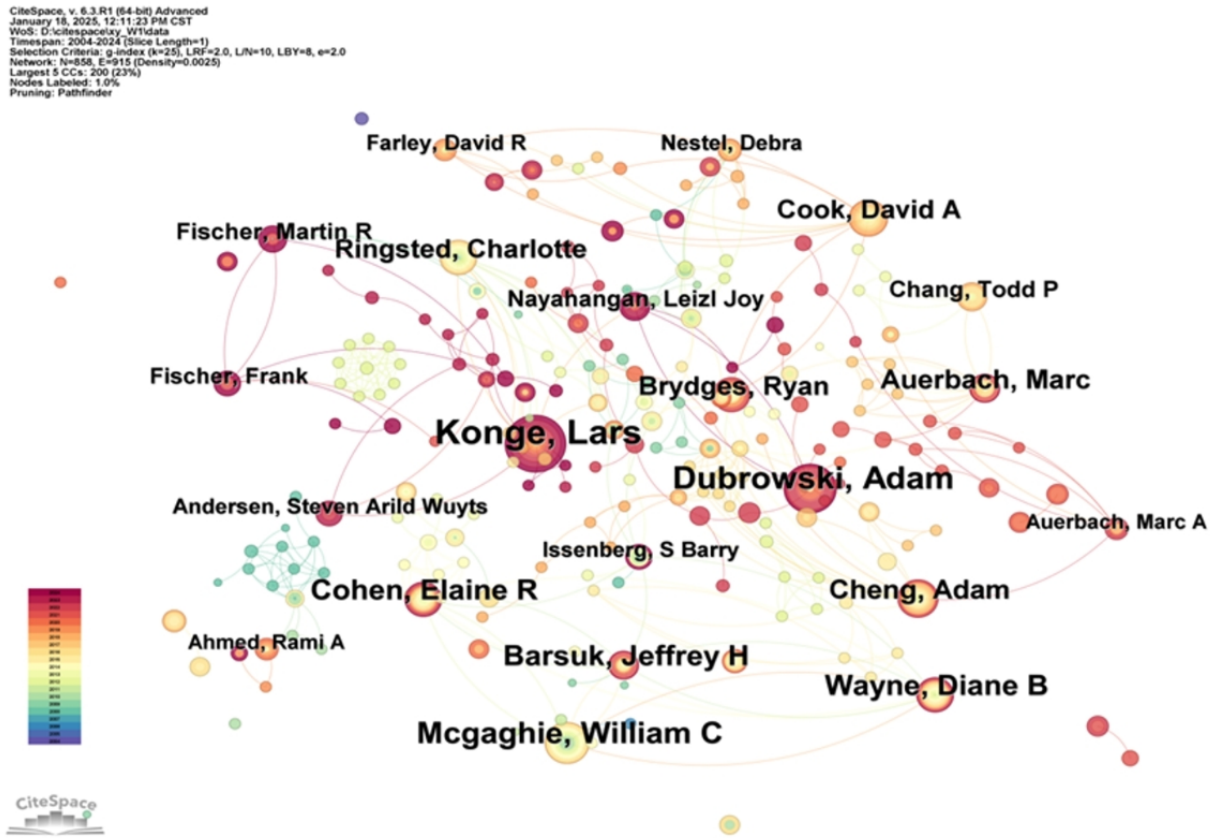
Author Collaborations

The sample data were processed using CiteSpace, and the resulting author co-occurrence map is shown in Figure 4. In this map, each node represents a different author. The size of the node indicates the author's publication frequency, meaning the larger the node, the more publications the author has. When nodes are presented in the form of annual rings, the bandwidth of the color band corresponding to a particular year represents the number of papers published by the author that year, with a wider ring indicating more publications. The lines between nodes represent collaborative relationships between organizations or authors, with the thickness of the lines indicating the degree of collaboration.

Among the 6743 core journal articles, a total of 858 authors were involved. The top-10 authors by publication volume are Konge, Lars (69 papers); Dubrowski, Adam (50 papers); McGaghie, William C (37 papers); Wayne, Diane B (33 papers); Cohen, Elaine R (33 papers); Barsuk, Jeffrey H (30 papers); Auerbach, Marc (26 papers); Cheng, Adam (24 papers); Cook, David A (23 papers); and Ringsted, Charlotte (22 papers). Authors with 7 or more publications, a total of 44 individuals, were classified as the core author group, which accounts for only 5.1% of the total authors. In addition, there are 915 collaboration lines among the authors on the map, with a collaboration density of 0.0025, indicating a low-density level. The number of lines is relatively sparse, and the collaboration network map shows a relatively dispersed pattern. The largest collaboration network system is formed by the research team centered around Dubrowski, Adam; Nayahangan, Leizl Joy; Cheng, Adam; Auerbach, Marc A; and Cook, David A. The scale of collaboration is mainly presented in the form of individual or small-scale research teams, indicating that the core research team in this field has yet to be fully established.

XSL•FO
RenderX

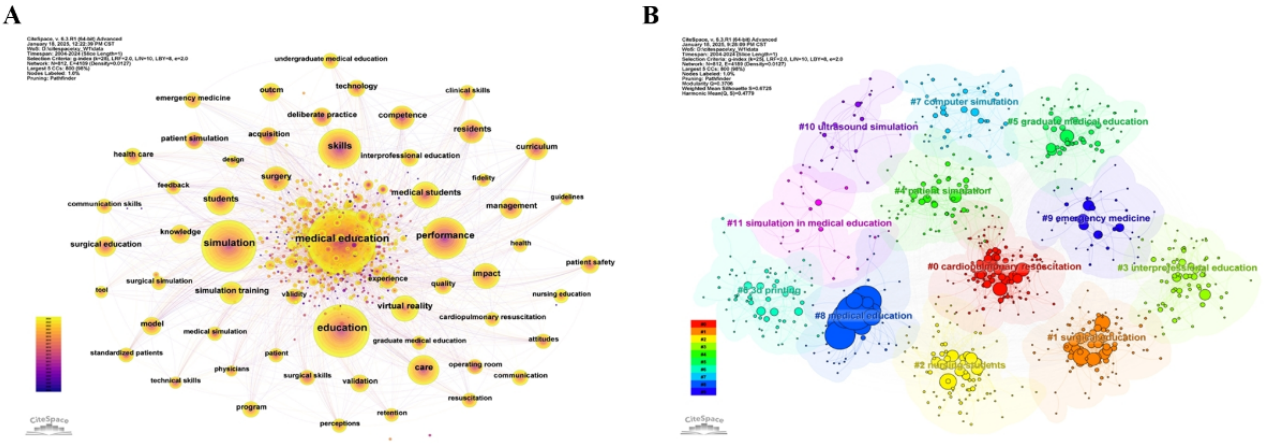
Figure 4. Network diagram of author collaborations. The bubble size represents the number of publications.



Keyword Co-Occurrence and Cluster

Using CiteSpace software to conduct a keyword co-occurrence analysis on the sample, the constructed keyword co-occurrence map is shown in Figure 5A. From the keyword co-occurrence analysis, a total of 812 common keywords were identified, forming 4189 connections, with a network density of 0.0127. The most frequently occurring keyword is "medical education," accounting for 7.9%. This is followed by "education" and "simulation," which account for 5.48% and 5.14%, respectively. The keywords "performance" and "skills" account for 3.49% and 3.29%, respectively. These keywords represent the current research hotspots and status in the field of simulation teaching in medical education.

Figure 5. Keyword co-occurrence and keyword clustering map. (A) Keyword co-occurrence map. (B) Keyword clustering map. The bubble size represents the number of publications.



Based on the keyword co-occurrence map, the log-likelihood ratio algorithm was used to cluster the keywords, resulting in a keyword clustering co-occurrence map. The Q value is 0.3707 (>0.3), and the S value is 0.6725 (>0.5), indicating a significant clustering structure and a high degree of clustering match. The map displays a total of 10 clustering areas, among which "cardiopulmonary resuscitation," "surgical education," "nursing students," "interprofessional education," and "patient simulation" are the five largest clusters (Figure 5B). Specifically, medical simulation teaching has become an important

XSL•FO
RenderX

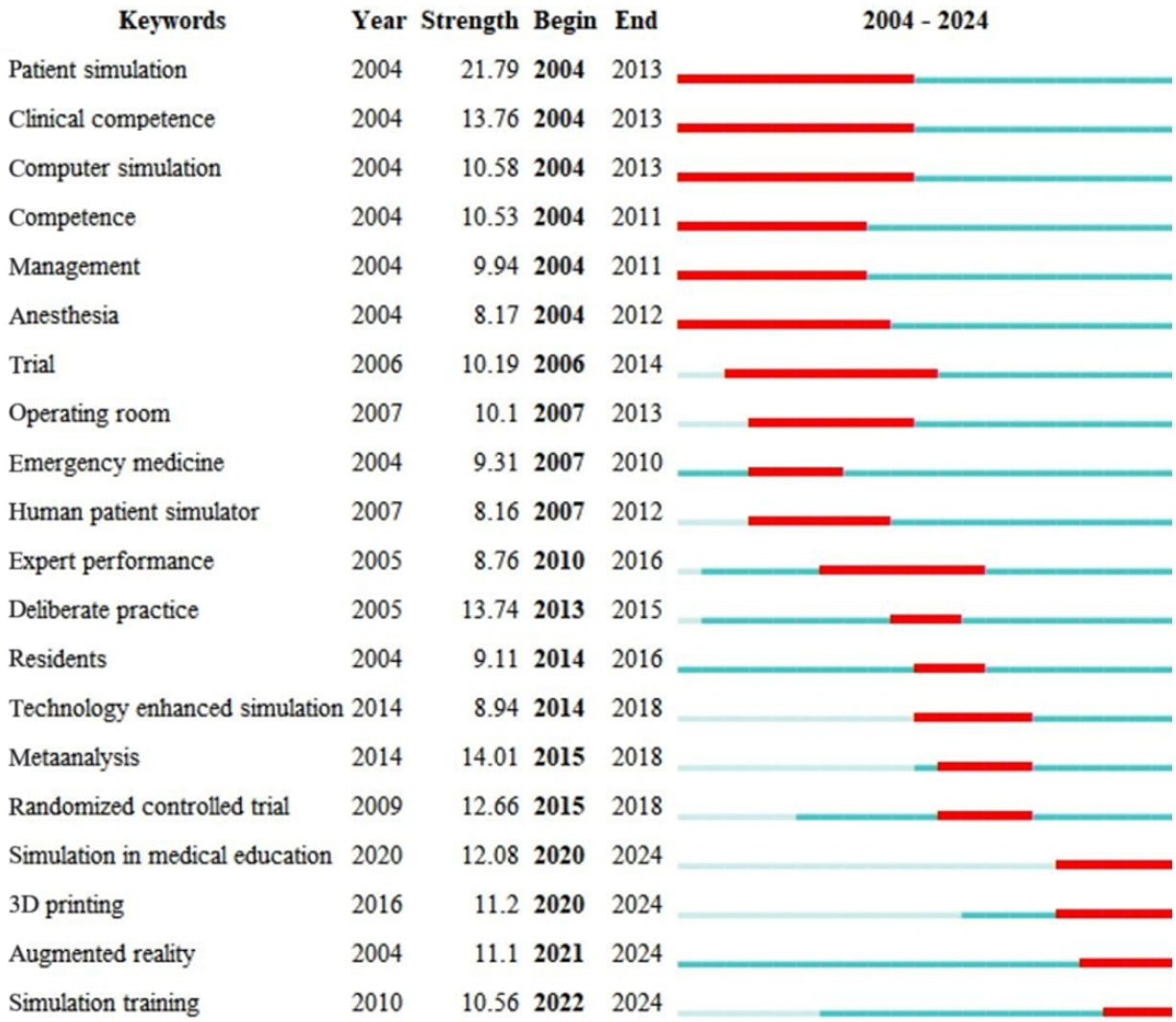
component of medical education, widely applied in various fields including cardiopulmonary resuscitation, surgical education, and nursing student training.

Keyword Citation Bursts

The keyword burst visualization analysis identified a total of 20 keywords in the field of simulation teaching in medical education from 2004 to 2024, along with their emergence intensity and start-end years. The relevant literature on keyword emergence is shown in Figure 6.

It can be observed that the keywords with high emergence intensity include "patient simulation," "randomized controlled trial," "clinical competence," and "deliberate practice." Meanwhile, the research area began to focus on "patient simulation" as early as 2004, which, along with "clinical competence" and "computer simulation," became one of the keywords with the longest duration of emergence. In recent years, researchers have increasingly focused on themes such as "trial" and "expert performance." The keywords that are still emerging represent the current research frontiers and trends, which include "simulation in medical education," "3D printing," "augmented reality," and "simulation training." The emergence of "3D printing" reflects the growing interest in using patient-specific anatomical models for surgical planning and training, offering a more personalized and immersive learning experience. Similarly, "augmented reality" signifies the integration of advanced technologies to create interactive and realistic training environments, enhancing the acquisition of clinical skills. These emerging trends highlight the transformative potential of technology in medical education, paving the way for more innovative and effective teaching methodologies.

Figure 6. Keyword burst graph (sorted by the beginning year of the burst). The blue bars denote the reference has been published; the red bars denote citation burstiness.



XSL•FO
RenderX

Discussion

Principal Findings

The results of our bibliometric analysis provide a comprehensive overview of the evolution, collaboration patterns, and thematic focus of simulation-based education research in the medical field. Key trends include a steady increase in publications from 2004 to 2024, particularly a surge after 2009, indicating a growing recognition of simulation's importance in medical education. By 2024, the publication count had peaked at 850, highlighting a transition of simulation from a novel approach to a staple in medical education. In addition, the United States emerged as the leading contributor with 3083 articles, reflecting substantial investment in education and research. Harvard University is a central hub for simulation-based medical education, despite a fragmented institutional landscape. Prominent authors like Lars Konge, Adam Dubrowski, and William C McGaghie drive the field, though the low density of collaborative networks suggests room for enhanced inter-institutional teamwork. Keyword analysis underscores the focus on competency-based education and practical skill acquisition, with emerging technologies like 3D printing and AR shaping future directions.

Comparison to Literature

Our findings are consistent with existing literature [16,17], which also highlights the increasing role of simulation in medical education over the past two decades. Previous studies have documented the rise in publications and the central role of the United States and key institutions like Harvard in advancing this field. However, our analysis provides a more granular look at the collaborative networks and thematic focuses, revealing a fragmented institutional landscape and the emergence of cutting-edge technologies that are less emphasized in earlier reviews.

Implications of Findings

The implications of these findings are multifaceted. The robust growth in simulation-based medical education research indicates a broad acceptance of its efficacy in improving medical training. The strong international collaboration suggests that best practices and innovative methodologies are being shared globally, potentially standardizing and enhancing simulation protocols. The emergence of new technologies like 3D printing and AR points to a future where simulation-based education will be more immersive and technologically advanced [18-20], which could significantly enhance learning outcomes and patient care. The integration of 3D printing and AR into simulation-based training can significantly improve clinical outcomes. 3D-printed anatomical models enable patient-specific simulations, allowing surgeons to practice complex procedures before operating on real patients, thus enhancing precision and reducing errors [21]. Similarly, AR creates immersive training environments, providing real-time feedback and interactive learning to enhance clinical skill acquisition [22]. However, challenges such as the high cost of equipment and the need for specialized training for educators and learners may limit their widespread adoption. Future research should explore cost-effective solutions to overcome these barriers and ensure broader access to these technologies in medical institutions.

The emergence of "cardiopulmonary resuscitation" as a major research hotspot reflects its critical importance in medical education and clinical practice. Cardiopulmonary resuscitation is a high-stakes procedure where errors can have severe consequences, making it an ideal candidate for simulation-based training [23]. Simulation allows learners to practice cardiopulmonary resuscitation in a controlled environment, receive immediate feedback, and refine their skills through repetitive practice [24]. This not only enhances individual competence but also improves team dynamics and communication during real-life emergencies.

Similarly, the focus on "surgical education" underscores the need for advanced training methods to prepare surgeons for complex procedures. Simulation-based training in surgical education has been shown to improve technical skills, reduce operative time, and enhance patient safety [25]. These findings highlight the transformative potential of simulation in addressing critical gaps in medical education and improving clinical outcomes.

Limitations

While the bibliometric analysis provides valuable insights, it has several limitations. First, the data might not capture all relevant publications, particularly those in non-English languages or those in less accessible databases, which could introduce selection bias. Second, the analysis relies on citation metrics, which may not fully reflect the quality or practical impact of the research. For instance, highly cited articles may not always represent the most impactful studies in terms of educational outcomes. Third, the low density of collaborative networks suggests that our findings might underrepresent the potential for interinstitutional synergy and innovation. Finally, a limitation of this study is the reliance on a single database (WoSCC), which may not capture all relevant publications. Future studies could expand the search to include additional databases such as PubMed and Scopus to enhance the robustness of the findings.

Suggestions

To address the identified limitations and enhance the impact of simulation-based education research, we suggest the following:

1. Increasing efforts to include diverse and international publications in future analyses.
2. Encouraging more interinstitutional collaborations to create a more cohesive research landscape.
3. Fostering larger, integrated research teams to deepen the scope of studies and drive innovation.
4. Embracing and further investigating emerging technologies to stay at the forefront of educational advancements.

Conclusions

In conclusion, the bibliometric analysis of simulation in medical education research reveals a dynamic field characterized by rapid growth, strong international collaboration, and evolving thematic focuses. The increasing trend in publications, significant contributions from leading countries and institutions,

XSL•FO
RenderX

and the integration of new technologies underscore the impactful nature of this research area. Moving forward, enhancing collaboration among institutions and expanding the core author network will be crucial. Future research should focus on integrating emerging technologies, such as 3D printing and AR, into medical education. For instance, studies could explore how 3D-printed anatomical models can enhance surgical training by providing realistic, patient-specific simulations.

Data Availability

All datasets generated for this study were included in the manuscript.

Authors' Contributions

HB conceived and designed the ideas for the manuscript. HB, LZ, XH, and SL participated in all data collection and processing. HB was the major contributor in organizing records and drafting the manuscript. All authors proofread and approved the manuscript.

Conflicts of Interest

None declared.

References

1. Motola I, Devine LA, Chung HS, Sullivan JE, Issenberg SB. Simulation in healthcare education: a best evidence practical guide. AMEE Guide No. 82. Med Teach 2013;35(10):e1511-e1530. [doi: 10.3109/0142159X.2013.818632] [Medline: 23941678]
2. So HY, Chen PP, Wong GKC, Chan TTN. Simulation in medical education. J R Coll Physicians Edinb 2019;49(1):52-57. [doi: 10.4997/JRCPE.2019.112] [Medline: 30838994]
3. Guze PA. Using technology to meet the challenges of medical education. Trans Am Clin Climatol Assoc 2015;126:260-270 [FREE Full text] [Medline: 26330687]
4. Kennedy CC, Cannon EK, Warner DO, Cook DA. Advanced airway management simulation training in medical education: a systematic review and meta-analysis. Crit Care Med 2014;42(1):169-178. [doi: 10.1097/CCM.0b013e31829a721f] [Medline: 24220691]
5. Hammond J. Simulation in critical care and trauma education and training. Curr Opin Crit Care 2004;10(5):325-329. [doi: 10.1097/01.ccx.0000140950.47361.c9] [Medline: 15385746]
6. Martín Parra JI, Manuel Palazuelos JC, Gómez Fleitas M. Pursuing quality in simulation-based surgical education. Cir Esp 2013;91(10):623-624. [doi: 10.1016/j.ciresp.2013.06.013] [Medline: 24143942]
7. Hepps JH, Yu CE, Calaman S. Simulation in medical education for the hospitalist: moving beyond the mock code. Pediatr Clin North Am 2019;66(4):855-866. [doi: 10.1016/j.pcl.2019.03.014] [Medline: 31230627]
8. Lu J, Cuff RF, Mansour MA. Simulation in surgical education. Am J Surg 2021;221(3):509-514. [doi: 10.1016/j.amjsurg.2020.12.016] [Medline: 33358139]
9. Walsh C, Lydon S, Byrne D, Madden C, Fox S, O Connor P. The 100 most cited articles on healthcare simulation: a bibliometric review. Simul Healthc 2018;13(3):211-220. [doi: 10.1097/SIH.0000000000000293] [Medline: 29613918]
10. Yao S, Tang Y, Yi C, Xiao Y. Research hotspots and trend exploration on the clinical translational outcome of simulation-based medical education: a 10-year scientific bibliometric analysis from 2011 to 2021. Front Med (Lausanne) 2021;8:801277 [FREE Full text] [doi: 10.3389/fmed.2021.801277] [Medline: 35198570]
11. Wu H, Li Y, Tong L, Wang Y, Sun Z. Worldwide research tendency and hotspots on hip fracture: a 20-year bibliometric analysis. Arch Osteoporos 2021;16(1):73. [doi: 10.1007/s11657-021-00929-2] [Medline: 33866438]
12. Vargas JS, Livinski AA, Karagu A, Cira MK, Maina M, Lu Y, et al. A bibliometric analysis of cancer research funders and collaborators in Kenya: 2007-2017. J Cancer Policy 2022;33:100331 [FREE Full text] [doi: 10.1016/j.jcpo.2022.100331] [Medline: 35792397]
13. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010;84(2):523-538 [FREE Full text] [doi: 10.1007/s11192-009-0146-3] [Medline: 20585380]
14. Chen C. CiteSpace: A Practical Guide for Mapping Scientific Literature. New York, NY: Nova Science Publishers; 2016.
15. Kleinberg J. Bursty and hierarchical structure in streams. Data Min Knowl Discov 2003;7:373-397. [doi: 10.1023/A:1024940629314]
16. Elendu C, Amaechi DC, Okatta AU, Amaechi EC, Elendu TC, Ezeh CP, et al. The impact of simulation-based training in medical education: a review. Medicine (Baltimore) 2024;103(27):e38813 [FREE Full text] [doi: 10.1097/MD.0000000000038813] [Medline: 38968472]
17. Higgins M, Madan C, Patel R. Development and decay of procedural skills in surgery: a systematic review of the effectiveness of simulation-based medical education interventions. Surgeon 2021;19(4):e67-e77. [doi: 10.1016/j.surge.2020.07.013] [Medline: 32868158]

XSL•FO
RenderX

18. Childs BS, Manganiello MD, Korets R. Novel education and simulation tools in urologic training. Curr Urol Rep 2019;20(12):81. [doi: 10.1007/s11934-019-0947-8] [Medline: 31782033]
19. Smith B, Dasgupta P. 3D printing technology and its role in urological training. World J Urol 2020;38(10):2385-2391. [doi: 10.1007/s00345-019-02995-1] [Medline: 31676911]
20. Sun Z, Wong YH, Yeong CH. Patient-specific 3D-printed low-cost models in medical education and clinical practice. Micromachines (Basel) 2023;14(2):464 [FREE Full text] [doi: 10.3390/mi14020464] [Medline: 36838164]
21. Ravi P, Chepelev LL, Stichweh GV, Jones BS, Rybicki FJ. Medical 3D printing dimensional accuracy for multi-pathological anatomical models 3D printed using material extrusion. J Digit Imaging 2022;35(3):613-622 [FREE Full text] [doi: 10.1007/s10278-022-00614-x] [Medline: 35237891]
22. Nagayo Y, Saito T, Oyama H. Augmented reality self-training system for suturing in open surgery: a randomized controlled trial. Int J Surg 2022;102:106650 [FREE Full text] [doi: 10.1016/j.ijsu.2022.106650] [Medline: 35525415]
23. Demirtas A, Guvenc G, Aslan, Unver V, Basak T, Kaya C. Effectiveness of simulation-based cardiopulmonary resuscitation training programs on fourth-year nursing students. Australas Emerg Care 2021;24(1):4-10. [doi: 10.1016/j.auec.2020.08.005] [Medline: 32933888]
24. Laco RB, Stuart W. Simulation-based training program to improve cardiopulmonary resuscitation and teamwork skills for the urgent care clinic staff. Mil Med 2022;187(5-6):e764-e769. [doi: 10.1093/milmed/usab198] [Medline: 34050365]
25. Zubair U, Zubair Z. Surgical resident training in Pakistan and benefits of simulation based training. J Pak Med Assoc 2020;70(5):904-908. [doi: 10.5455/JPMA.282116] [Medline: 32400750]

Abbreviations

AR: augmented reality
WoSCC: Web of Science Core Collection

Edited by L Tudor Car; submitted 27.01.25; peer-reviewed by G Huang, W Huang; comments to author 27.02.25; revised version received 06.03.25; accepted 07.03.25; published 26.03.25.

Please cite as:
Ba H, Zhang L, He X, Li S
Knowledge Mapping and Global Trends in Simulation in Medical Education: Bibliometric and Visual Analysis
JMIR Med Educ 2025;11:e71844
URL: https://mededu.jmir.org/2025/1/e71844
doi:10.2196/71844
PMID:

©Hongjun Ba, Lili Zhang, Xiufang He, Shujuan Li. Originally published in JMIR Medical Education (https://mededu.jmir.org), 26.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of an Interactive Web-Based Clinical Practice Monitoring System on Enhancing Motivation in Clinical Learning Among Undergraduate Nursing Students: Longitudinal Quasi-Experimental Study in Tanzania

Patricia Herman[1], MSc; Stephen M Kibusi[2], Prof Dr; Walter C Millanzi[3], PhD

[1]Department of Nursing, College of Health and Allied Sciences, Ruaha Catholic University, Iringa, United Republic of Tanzania

[2]Department of Public Health and Community Health Nursing, School of Nursing and Public Health, The University of Dodoma, Dodoma, United Republic of Tanzania

[3]Department of Nursing Management and Education, School of Nursing and Public Health, The University of Dodoma, Dodoma, United Republic of Tanzania

Corresponding Author:
Patricia Herman, MSc
Department of Nursing
College of Health and Allied Sciences
Ruaha Catholic University
Box 774
Iringa
United Republic of Tanzania
Phone: 255 788315184
Email: patriciaz1006@gmail.com

Abstract

Background: Nursing students' motivation in clinical learning is very important not only for their academic and professional achievement but also for making timely, informed, and appropriate decisions in providing quality and cost-effective care to people. However, the increased number of students and the scarcity of medical supplies, equipment, and patients, just to mention a few, have posed a challenge to educators in identifying and navigating the best approaches to motivate nursing students to learn during their clinical placements.

Objective: This study primarily used descriptive and analytical methods to examine undergraduate nursing students' desire for clinical learning both before and after participating in the program.

Methods: An uncontrolled longitudinal quasi-experimental study in a quantitative research approach was conducted from February to March 2021 among 589 undergraduate nursing students in Tanzania. Following a baseline evaluation, nursing students were enrolled in an interactive web-based clinical practice monitoring system by their program, institution, names, registration numbers, and emails via unique codes created by the lead investigator and trainers. The system recorded and generated feedback on attendance, clinical placement unit, selected or performed clinical nursing procedures, and in-between and end-of-shift feedback. The linear regression was used to assess the effect of the intervention (interactive web-based clinical practice monitoring system) controlled for other correlated factors on motivation in clinical learning (outcome) among nursing students. Nursing students' sociodemographic characteristics and levels of motivation in clinical learning were analyzed descriptively while a 2-tailed paired sample *t* test established a comparative mean difference in motivation in clinical learning between the pretest and the posttest. The association between variables was determined using regression analysis set at a 95% CI and 5% statistical significance.

Results: The mean age of study participants (N=589) was 23 (SD 2.69) years of which 383 (65.0%) were male. The estimated effect (β) of a 3-week intervention to improve nursing students' motivation in clinical learning was 3.041 (P=.03, 95% CI 1.022-7.732) when controlled for other co-related factors. The mean score for motivation in clinical learning increased significantly from the baseline (mean 9.31, SD 2.315) to the postintervention (mean 20.87, SD 5.504), and this improvement presented a large effect size of 2.743 (P<.001, 95% CI 1.011-4.107).

Conclusions: Findings suggest that an interactive web-based clinical practice monitoring system is viable and has the potential to improve undergraduate nursing students' motivation for clinical learning. One alternative clinical pedagogy that educators in

XSL•FO
RenderX

nursing education can use to facilitate clinical learning activities and develop motivated undergraduate nursing students is the integration of such technology throughout nursing curricula.

(JMIR Med Educ 2025;11:e45912) doi:10.2196/45912

KEYWORDS

clinical monitoring system; clinical practice; motivation in clinical learning; nursing students; smartphone; mobile phone; Ruaha Catholic University; web-based teaching

Introduction

Owing to changes in population, technology, and communicable and noncommunicable diseases, the uncertain nature of health care provision necessitates improvements in health care systems and the teaching and learning environment across health disciplines, including nursing [1]. Clinical nursing education, in particular, has become a fundamental aspect of the nursing profession that informs educators about the best and most innovative pedagogical strategies that are navigated into technology to increase nursing students' motivation in clinical learning and thus ensure consistent clinical attendance and passionate clinical learning around the world [2]. In clinical learning, motivation refers to an individual's inner drive toward an activity or behavior that defines his or her achievements, in this case, the attainment of skills and competence required in the nursing profession among nursing students [3]. Scholars have shown that motivated students can demonstrate metacognition, metacompetencies, and a sense of independence when providing high-quality, cost-effective health care to a diverse population [4-6].

Educators in nursing education are important individuals in both classroom and clinical teaching and learning activities to enhance nursing students' motivation in clinical learning during clinical placements for meta-competencies in diagnosing and making informed and appropriate decisions in providing quality and cost-effective care to people. Educators in nursing education are expected to be developed and empowered with pedagogical competencies to cope with advanced science and technology, increased rates of nursing student enrollments in middle and higher education institutions, scarcity of medical supplies and equipment, unimproved clinical teaching and learning environments, and availability of clients or patients as important individuals in clinical nursing education [7,8]. Empowering educators with knowledge and skills for facilitating clinical teaching and learning activities for nursing students may be intimately related to their pedagogical competencies in creating, supporting, mentoring, coaching, supervising, monitoring, and assessing nursing students [9,10].

Nursing students, developed by competent educators in nursing education, are believed to share their learning experiences with both educators and peers. They show motivation in their clinical learning activities and client or patient care provision as they demonstrate a sense of independence, a positive professional identity, and the attitude, skills, and values of lifelong learners and professionals [11]. Motivating nursing students to learn has risen to be prominent in clinical nursing education because it is regarded as a critical component in promoting consistent clinical attendance and learning [12]. Investing in the motivation of nursing students also ensures the development of competent nursing graduates who can demonstrate safe, ethical, and legal practices, which are foundational and essential aspects of clinical nursing education [6,13]. Contrary to what is expected of educators in nursing education in the twenty-first century, their current clinical nursing education practices are more traditional, with pedagogics such as bedside tutorials, lectures, demonstrations, discussions, case studies, portfolios, and nursing meetings, to name a few, being widely used [11,14,15]. However, due to the increased enrollment rate of nursing students with the unchanging pedagogical trend, a shortage of academic faculty, and a limited number of trained clinical instructors, the aforementioned clinical pedagogics are doubted to be able to enhance interactive communication between educators in nursing education and students.

Nevertheless, they are doubted on their abilities to establish teaching and learning feedback or experiences from students and nursing students' motivation in clinical learning [16]. Moreover, they demonstrate weaknesses in not developing them with clinical meta-competencies for providing quality and cost-effective care to people [17]. Scholars have linked the permanent implementation of conventional clinical nursing education pedagogics to a lack of interactive communication, teaching and learning feedback from students, and clinical absenteeism as remarkable signs of unmotivated nursing students worldwide [18]. The work by Rahman et al [12] has demonstrated that clinical absenteeism has been linked to an unsatisfactory clinical learning environment as well as a shortage of competent educators in nursing education. Nevertheless, the implementation of conventional clinical supervision, mentorships, support, monitoring, and evaluation measures such as registration books and follow-up books fail to manage a large group of nursing students in clinical settings [19]. Nursing students' avoidance and lack of enthusiasm in attending their daily practical activities during their clinical placements result in substandard care delivered to clients or patients alongside unethical professional conduct and poor customer care [20].

This work is based on the belief that understanding and implementing novel clinical pedagogical strategies will assist educators in nursing education in grasping and navigating the best ways to inspire unmotivated students to learn in a clinical context over conventional pedagogies. Adoption and integration of technology have been prioritized and have proven to be timely, quick, cost-efficient, long-term, and beneficial in enhancing students' motivation in their learning activities [21,22]. Authors of this work agree with other scholars that it appears to be timely for clinical nursing education to transform its pedagogics to technology-based ones to fill educational gaps demonstrated by conventional pedagogics for enhancing nursing

XSL•FO
RenderX

students' motivation in clinical learning, particularly in low- and middle-income countries such as Tanzania [23]. As it has worked elsewhere, web-based learning is currently becoming an increasingly vital instructional tool in nursing education, as it offers the potential to promote motivation to learn among students [24].

Scholars including Mico et al [25] have highlighted that the web-based learning approach, which has been endorsed as an essential educational tool, responds effectively and efficiently to nurses' needs and experiences in their clinical practices. Web-based technology in education is linked with a networking interactive model to promote communication with a feedback mechanism to enhance active learning between the users [26]. However, little about the integration of technology in clinical nursing education has been documented in Tanzania to mentor, supervise, support, monitor, and evaluate nursing students during their clinical placement [27]. If the situation remains unattended, nursing students will continue to be developed conventionally and continue to be unmotivated in their clinical learning, gaining little clinical competencies to work independently and confidently to deliver ethical, quality, and cost-effective care to people.

Therefore, this study intended to fill the gap by primarily using descriptive and analytical methods to examine undergraduate nursing students' desire for clinical learning both before and after participating in the program.

Methods

This study was conducted by taking into consideration international and national research standards and ethics. Moreover, it was informed by the institutional postgraduate guidelines and regulations of The University of Dodoma [28].

Study Location

The study was conducted in Dodoma Regional Referral Teaching Hospital in Tanzania, which accommodates a large number of nursing students from different nursing training institutions within Dodoma region. Some previous scholars [11,29,30] suggest that proximity to the research environment where the intervention is being piloted or implemented aids in receiving rapid input from the consulted experts, trainers, and participants, and ensuring periodic monitoring of its integrity. Aside from the availability of nursing students, the location was chosen because of the availability and accessibility of the consulted experts and trainers for their evaluation and appraisal of the web-based clinical monitoring system throughout the design and piloting procedures.

Study Design and Approach

As it has also been used by some previous scholarly works [31], this study used an uncontrolled longitudinal quasi-experimental design (pre-post tests) with a quantitative research approach among 589 randomly selected undergraduate nursing students in Tanzania from February 2021 to March 2021. The study began with a screening method to choose individuals who satisfied the inclusion criteria and were willing to engage in the study, followed by a baseline assessment to determine their initial level of motivation in clinical learning. Following the intervention, the same participants were exposed to the system for 3 weeks, with 1 week set apart for the end line evaluation (posttest) after the intervention as a follow-up assessment.

Study Population

To maximize the diversities of an intervention's effects, the study recruited undergraduate nursing students in diploma and bachelor's degree programs from 2 middle and 2 higher training institutions in the Dodoma region of Tanzania's central region.

Sample Size Determination

The minimum sample size of this study was determined based on the findings from previous studies [5]. Their findings revealed a baseline mean score of knowledge about how to plan learning activities of 53.10, whereas the end line mean score was 54.21. The following formula was used to determine the minimum sample size as suggested by previous studies [29,31] to be used when researchers wish to conduct an uncontrolled quasi-experimental study design:

where n=minimum sample size

$Z\alpha$: Tabulated Z value set at 95% (1.96) CI

$Z\beta$: Tabulated Z value set at 80% (0.84) power to demonstrate a statistical difference between pretest and posttest.

σ: Polled SD = 7.511399669835177

SD_1: SD 1 (from previous studies = 10.21)

SD_2: SD 2 (from previous = 11.02)

δ: Mean difference $(M_2 - M_1)^2 = 0.11$

M_1: Mean 1 (from previous studies = 53.10)

M_2: Mean 2 (from previous studies = 53.21)

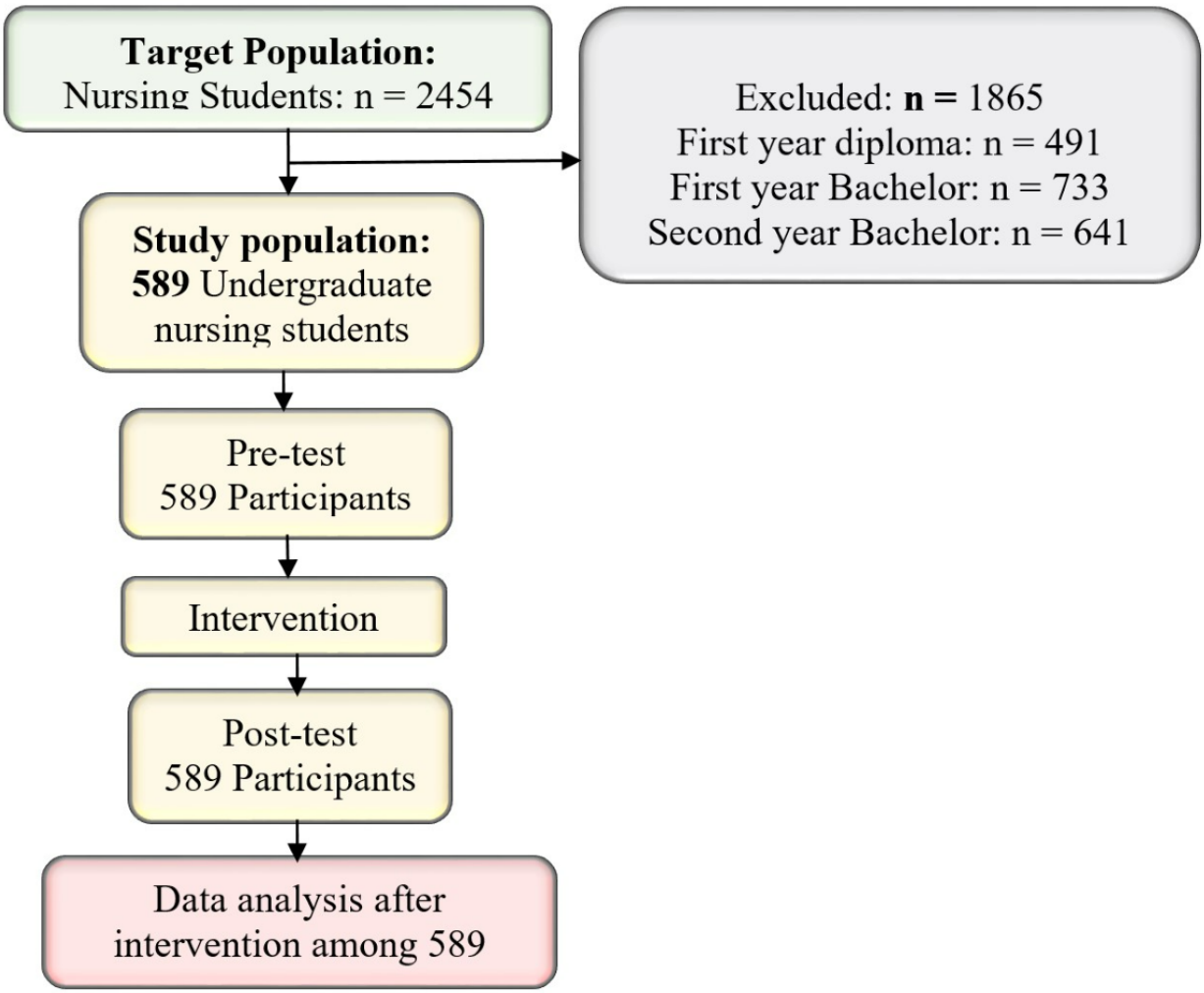
With the addition of a 10% attrition adjustment of the calculated sample size (n=54) = (535+54) = 589. Therefore, the minimum sample size of this study was 589 nursing students.

Recruitment Procedures of the Study Participants

The framework of recruiting study participants has been benchmarked from some previous scholarly works [11,32-37]. As shown in Figure 1, 2454 nursing students were eligible to join the study. However, 589 nursing students met the inclusion criteria and participated in the study, assessed at baseline, at end line, and their data analyzed. Sums (1865/2454, 76%) were excluded due to various reasons including not having started clinical placements (first-year diploma in nursing [n=491], a first-year bachelor of science in nursing [n=733], and a second-year bachelor of science in nursing [n=641]). There was no loss to follow-up among nursing students who joined the study and thus the completion rate was 100%.

XSL•FO
RenderX

Figure 1. A flow pattern of sampling procedure among nursing students. From field data (2021).



Eligibility Criteria

Inclusion Criteria

Nursing students were recruited for this study based on their willingness after being informed about the purpose, benefits, and drawbacks of participating in this study. The study recruited second-, third-, and fourth-year undergraduate nursing students who were in clinical placements at the time of this study. In addition, undergraduate nursing students with Information and Communication Technology literacy and those with iPhones or iPads were recruited for the study.

Exclusion Criteria

Nursing students who reported being unwell and unable to converse or participate in the study were excluded from the study. Students without institutional registration numbers, those unable to use computers or smartphones, and nursing students recruited for other studies or projects were not eligible to participate in this study.

Sampling Procedures

As recommended by previous scholars [38-41], probability sampling techniques through multistage sampling methods were used to reach and study nursing students in this study. Stage 1: a simple random sampling technique by lottery method was used to select regions and districts. Stage 2: As shown in Tables 1 and 2, stratified sampling methods were used to select higher training institutions (institutions A and B). On the other hand, lower training institutions (institutions C and D) were randomly selected using a stratified sampling technique.

Nevertheless, a stratified random sampling technique was used to select the nursing program (diploma in nursing and bachelor of science in nursing) and classes (who are in their clinical placement). The training health facility was selected purposely because it is only a major regional referral hospital in the Dodoma region used by multiple training institutions for training nursing students in clinical settings. Stage 3: systematic sampling methods were used to select the study participants. As shown in Tables 1 and 2, a proportionate formula: $n=[P_i \times (n/p)]$ was used to determine a minimum sample size of nursing students per training institution and their program, respectively, whereas n_i is the proportional sample size, P_i is the total targeted population, n is the number of nursing students, and P is the minimum sample size of the study.

XSL•FO
RenderX

Table 1. Proportionate sample size by the sampled training institutions[a].

Name of training institution	Total population, n	Proportionate, n
Institution A	968	232
Institution B	998	239
Institution C	367	88
Institution D	121	30

[a]From Study plan (2021). Total number of nursing students among the sampled training institutions = 2454 ($P_i \times n/P$).

Table 2. Proportionate sample size by training nursing program[a].

Training institution	Nursing program	Total population, n	Proportionate ($P_i \times n/P$)
Institution A	Diploma in nursing	230	55
Institution A	Bachelor in nursing	738	177
Institution B	Bachelor in nursing	828	198
Institution B	Diploma in nursing	171	41
Institution C	Diploma in nursing	367	88
Institution D	Diploma in nursing	121	30

[a]From Study plan (2021).

Intervention

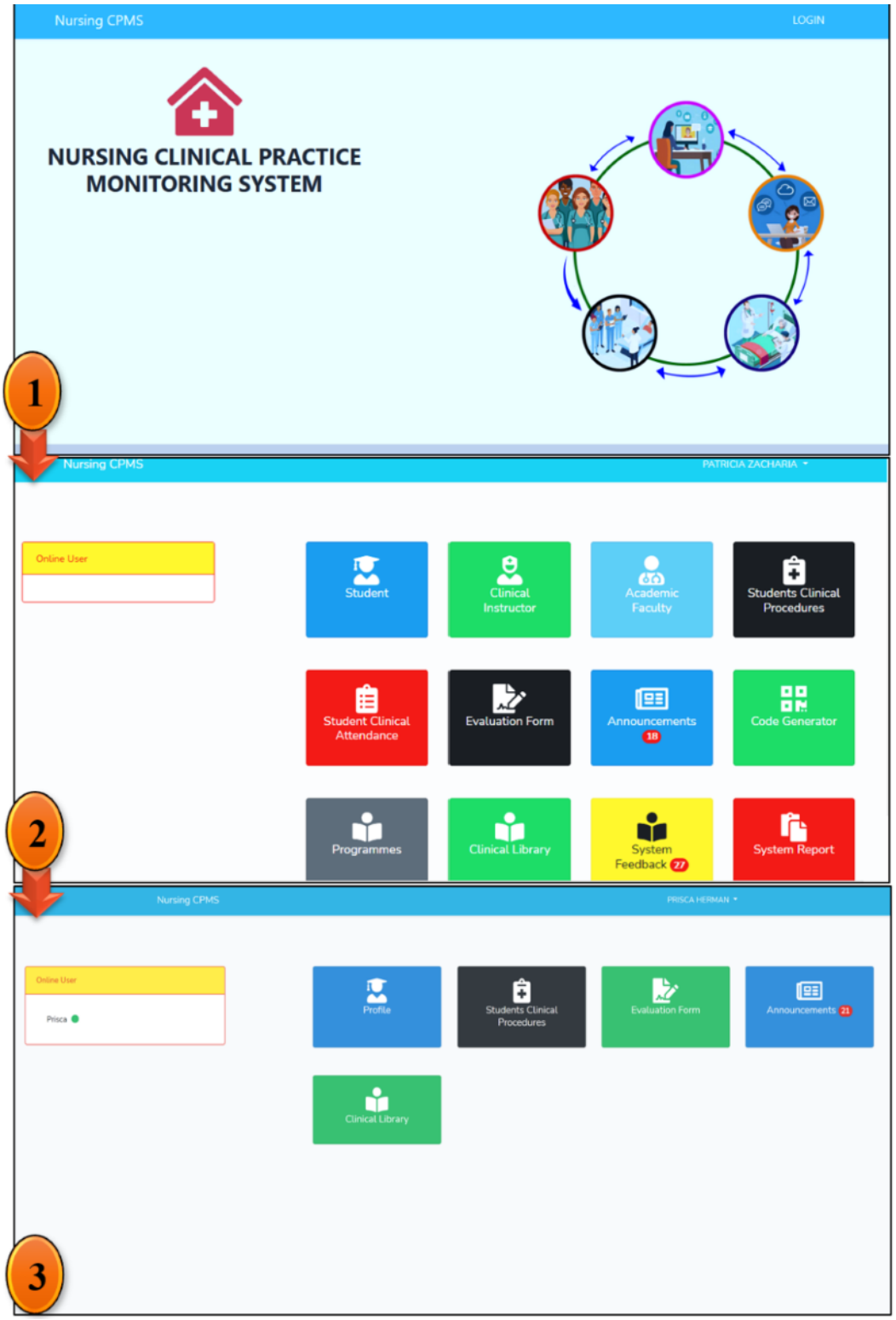
The intervention was completed in 3 weeks, with 1 week set aside for the final evaluation (posttest). It was carried out on a daily basis (6 hours a day), per the clinical rotation and placement schedules established by the nursing students' respective training institutions. Within the 3 weeks of the intervention, all students were expected to perform and complete the assigned clinical learning task or activities and models on a daily basis and within 6 hours of their duty shifts. The primary aim of the intervention was to descriptively and analytically measure a change in motivation in clinical learning among undergraduate nursing students before and after participating in the program. Trained research trainers who also had clinical nursing education competence implemented the intervention. As illustrated in Figure 2, the system included a welcoming window as well as several menus or nodes such as clinical instructors, academic staff, clinical nursing procedures, clinical attendance, evaluation forms, announcements, code generator, professional programs, clinical library, system feedbacks and reports, and students' node, which included students' profiles, clinical notes, attendance, evaluation forms, and announcement menus.

Referring to Figure 3, nursing students were required to arrive at their shift very early each day before the tuned time in the system, which was 7:30 AM (East African time), to be assigned codes of a specific day or date generated by the trainers. The generated codes allowed them to log in to the system and sign out in the presence of the trainer at the end of the shift, which was tuned at 1:30 PM (East African time). Any student who arrived late for his or her shift and who did not provide web-based feedback on his or her performance of the chosen clinical nursing procedure in the presence of clinical instructors or trainers in between shifts, or left the shift before the fixed time of signing out was considered absent in that particular shift. The system provided a daily shift attendance report that included the dates, students' names and registration numbers, department, ward or unit, duty shift, clinical nursing procedure executed, and evaluation score for each student.

Nonetheless, as shown in Figure 4, several nursing clinical procedures were imported into the system, including giving and receiving reports; patients admission; changing patients' positions on a bed; counseling; bed making; administration of intravenous, intermuscular, and oral medications; catheterization; kangaroo mother care; administration of oxygen to patients (oxygen therapy); cardiopulmonary resuscitation; management of patients with preeclampsia or eclampsia; blood transfusion; per-vagina examination; wound dressing; mouth care; management of postpartum hemorrhage; assessment of placenta; management of pregnant women in a first and second stage of labor; assessment of new-born babies; history taking during labor; and vital sign assessment. After selecting a clinical nursing procedure for a particular duty shift, nursing students had to adhere to the 6 stages of conducting it including identification of the patient alongside getting informed consent; preparation of the environment; one-self; equipment appropriate for the chosen procedure; and performing and finishing the procedure. Each clinical nursing procedure was featured in several activities, which nursing students were supposed to follow, adhere, and implement chronologically.

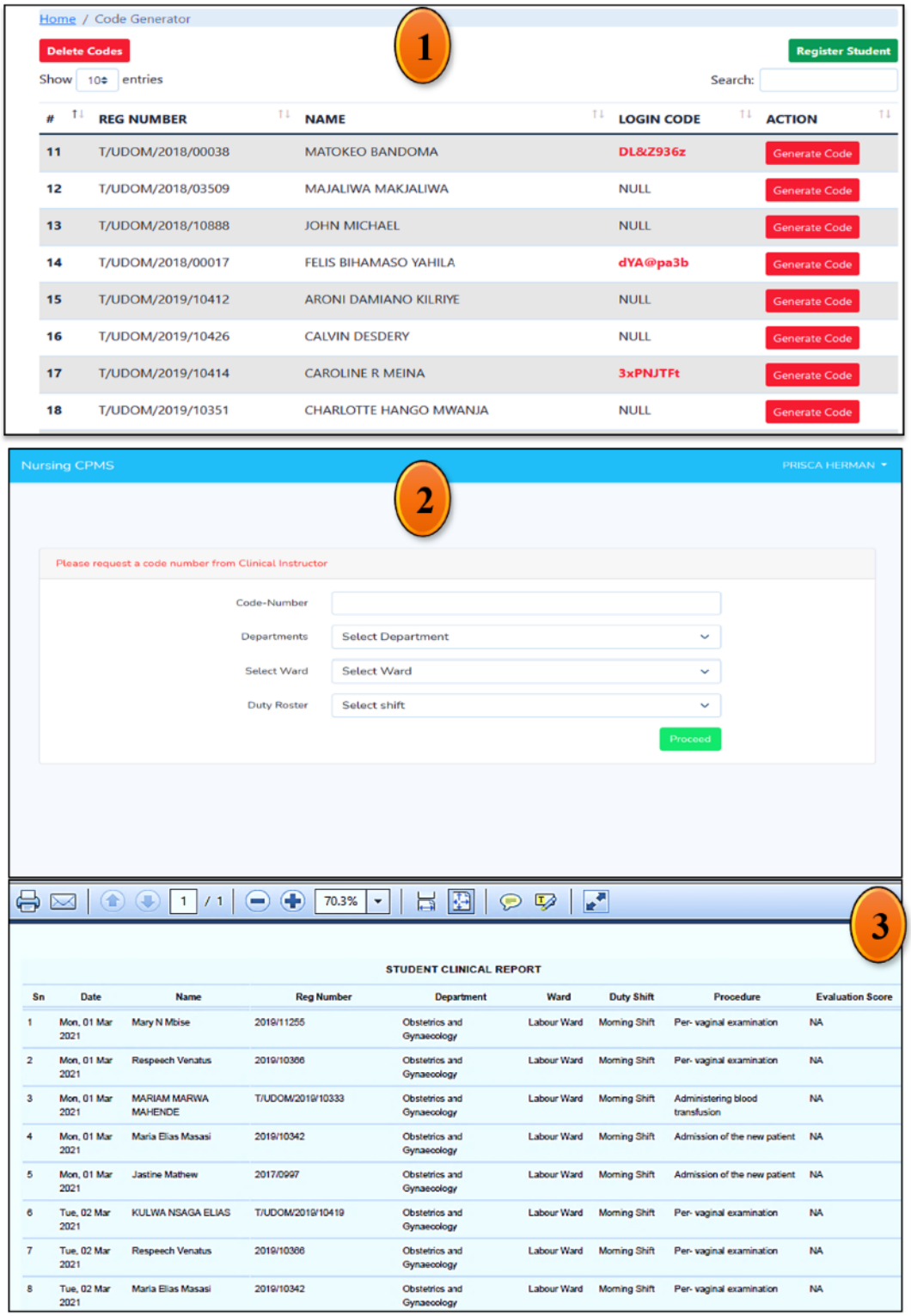
XSL•FO
RenderX

Figure 2. A welcoming window, menu window, and nursing students' node with various menus in the system. From field data (2021).



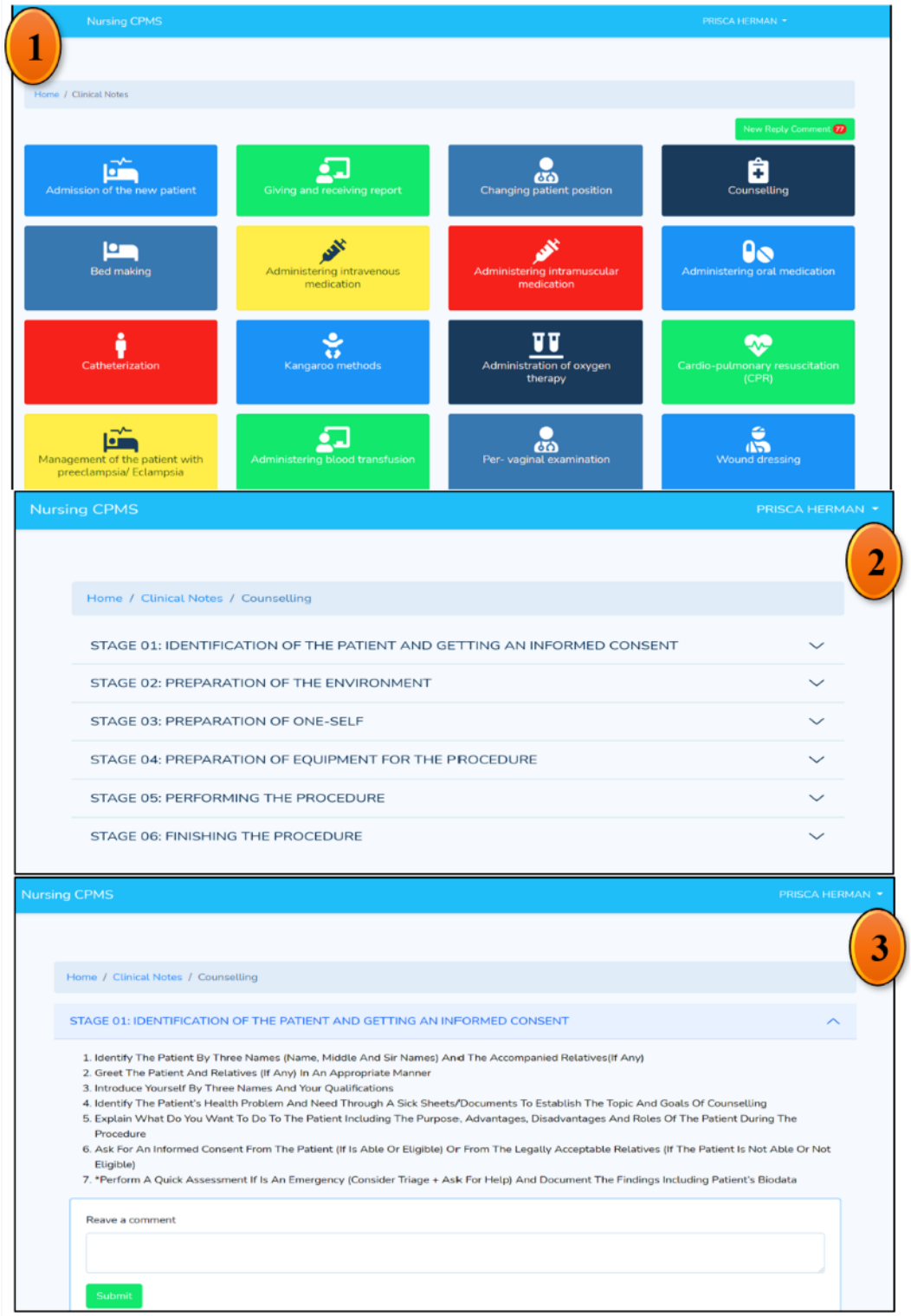
XSL•FO
RenderX

Figure 3. Code generator window, nursing students' login window, and an example of a system-generated report. From field data (2021).



XSL•FO
RenderX

Figure 4. Variety of clinical nursing procedures, stages, and activities imported into the system for nursing students. From field data (2021).



Procedures for Getting Nursing Students Into the System, Supervising, Supporting, Monitoring, and Evaluating Them

Undergraduate nursing students were recruited based on their units of clinical placements including the emergency unit, labor and postnatal ward, pediatric ward, and male medical ward within Dodoma Regional Referral Hospital. Resources such as computers, smartphones or iPads, electricity, web connectivity, clinical nursing procedure checklists, duty rosters, notebooks, pens, and shift objectives were the requirements for the intervention. Nursing students' institutional registration numbers were used as username identities during the login procedures into the system.

Clinical nursing procedures and evaluation checklists were imported into the system during the intervention to allow nursing students and clinical instructors to access them and have an interactive room to identify and discuss them before starting to

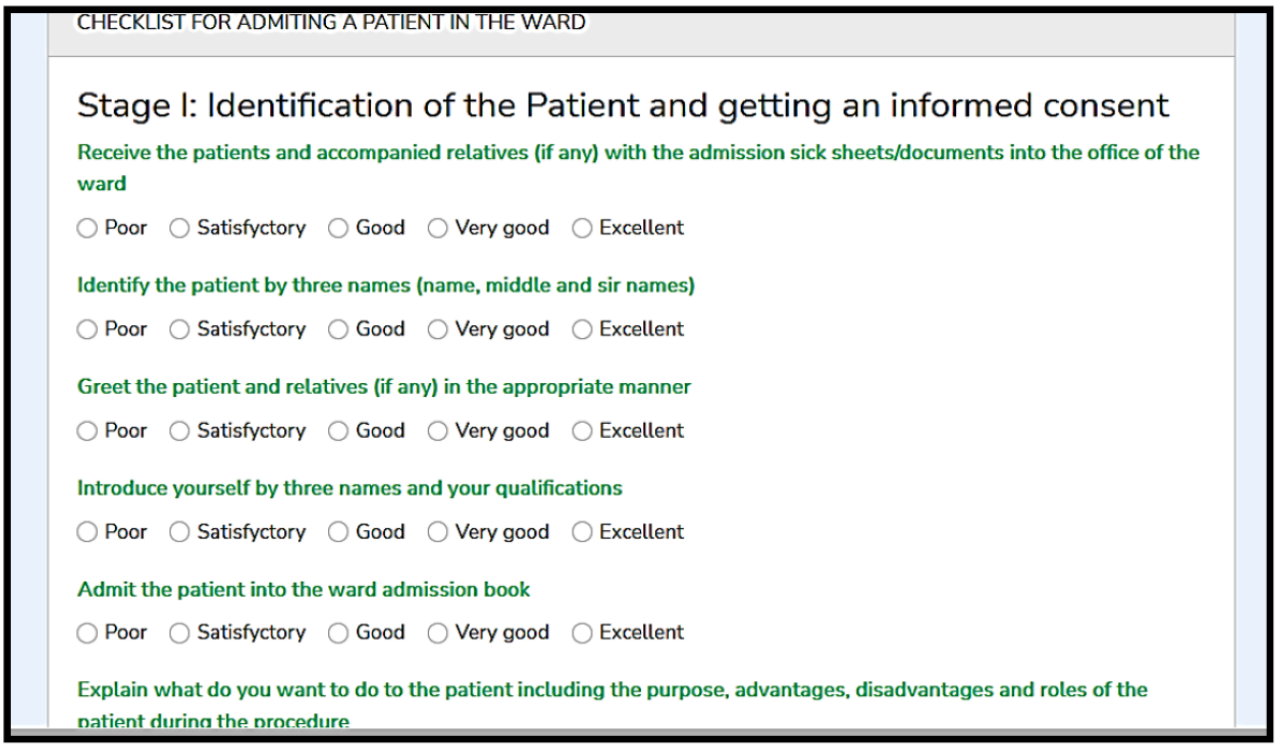
XSL•FO
RenderX

care for the patients on that particular day. Nursing students had to ask for a system-generated code in the presence of a (system implementers) clinical instructor to be able to log in to the system and counted to have reported timely in a particular duty shift as the mode of monitoring and keeping clinical attendance reports among them. Furthermore, after the successful registration of nursing students into the system, system implementers had to confirm the presence of nursing students throughout the specified duty shift.

A shift progress Min-reports (brief or summary report of what a student has done for a particular time of a shift) was posted by nursing students in the system including patients' progress after performing the procedure. On the other hand, trained research trainers had the role of periodically posting announcements into the system to enhance interactive and reciprocal communication among nursing students and academic faculty. They also had the role of summing up nursing students' clinical signs of progress and filling the evaluation forms at the end of a duty shift. All clinical teaching and learning activities performed by the trained research trainers and nursing students were tracked, monitored, and sometimes addressed by the principal investigator (PZH) through a system WebTop. All self-rated evaluations on the performed clinical procedures, day feedback, and experiences of the interactive web-based clinical practice monitoring system among nursing students and system implementers were captured through the system before they signed out at the end of the duty shift.

Only 1 session (morning duty shift) was preferred for the intervention as negotiated by hospital administration and system implementers. The intervention was implemented for 6 hours equivalent to 1 duty shift in a day with a duration ranging from 7:30 AM to 1:30 PM (East African time) among nursing students. Each student had to carry out and finish one of the aforementioned clinical nursing procedures every day for 3 weeks for learning to take place. This amounted to about 20 clinical nursing procedures in total. All students finished the same clinical nursing procedures during the intervention's 3-week duration including giving and receiving reports; patients admission; changing patients' positions on a bed; counseling; bed making; administration of intravenous, intermuscular, and oral medications; catheterization; kangaroo mother care; administration of oxygen to patients (oxygen therapy); cardiopulmonary resuscitation; management of patients with preeclampsia or eclampsia; blood transfusion; per-vagina examination; wound dressing; mouth care; management of postpartum hemorrhage; assessment of placenta; management of pregnant women in first and second stages of labor; assessment of newborn babies; history taking during labor; and vital sign assessment. Assessment of nursing students' motivation in clinical learning was performed using academic motivation questionnaires that were administered to them as pretests and posttests. As shown in Figure 5, the evaluation of the system was performed through nursing students' and system implementers' experiences feedback inventory that was posted in the system and filled out before signing out of the system.

Figure 5. An example of a self-rated evaluation form imported into the system for nursing students. From field data (2021).



Strategies Used to Maintain Adherence to the Intervention Among Nursing Students

This section describes the measures used to increase nursing students' adherence during intervention implementation in the field. Throughout the intervention, nursing students' registration numbers were used as identifiers during system deployment. Being unidentified and acknowledged by name would most likely make nursing students feel at ease and confident in their clinical learning privacy, allowing them to feel free and inspired

XSL•FO
RenderX

to follow the intervention regimen. Furthermore, with the assistance of academic faculty and clinical instructors, the study was carried out under the training institutional clinical rotation schedule, which most likely aided this study in ensuring that nursing students adhered to the intervention because they had to attend their shifts accordingly. The use of technology to assist clinical learning among nursing students appeared to impress them, particularly the ways interactive communications and physical contact were assured to them. Nursing students and system implementers have enhanced adherence to the intervention among nursing students by posting various announcements, peer teaching and learning, and self-evaluation forms.

Data Collection Procedures

The focus of the current effort was mostly on presenting the study's quantitative findings. Quantitative data were gathered at two-time intervals, baseline, and end line (directly postintervention), in different unoccupied rooms, as agreed upon by the administrative authority of the respective training institutes. The research assistants were chosen on purpose based on their eligibility criteria, which included at least 1 year of data-gathering experience. Nursing students were given a brief explanation of completing the questionnaire before having copies provided by research assistants. The researcher and assistants were present throughout the process to supervise and address the overstretched immediate and long-term problems before collecting all copies of completed surveys and protecting them as part of nursing students' confidentially. Before leaving the room, nursing students were informed of the timetable and method for intervention.

Research Tools and Instruments

The data collection tool for the quantitative part of the work consisted of 2 parts: the sociodemographic characteristics profiles and the part that measured nursing students' motivation in clinical learning. Gender, age, marital status, interest, motive for joining the nursing profession, accommodation, and marital status were all part of the sociodemographic profile. The study used a 5-point Likert scale to assess motivation in clinical learning using an Academic Motivation Scale composed of 28 items adopted from previous scholarly works [5,42-44]. The 5-point Likert scale ranged from 1=strongly disagree to 5=strongly agree. Its subscales included intrinsic motivation which was assessed by 12 items, extrinsic motivation (12 items), and amotivation (4 items) of which the findings were dichotomized into an "Agree" (assigned a value of "1") response that described the action to have been performed by the participant; otherwise, "Disagree" ("0" value) for unperformed action or behavior.

Before data collection, the principal component analysis was performed at a measure of Keiser-Meyer-Olkin and Bartlett test value of >0.5 and a significance level of ≤5% to measure the weight of each item. The findings revealed that all 28 items scored >0.3, and thus, they were all retained for data collection. Scoring the variable of motivation in clinical learning was adopted in a study conducted by Millanzi and Kibusi [6] that nursing students who scored 0-16 were categorized into low motivation in clinical, those who scored 17-24 had moderate learning motivation, and nursing students who scored 25 and above demonstrate high motivation in clinical learning.

Validity and Reliability of the Study

The validity and reliability tests were performed first before subjecting the tool to the actual field for data collection. The tool was shared with the subject matter and statistician for suitability of the items, reliability, and ambiguity to fit for knowledge to undergraduate nursing students. The pretest of research tools was conducted among 60 nursing students at a location that was different from the sampled study setting. The finding of a pretest was subjected to the exploratory factor analysis to determine the weight of each item to assess the outcome of interest and the reliability scale analysis to determine the internal consistency of the tool and presented using Cronbach α values as recommended by previous studies [45]. The findings of the scale analysis on motivation in clinical learning were found to be Cronbachα=0.840 (>0.7), which was statistically reliable for the actual data collection.

Data Analysis

An SPSS software program (version 23; IBM Corp) was used to analyze data. Before data analysis, cleaning was done to ensure the completeness, accuracy, and clarity of the information in the questionnaires. A normality test was performed to determine the distribution of data to opt for the mode of data analysis. Findings of the normality test revealed that data motivation in clinical learning was approximately normally distributed, and thus, parametric statistical measurements were adopted. Descriptive statistic was used to establish participants' sociodemographic characteristics profiles of the study participants such as age, sex, program, and year of study, just to mention a few, and the motivation in clinical learning. For inferential statistics, a 2-tailed paired *t* test was performed to determine a comparative mean score change and difference in motivation in clinical learning among nursing students between the pretest and the posttest.

The multiple linear regression analysis model by considering the control of other factors as independent variables such as sociodemographic characteristics profiles and the intervention (interactive web-based clinical practice monitoring system) was performed to establish the extent of association with the outcome variable of interest (motivation in clinical learning). The multiple linear regression was opted for because several factors were controlled during analysis and the outcome variable of interest was treated as a scale variable. The goal of this study was to determine the net effect of the intervention by taking into account and controlling for the sociodemographic characteristic profiles of nursing students, as it was hypothesized that these profiles would also have an impact on the outcome of interest (motivation in clinical learning) over the intervention. Findings of inferential statistics were presented in tabular forms by means, of SDs, *t* values, significance level (<5%), and 95% CI. The following logistic regression model was used:

XSL•FO
RenderX

where Y is dependent variable; a is intercept; b, c, d, and e are slopes; X_1, X_2, X_3, and X_4 are independent (explanatory) variables; and e is residual (error).

Ethical Considerations

This study conformed to The University of Dodoma institution's postgraduate guidelines and regulations after being approved and given an ethical clearance (number MA.84/261/02/81) by the institutional research ethics review committee. All participants provided written informed consent after being explained about the study and their freedom to participate in it. Data collection procedures were performed in separate and unoccupied venues that were available in the respective institution premises. Participants' names were not included in the data collection tools and their information was secured by the principal investigator (PZH) using folders with passwords. Given that participants were in their academic clinical rotations calendar for clinical learning activities, there were no compensations of either time or monetary incentives throughout the study.

Results

Proportional Distribution of Nursing Students by Their Sociodemographic Characteristics Profiles

The completion rate of the study was 100% of the studied participants. Findings in Table 3 indicate that 65.0% (383/589) of nursing students were males while 79.6% (469/589) of the sample were younger than 24 years, with a mean age of 23 (SD 2.689, range 19-50) years. Accommodated participants in their respective training institutions' hostels accounted for 71.5% (421/589) while 63.7% (375/589) of them were enrolled in bachelor of science in nursing and 33.6% (198/589) and 30.1% (177/589) of them were in their fourth and third year of studies, respectively. A majority of nursing students 69.4% (406/589) were not interested in joining nursing programs. However, those who were interested in joining nursing education were driven by a belief that it is a secure profession (567/589, 96.3%), caring to save peoples' lives (491/589, 83.4%), autonomy to practice (478/589, 81.2%), and generous salary and employment benefits (438/589, 74.4%). Other findings were found as shown in the table.

XSL•FO
RenderX

Table 3. Proportional distribution of nursing students by their sociodemographic characteristics (n=589)[a].

Variable	Values
Age (years), mean (SD; range)	23 (2.689; 19-50)
Age (years), n (%)	
<24	469 (79.6)
25-34	115 (19.5)
>35	5 (0.9)
Institution, n (%)	
Training institution A	232 (39.4)
Training institution B	239 (40.6)
Training institution C	88 (14.9)
Training institution D	30 (5.1)
Sex, n (%)	
Male	383 (65.0)
Female	206 (35.0)
Marital status, n (%)	
Single	543 (92.2)
Married	46 (7.8)
Accommodation, n (%)	
In-campus	421 (71.5)
Off-campus	168 (28.5)
Program of study, n (%)	
Diploma in nursing and midwifery	214 (36.3)
Bachelor of science in nursing	375 (63.7)
Year of study , n (%)	
Second-year diploma in nursing	89 (15.1)
Third-year diploma in nursing	125 (21,2)
Third-year bachelor of science in nursing	177 (30.1)
Fourth-year bachelor of science in nursing	198 (33.6)
Interested to join the nursing profession, n (%)	
No	409 (69.4)
Yes	180 (30.6)
Reason to join the nursing profession	
Generously salary and employment benefits, n (%)	
Yes	438 (74.4)
No	151 (25.6)
A secured profession, n (%)	
Yes	567 (96.3)
No	22 (3.7)
Autonomy to practice, n (%)	
No	478 (81.2)
Yes	111 (18.8)
Caring to save people's lives, n (%)	
No	491 (83.4)

XSL•FO
RenderX

Variable	Values
Yes	98 (16.6)
Opportunity to travel worldwide, n (%)	
Yes	421 (71.5)
No	168 (28.5)
Job availability, n (%)	
Yes	398 (67.5)
No	191 (32.5)

[a]From field data (2021).

The Overall Distribution of the Level of Motivation in Clinical Learning and Their Domains Among Nursing Students

The findings from Table 4 revealed that there was no significant difference between the proportion of nursing students with low and moderate motivation in clinical learning (261/589, 44.3%; and 328/589, 55.7%, respectively). However, baseline findings indicated that none of the nursing students demonstrated high motivation in clinical learning 0.0% (n=0). On the other hand, baseline findings of the motivation domains indicated that 94.7% (558/589) of nursing students were not intrinsically motivated in clinical learning contrary to the end line findings, which indicated that 90.5% (533/589) demonstrated inner motive in clinical learning. Highly motivated nursing students to learn in clinical settings accounted for 67.7% (399/589) while only 4.9% (29/589) of them demonstrated lower motivation in clinical learning. Other findings were observed as shown in Table 4.

Table 4. Overall distribution of the level of motivation in clinical learning and their domains among nursing students in the Dodoma region (N=589)[a].

Variable	Pretest	Posttest
Overall motivation in clinical learning, n (%)		
High learning motivation	0 (0)	399 (67.7)
Moderate learning motivation	328 (55.7)	160 (27.2)
Low learning motivation	261 (44.3)	29 (4.9)
Motivation subscales		
Intrinsic motivation in clinical learning, n (%)		
No	558 (94.7)	56 (9.5)
Yes	31 (5.3)	533 (90.5)
Extrinsic motivation in clinical learning, n (%)		
No	506 (85.9)	106 (18.0)
Yes	83 (14.1)	482 (81.8)
Amotivation, n (%)		
Yes	475 (80.6)	313 (53.1)
No	114 (19.4)	276 (46.9)

[a]Field data (2021).

Overall Mean Score Change and Mean Difference in Motivation in Clinical Learning Between Pretest and Posttest Among Nursing Students

As shown in Table 5, there was a statistically significant increase in mean scores changes of motivation in clinical learning from mean 9.31 (SD 2.315) at baseline to mean 20.87 (SD 5.504) at the end line. A comparative analysis of motivation performance among nursing students between pretest and posttest was found to be statistically significant (mean 11.566, SD 5.667; t_{588}=49.496; P<.001; 95% CI 11.107-12.025). The findings suggest that nursing students scored high on motivation in clinical learning in the posttest as compared with the pretest. Moreover, the findings in Table 5 indicated that there was an increase in mean scores among nursing students per domain of motivation to clinical learning between the pretest (mean 3.74, SD 1.231) and the posttest (mean 9.53, SD 2.762).

XSL•FO
RenderX

Table 5. Overall mean score change and mean difference in motivation in clinical learning between pretest and posttest among nursing students (N=589)[a].

	Pretest, mean (SD)	Posttest, mean (SD)	Mean difference, mean (SD)	*t* test (*df*)	*P* value	95% CI	Effect size (Cohen *d*)	95% CI
Motivation in clinical learning	9.31 (2.315)	20.87 (5.504)	11.56 (5.667)	49.496 (588)	.001	11.107-12.025	2.743	1.011-4.107
Domains of motivation in clinical learning								
Intrinsic	3.74 (1.231)	9.53 (2.762)	5.800 (2.968)	47.421 (588)	.001	5.559-6.040	N/A[b]	N/A
Extrinsic	4.49 (1.42)	8.77 (3.325)	4.276 (3.474)	29.845 (588)	.001	3.994-4.557	N/A	N/A
Amotivation	2.57 (1.1871)	1.08 (1.392)	1.492 (1.644)	22.031 (588)	.001	1.359-1.625	N/A	N/A

[a]From field data (2021).

[b]N/A: not applicable.

Moreover, they also demonstrate higher scores in their extrinsic motivation to learning in clinical settings at the end line (mean 8.77, SD 3.325) than at baseline (mean 4.49, SD 1.42) while amotivation performance decreased from mean 2.57 (SD 1.187) at baseline to mean 1.08 (SD 1.392) at the end line. The effect size of the intervention on motivation in clinical learning among nursing students was computed using Cohen *d* formula (mean 2 minus mean 1 divided by a pooled SD). Findings showed that the intervention demonstrated an effect size of 2.74 ($P<.001$; 95% CI 1.011-4.107), which is a high effect size based on Cohens *d* classifications of effect sizes [46].

The Estimated Effect of an Intervention (Interactive Web-Based Clinical Practice Monitoring System) Controlled for Other Co-Related Factors on Motivation in Clinical Learning Among Nursing Students at Posttest

About 58.5% variation in motivation in clinical learning scores is explained by the explanatory variables included in the model. The overall model was statistically significant ($f=25.6$; $P=.001$). Findings in Table 6 indicate that reasons to join the nursing profession such as due to the opportunity to demonstrate autonomy (β=1.590; $P=.02$; 95% CI 0.279-3.901), the opportunity to travel around the world (β=1.648; $P=.04$; 95% CI 0.583-4.713), job availability (β=1.409; $P=.001$; 95% CI 1.046-5.772), and other corelated factors were statistically significantly associated with motivation in clinical learning among nursing students against their counterparts. The estimated effect (β) of a 3-week intervention to improve nursing students' motivation in clinical learning was 3.041 ($P=.03$, 95% CI 1.022-7.732) when controlled for other correlated factors.

XSL•FO
RenderX

Table 6. The estimated effect of an intervention (interactive web-based clinical practice monitoring system) controlled for other corelated factors on motivation in clinical learning among nursing students at posttest (N=589)[a,b].

Variable	Estimate (β)	SE	*P* value	95% CI
Intervention				
Pretest	1	N/A[c]	N/A	N/A
Posttest	3.041	0.308	.03[d]	1.022-7.732
Institutions	N/A	N/A	N/A	N/A
Institution A	1	N/A	N/A	N/A
Institution B	.729	0.883	.23	1.956-0.467
Institution C	.932	0.794	.65	6.052-3.753
Institution D	.312	1.117	.12	0.194-1.757
Programs				
Diploma	1	N/A	N/A	N/A
Bachelor	.281	0.483	.56	1.229-0.668
Year of study				
Second-year diploma	1	N/A	N/A	N/A
Third-year diploma	.593	0.809	.36	1.806-0.655
Third-year bachelor	.936	0.926	.71	5.315-3.611
Fourth-year bachelor	.744	0.617	.01[d]	0.431-3.417
Age group (years)				
<24	1	N/A	N/A	N/A
24-34	1.149	2.496	.49	0.669-1.407
>35	.782	0.497	.42	1.827-0.768
Sex				
Female	1	N/A	N/A	N/A
Male	.434	0.905	.58	0.851-1.512
Marital status				
Single	1	N/A	N/A	N/A
Married	.575	0.627	.39	1.184-3.007
Accommodation				
In-campus	1	N/A	N/A	N/A
Off-campus	.852	2.272	.73	1.981-1.385
Interested to join the nursing profession				
No	1	N/A	N/A	N/A
Yes	.250	0.043	.051[d]	0.335-0.165
Reason to join nursing				
Autonomy to practice				
Yes	1.590	0.667	.02[d]	0.279-3.901
No	1	N/A	N/A	N/A
Caring patients				
Yes	1.107	0.679	.10	0.226-2.441
No	1	N/A	N/A	N/A
Opportunity to travel around worldwide				

XSL•FO
RenderX

Variable	Estimate (β)	SE	*P* value	95% CI
Yes	1.648	1.512	.04[d]	0.583-4.713
No	1	N/A	N/A	N/A
It is the secured profession				
Yes	.563	0.551	.31	0.519-1.645
No	1	N/A	N/A	N/A
Reasonable payment				
Yes	.033	0.538	.95	1.089-1.023
No	1	N/A	N/A	N/A
Job availability				
Yes	1.409	0.694	.001[d]	1.046-5.772
No	1	N/A	N/A	N/A

[a]From field data (2021).

[b]R^2=0.865, f=116; P<.001, significant at P<.05, and significant at P<.001.

[c]N/A: not applicable.

[d]Variables that are significantly associated with the outcome variable.

Discussion

Principal Findings

In terms of the study's focus and objective, the implementation of an interactive web-based clinical practice monitoring system for nursing students' motivation in clinical learning was feasible and practical in a clinical setting with consistent electricity supply and web connectivity. Nursing students indicated moderate to high levels of motivation in clinical learning after 3 weeks of system implementation, compared with when they were not exposed to it. Nursing students demonstrated a capacity to plan, identify, and access academic resources and help, as well as participate in clinical practices with minimal support from clinical instructors, trainers, and academic faculty, according to the end line assessment. Nonetheless, contrary to the existing practices where nursing students are not allowed formally to use electronic devices in clinical settings, the use of electronic devices while students are in clinical settings such as smartphones, iPads, and computers would most likely extrinsically motivate nursing students to attend their daily duty shifts.

The posttest results show that nursing students who are mentored, supported, monitored, supervised, and evaluated using an interactive web-based clinical practice monitoring system are more efficient in terms of clinical attendance and completing clinical activities on time. The findings of this study indicated that nursing students showed a readiness to stick to their clinical duty roster, report on the clinical environment on time, receive and give reports, and complete their assigned responsibilities using an interactive web-based clinical practice monitoring system. Furthermore, students expressed a desire to ask questions and locate clinical resources to help them learn not just to win a prize or a grade but also to broaden their knowledge and skills. Despite a significant change in clinical learning motivation among nursing students, the system improved clinical attendance as an indicator that it motivated and enhanced their interest and willingness with a sense of being confident, independent, and autonomous to engage in clinical learning activities than when conventional clinical pedagogies such as attendance books, bed tutorials, or assignments were predominantly used.

Similarly to the findings from other previous scholars [2,21], the implementation of an interactive web-based clinical practice monitoring system would allow nursing students to interact with one another while performing clinical nursing procedures, as well as interact instantly and timely with trainers, clinical instructors, or academic faculty for any support or mentorship. Nonetheless, as argued by some previous scholars [5,47] that individuals' behaviors are sometimes shaped by their personalities, this study found that nursing students' motivation in clinical learning was partly attributed to what motivated them to join nursing education programs, such as challenges in job availability, opportunities to demonstrate autonomy in nursing, and opportunities to travel around the world when individuals enrolled in the nursing profession. Such corelated aspects to the intervention would most likely drive nursing students to study nursing programs extrinsically rather than intrinsically to match their academic and living goals.

Referring to the findings from some previous scholarly works on students' abilities to demonstrate an interest in their learning activities becomes a precursor for them to be motivated to identify and locate learning resources and thus engage in learning activities actively [48,49]. Similarly, scholars' [6,50,51] low motivation in clinical learning has been linked to uninteresting clinical teaching techniques, an uncomfortable learning environment, a lack of interest, and unclear clinical objectives, all of which contribute to clinical absenteeism among nursing students. This study's findings on motivation to learn are consistent with those found by Millanzi and Kibusi [6], for example, who argued that innovative pedagogy has the potential of improving learning motivation among nursing students.

XSL•FO
RenderX

Nevertheless, Allvin et al [52], claimed that while the clinical environment is important for nursing students' academic achievement, students' learning motivation is positively correlated with an adequate number of competent and qualified clinical instructors who use innovative clinical pedagogies that enhance their motivation in clinical learning, including the prescription and integration of interactive web-based clinical practice monitoring system in nursing curricula. In the same way, Aghajari et al [53] observed that the majority of nursing students lack academic motivation during clinical placement practices because the clinical learning environment, as well as clinical teaching and learning pedagogies, does not motivate them to learn and meet their clinical academic potential as lifelong learners. This study's and prior research findings indicate to underline that a favorable, learner-centered, and technology-based innovative clinical pedagogy may positively boost clinical learning motivation among nursing students in nursing education.

Limitations of the Study

The study did not involve a control group to maximize the validity of findings on the efficacy of interactive web-based clinical practice monitoring systems on the outcomes of interest. The use of a single group may obscure the interpretation of the effect size on the outcome variables because standard clinical teaching pedagogies would also produce effects on the outcome of interest. Therefore, findings on the effect of interactive web-based clinical practice monitoring systems need to be interpreted cautiously by considering this limitation. The study suffers from a methodological limitation, as it did not adopt a randomized controlled trial (a true intervention) to estimate the random effect of the interactive web-based clinical practice monitoring system on the outcome variables over the standard clinical teaching pedagogies. Therefore, findings need to be treated and interpreted cautiously by considering that with a quasi-experimental design random effect of the intervention on the outcome, variables would not be established without outweighing its effect over the standard clinical teaching pedagogy (control group) if it could be involved.

Conclusions

The findings of this study demonstrate that it is possible to teach, mentor, supervise, support, monitor, and assess nursing students throughout their clinical placements by adopting, implementing, and evaluating an interactive web-based clinical practice monitoring system. As a minimum exposure of at least 3 weeks of the interactive web-based clinical practice monitoring system, nursing students may demonstrate nursing students' attendance and motivation in clinical learning by their will than is now performed, where sanctions and other associated techniques are used to force them to attend their daily clinical duty shifts accordingly. Incorporating technology into clinical nursing education pedagogics nursing curricula can be an alternative educational technique for educators in nursing education to facilitate clinical learning activities to develop motivated and passionate undergraduate nursing students to engage and learn efficiently and effectively during their clinical placements. Nonetheless, managing a big group of nursing students was proven to be achievable with the use of an interactive web-based clinical practice monitoring system.

Furthermore, an interactive web-based clinical practice monitoring system was feasible to not only mentor, supervise, monitor, and support nursing students but also record students' clinical attendance and the type and number of clinical nursing procedures learned and practiced, as well as generate clinical formative evaluation reports. In other words, an interactive web-based clinical practice monitoring system can be used as an innovative clinical pedagogical approach in clinical teaching and learning to improve nursing students' motivation in clinical learning as a precursor to clinical competence for quality and cost-effective care to people.

Acknowledgments

This study would not have been possible without the support of The University of Dodoma and the Dodoma Regional Referral Training Hospital. On the other hand, the publication of this work would not be possible without support from JMIR Publications for granting a discount of article-processing charges for this work to be published. The work was privately sponsored.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request via patriciaz1006@gmail.com.

Authors' Contributions

PH participated in the conceptualization, methods and materials, resources, investigation, writing the original draft, revision, and editing the draft of the work. SK contributed to the conceptualization, methodology, supervision, revision, and editing of the draft of the work. WM participated in the conceptualization, methods and materials, data curation, analysis, writing, and editing the original draft of the work. The authors have read and approved the manuscript.

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. Oguguo BC, Ajuonuma JO, Azubuike R, Ene CU, Atta FO, Oko CJ. Influence of social media on students' academic achievement. Int J Eval Res Educ 2020;9(4):1000.
2. Ghasemi MR, Moonaghi HK, Heydari A. Strategies for sustaining and enhancing nursing students' engagement in academic and clinical settings: a narrative review. Korean J Med Educ 2020;32(2):103-117 [FREE Full text] [doi: 10.3946/kjme.2020.159] [Medline: 32486620]
3. Soroush A, Andaieshgar B, Vahdat A, Khatony A. The characteristics of an effective clinical instructor from the perspective of nursing students: a qualitative descriptive study in Iran. BMC Nurs 2021;20(1):36 [FREE Full text] [doi: 10.1186/s12912-021-00556-9] [Medline: 33663461]
4. Carey M, Kent B, Latour J. Experiences of undergraduate nursing students in peer assisted learning in clinical practice: a qualitative systematic review. JBI Database System Rev Implement Rep 2018;16(5):1190-1219 [FREE Full text] [doi: 10.11124/JBISRIR-2016-003295]
5. Millanzi W, Kibusi S. Exploring the effect of problem based facilitatory teaching approach on motivation to learn: a quasi-experimental study of nursing students in Tanzania. BMC Nurs 2021;20(1):3 [FREE Full text] [doi: 10.1186/s12912-020-00509-8] [Medline: 33397332]
6. Millanzi W, Kibusi S. Exploring the effect of problem-based facilitatory teaching approach on metacognition in nursing education: a quasi-experimental study of nurse students in Tanzania. Nurs Open 2020;7(5):1431-1445 [FREE Full text] [doi: 10.1002/nop2.514] [Medline: 32802363]
7. Forman T, Armor D, Miller A. A review of clinical informatics competencies in nursing to inform best practices in education and nurse faculty development. Nurs Educ Perspect 2020;41(1):E3-E7. [doi: 10.1097/01.NEP.0000000000000588] [Medline: 31860501]
8. Chewaka Gamtessa L. Correlation between academic and clinical practice performance of nursing students at a pediatrics and child health nursing course; Mizan-Tepi University, Ethiopia. Adv Med Educ Pract 2021;12:155-162 [FREE Full text] [doi: 10.2147/AMEP.S294650] [Medline: 33623467]
9. Kaphagawani N, Useh U. Clinical supervision and support: exploring pre-registration nursing students' clinical practice in Malawi. Ann Glob Health 2018;84(1):100-109 [FREE Full text] [doi: 10.29024/aogh.16] [Medline: 30873795]
10. National Association of School Nurses. Framework for 21st century school nursing practice™: clarifications and updated definitions. NASN Sch Nurse 2020;35(4):225-233. [doi: 10.1177/1942602X20928372] [Medline: 32491974]
11. Millanzi WC, Herman PZ, Hussein MR. The impact of facilitation in a problem-based pedagogy on self-directed learning readiness among nursing students: a quasi-experimental study in Tanzania. BMC Nurs 2021;20(1):242 [FREE Full text] [doi: 10.1186/s12912-021-00769-y] [Medline: 34872553]
12. Rahman T, Ghani DM, Kausar S. Factors contributing to absenteeism in undergraduates nursing students. Pakistan J Educ 2021;38(2).
13. Cusack L, Thornton K, Drioli-Phillips P, Cockburn T, Jones L, Whitehead M, et al. Are nurses recognised, prepared and supported to teach nursing students: mixed methods study. Nurse Educ Today 2020;90:104434. [doi: 10.1016/j.nedt.2020.104434] [Medline: 32315837]
14. Lenouvel E, Chivu C, Mattson J, Young JQ, Klöppel S, Pinilla S. Instructional design strategies for teaching the mental status examination and psychiatric interview: a scoping review. Acad Psychiatry 2022;46(6):750-758 [FREE Full text] [doi: 10.1007/s40596-022-01617-0] [Medline: 35318592]
15. Sommer I, Larsen K, Nielsen CM, Stenholt BV, Bjørk IT. Improving clinical nurses' development of supervision skills through an action learning approach. Nurs Res Pract 2020;2020:9483549 [FREE Full text] [doi: 10.1155/2020/9483549] [Medline: 32148957]
16. Tesfaye TS, Alemu W, Mekonen T. Perceived clinical practice competency and associated factors among undergraduate students of medicine and health science collage in Dilla University, SNNPR, Ethiopia. Adv Med Educ Pract 2020;11:131-137 [FREE Full text] [doi: 10.2147/AMEP.S235823] [Medline: 32110133]
17. Mashala YL. The impact of the implementation of free education policy on secondary education in Tanzania. Int J Acad Multidiscip Res Internet 2019;3(1):6-14 [FREE Full text]
18. Thirumoorthy G. Outcome based education (OBE) is need of the hour. Int J Res GRANTHAALAYAH 2021;9(4):571-582.
19. Mlaba ZP, Emmamally W. Describing the perceptions of student nurses regarding barriers and benefits of a peer-mentorship programme in a clinical setting in KwaZulu-Natal. Health SA 2019;24:1118 [FREE Full text] [doi: 10.4102/hsag.v24i0.1118] [Medline: 31934419]
20. Agarwal G, Gaber J, Richardson J, Mangin D, Ploeg J, Valaitis R, et al. Pilot randomized controlled trial of a complex intervention for diabetes self-management supported by volunteers, technology, and interprofessional primary health care teams. Pilot Feasibility Stud 2019;5:118 [FREE Full text] [doi: 10.1186/s40814-019-0504-8] [Medline: 31673398]
21. Kinnunen U, Heponiemi T, Rajalahti E, Ahonen O, Korhonen T, Hyppönen H. Factors related to health informatics competencies for nurses—results of a national electronic health record survey. Comput Inform Nurs 2019;37(8):420-429. [doi: 10.1097/CIN.0000000000000511] [Medline: 30741730]
22. Yuliawan D, Widyandana D, Nur Hidayah R. Utilization of Nursing Education Progressive Web Application (NEPWA) media in an education and health promotion course using Gagne's model of instructional design on nursing students:

XSL•FO
RenderX

quantitative research and development study. JMIR Nurs 2020;3(1):e19780 [FREE Full text] [doi: 10.2196/19780] [Medline: 34345790]
23. Gemuhay H, Kalolo A, Mirisho R, Chipwaza B, Nyangena E. Factors affecting performance in clinical practice among preservice diploma nursing students in Northern Tanzania. Nurs Res Pract 2019;2019:3453085 [FREE Full text] [doi: 10.1155/2019/3453085] [Medline: 30941212]
24. Anisenkov A, Zhadan D, Logashenko I. A Web-based control and monitoring system for DAQ applications. EPJ Web Conf 2019;214:01049. [doi: 10.1051/epjconf/201921401049]
25. Mico O, Santos P, Caldo R. Web-based smart farm data monitoring system?: a prototype. Int J Agric Environ Food Sci 2016;3(3):85-96.
26. Kotcherlakota S, Pelish P, Hoffman K, Kupzyk K, Rejda P. Augmented reality technology as a teaching strategy for learning pediatric asthma management: mixed methods study. JMIR Nurs 2020;3(1):e23963 [FREE Full text] [doi: 10.2196/23963] [Medline: 34406970]
27. El Idrissi WEM, Chemsi G, Kababi KE, Radid M. Assessment practices of student's clinical competences in nurse education. Open Nurs J 2021;15(1):47-54. [doi: 10.2174/1874434602115010047]
28. UDOM. Regulations and Guidelines for Postgraduate Programmes [Internet]. 3rd ed. In: The University of Dodoma. Dodoma: Office of the Deputy Vice-Chancellor; Academic, Research and Consultancy; 2019.
29. Salim MA, Gabrieli P, Millanzi WC. Enhancing pre-school teachers' competence in managing pediatric injuries in Pemba Island, Zanzibar. BMC Pediatr 2022;22(1):691 [FREE Full text] [doi: 10.1186/s12887-022-03765-6] [Medline: 36461011]
30. Cleophace W. The effectiveness of integrated reproductive health lesson materials in a problem-based pedagogy on knowledge, soft skills, and sexual behaviors among adolescents in Tanzania. In: The University of Dodoma. Dodoma, Tanzania: The University of Dodoma; 2021:1-538.
31. Haramba S, Millanzi W, Seif S. Enhancing nursing student presentation competences using Facilitatory Pecha Kucha presentation pedagogy: a quasi-experimental study protocol in Tanzania. BMC Med Educ 2023;23(1):628 [FREE Full text] [doi: 10.1186/s12909-023-04628-z] [Medline: 37661279]
32. Shitindi G, Millanzi W, Herman P. Perceived motivators, knowledge, attitude, self-reported and intentional practice of female condom use among female students in higher training institutions in Dodoma, Tanzania. Contracept Reprod Med 2023;8(1):16 [FREE Full text] [doi: 10.1186/s40834-022-00208-6] [Medline: 36750970]
33. Mwanja C, Herman P, Millanzi W. Prevalence, knowledge, attitude, motivators and intentional practice of female genital mutilation among women of reproductive age: a community-based analytical cross-sectional study in Tanzania. BMC Womens Health 2023;23(1):226 [FREE Full text] [doi: 10.1186/s12905-023-02356-6] [Medline: 37138247]
34. Millanzi W, Herman P, Mtangi S. Knowledge, attitude, and perceived practice of sanitary workers on healthcare waste management: a descriptive cross-sectional study in Dodoma region,Tanzania. SAGE Open Med 2023;11:20503121231174735 [FREE Full text] [doi: 10.1177/20503121231174735] [Medline: 37223674]
35. Millanzi WC, Herman PZ, Ambrose BA. Feeding practices, dietary adequacy, and dietary diversities among caregivers with under-five children: a descriptive cross-section study in Dodoma region, Tanzania. PLoS One 2023;18(3):e0283036 [FREE Full text] [doi: 10.1371/journal.pone.0283036] [Medline: 36947536]
36. Millanzi W, Osaki K, Kibusi S. Non-cognitive skills for safe sexual behavior: An exploration of baseline abstinence skills, condom use negotiation, self-esteem, and assertiveness skills from a controlled problem-based learning intervention among adolescents in tanzania non-cognitive skills for safe sexual behavior: an exploration of baseline abstinence skills, condom use negotiation, self-esteem, and assertiveness skills from a controlled problem-based learning intervention among adolescents in Tanzania. Glob J Med Res Internet 2020;20:43-50 [FREE Full text]
37. Millanzi W, Osaki K, Kibusi S. Attitude and prevalence of early sexual debut and associated risk sexual behavior among adolescents in Tanzania; evidence from baseline data in a Randomized Controlled Trial. BMC Public Health 2023;23(1):1758 [FREE Full text] [doi: 10.1186/s12889-023-16623-6] [Medline: 37689638]
38. Millanzi W, Kibusi SM, Osaki KM. Effect of integrated reproductive health lesson materials in a problem-based pedagogy on soft skills for safe sexual behaviour among adolescents: a school-based randomized controlled trial in Tanzania. PLoS One 2022;17(2):e0263431 [FREE Full text] [doi: 10.1371/journal.pone.0263431] [Medline: 35192640]
39. Millanzi W, Kibusi S, Osaki K. Effect of integrated reproductive health lesson materials in a problem-based pedagogy on soft skills for safe sexual behaviour among adolescents: a school-based randomized controlled trial in Tanzania. PLoS One 2022;17(2):e0263431 [FREE Full text] [doi: 10.1371/journal.pone.0263431] [Medline: 35192640]
40. Millanzi W, Osaki K, Kibusi S. The effect of educational intervention on shaping safe sexual behavior based on problem-based pedagogy in the field of sex education and reproductive health: clinical trial among adolescents in Tanzania. Health Psychol Behav Med 2022;10(1):262-290 [FREE Full text] [doi: 10.1080/21642850.2022.2046474] [Medline: 35251774]
41. Millanzi WC. Adolescents' World: Know One Tell One Against Unsafe Sexual Behaviours, Teenage Pregnancies and Sexually Transmitted Infections Including Chlamydia. In: IntechOpen. London, UK: IntechOpen; 2022:1541-1548.
42. Lucidi FF. The Academic Motivation Scale (AMS): factorial structure, invariance, and validity in the Italian context. TPM Test Psychom Methodol Appl Psychol 2008;15(4):211-220.
43. Utvær BK, Haugan G. The academic motivation scale: dimensionality, reliability, and construct validity among vocational students. J Vocat Educ Train 2016;6(2):17-45.

XSL•FO
RenderX

44. Taghipour H, Gilaninia S, Jalali M, Azizipour H, Javad S, Razaghi R. Standardizing of academic motivation scale. J Basic Appl Sci Res 2012;2(2):1186-1192.
45. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. Res Sci Educ 2017;48(6):1273-1296 [FREE Full text] [doi: 10.1007/s11165-016-9602-2]
46. Brydges C. Effect size guidelines, sample size calculations, and statistical power in Gerontology. Innov Aging 2019;3(4):igz036 [FREE Full text] [doi: 10.1093/geroni/igz036] [Medline: 31528719]
47. Ibrahem S, Elsayed N, Mohamed H, Shereda A. Impact of clinical simulation on student's levels of motivation, satisfaction and self-confidence in learning psychiatric mental health nursing. Int J Spec Educ 2022;37(3):6214-6228.
48. Taylor I, Bing-Jonsson P, Wangensteen S, Finnbakk E, Sandvik L, McCormack B, et al. The self-assessment of clinical competence and the need for further training: a cross-sectional survey of advanced practice nursing students. J Clin Nurs 2020;29(3-4):545-555. [doi: 10.1111/jocn.15095] [Medline: 31714619]
49. Konttila J, Siira H, Kyngäs H, Lahtinen M, Elo S, Kääriäinen M, et al. Healthcare professionals' competence in digitalisation: a systematic review. J Clin Nurs 2019;28(5-6):745-761. [doi: 10.1111/jocn.14710] [Medline: 30376199]
50. Alaagib NA, Musa OA, Saeed AM. Comparison of the effectiveness of lectures based on problems and traditional lectures in physiology teaching in Sudan. BMC Med Educ 2019;19(1):365 [FREE Full text] [doi: 10.1186/s12909-019-1799-0] [Medline: 31547817]
51. Solomon Y. Comparison between problem-based learning and lecture-based learning: effect on nursing students' immediate knowledge retention [Response To Letter]. Adv Med Educ Pract 2021;12:163-164 [FREE Full text] [doi: 10.2147/AMEP.S305514] [Medline: 33623468]
52. Allvin R, Bisholt B, Blomberg K, Bååth C, Wangensteen S. Self-assessed competence and need for further training among registered nurses in somatic hospital wards in Sweden: a cross-sectional survey. BMC Nurs 2020;19:74 [FREE Full text] [doi: 10.1186/s12912-020-00466-2] [Medline: 32774153]
53. Aghajari Z, Loghmani L, Ilkhani M, Talebi A, Ashktorab T, Ahmadi M, et al. The relationship between quality of learning experiences and academic burnout among nursing students of Shahid Beheshti University of Medical Sciences in 2015. Electron J Gen Med 2018;15(6). [doi: 10.29333/ejgm/93470]

Edited by T de Azevedo Cardoso, T Leung; submitted 22.01.23; peer-reviewed by RJ Medina, L Nunes, MF Deschênes; comments to author 03.08.23; revised version received 11.10.23; accepted 15.03.24; published 23.04.25.

Please cite as:
Herman P, M Kibusi S, C Millanzi W
Effectiveness of an Interactive Web-Based Clinical Practice Monitoring System on Enhancing Motivation in Clinical Learning Among Undergraduate Nursing Students: Longitudinal Quasi-Experimental Study in Tanzania
JMIR Med Educ 2025;11:e45912
URL: https://mededu.jmir.org/2025/1/e45912
doi:10.2196/45912
PMID:

©Patricia Herman, Stephen M Kibusi, Walter C Millanzi. Originally published in JMIR Medical Education (https://mededu.jmir.org), 23.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Media-Induced and Psychological Factors That Foster Empathy Through Virtual Reality in Nursing Education: 2×2 Between-Subjects Experimental Study

Kuo-Ting Huang[1], PhD; Zexin Ma[2], PhD; Lan Yao[3], PhD

[1]Department of Information Culture and Data Stewardship, University of Pittsburgh, 135 N Bellefield Ave, Room 616, Pittsburgh, PA, United States

[2]Department of Communication, University of Connecticut, Storrs, CT, United States

[3]School of Nursing, Oakland University, Rochester, MI, United States

Corresponding Author:
Kuo-Ting Huang, PhD
Department of Information Culture and Data Stewardship, University of Pittsburgh, 135 N Bellefield Ave, Room 616, Pittsburgh, PA, United States

Abstract

Background: Virtual reality (VR) has emerged as a promising tool in medical education, particularly for fostering critical skills such as empathy. However, how VR, combined with perspective-taking, influences affective empathy in nursing education remains underexplored.

Objective: This study investigates the influence of VR and perspective-taking on affective empathy in nursing education, focusing on 4 psychological factors: perceived self-location, narrative transportation, emotional engagement, and affective empathy.

Methods: A 2×2 between-subjects design was used, involving 69 nursing undergraduates from two Midwest universities. The participants engaged with a narrative-focused video game, That Dragon, Cancer, in either VR or non-VR conditions and from the perspective of either parents or clinicians.

Results: VR significantly enhanced perceived self-location (P=.01), while adopting a clinician's perspective amplified emotional engagement (P=.03). However, VR did not significantly influence narrative transportation (P=.35). An interaction effect was found between the platform and player's perspective on narrative transportation (P=.04). Several indirect effects of media elements on affective empathy were observed via other psychological factors, though the direct effect of VR on affective empathy was not significant (P=.84).

Conclusions: These findings underscore the potential of VR in medical education, suggesting that perspective-taking should be carefully considered when designing immersive learning experiences. The study advocates for broader integration of VR technologies into medical curricula to enhance instruction quality and patient-centered care.

(JMIR Med Educ 2025;11:e59083) doi:10.2196/59083

KEYWORDS

nursing education; narrative transportation; presence; virtual reality; game-based learning; affective empathy

Introduction

Background

The domain of health care education is currently undergoing a monumental shift, facilitated by advancements in immersive technologies such as virtual reality (VR). Immersive technologies have demonstrated significant potential in transforming medical and health education, including enhanced training in surgical procedures [1,2], improved understanding of complex biomedical processes through immersive visualization [3,4], and more empathetic patient care through simulated patient interactions [5,6]. Despite its promise, VR technology faces challenges such as system fidelity and presence, which can impact user experience and learning outcomes [7]. Addressing these challenges is essential for the effective design and implementation of VR training modules.

Moreover, VR provides a safe and controlled environment for students to practice and make mistakes without direct consequences on actual patients, increasing their confidence and proficiency before real-world clinical scenarios [8,9]. Concurrent with this paradigm shift, immersive technologies like VR have emerged as an effective tool in the instruction of empathy among nursing students [10-12]. Empathy, being a fundamental aspect of the nursing profession, has been shown to improve patient adherence to treatment plans [13], satisfaction levels [14], and overall health outcomes [15]. Specifically, being

XSL•FO
RenderX

immersed in virtual narratives allows students to navigate and process their own emotions, as well as respond appropriately within simulated scenarios [16]. This innovative approach offers a safe space for students to handle the emotional complexities associated with patient care, thus better preparing them for future clinical encounters involving nuanced emotional interactions.

VR simulation has also been studied in the context of nursing education [17-21]. In our recent study [22], we found that VR-based role-playing games enhanced cognitive empathy among nursing students. However, affective empathy remains underexplored in VR-based nursing education. Cognitive empathy involves understanding another's perspective, whereas affective empathy refers to sharing and responding to another's emotional states [23]. A recent meta-analysis revealed that VR has a significant effect on perspective-taking outcomes (cognitive empathy) but lacks impact on affective empathy [23]. Therefore, the objective of the current research is to investigate the potential of VR to influence affective empathy among nursing students by exploring narrative-related psychological factors.

Research on immersive media indicates that perceived self-location (ie, being there in a virtual environment) is a key mechanism that explains the impact of VR on empathy [24,25]. VR-based empathy training programs often contain a story with plots and characters to help users experience a situation first-hand [26]. They comprise both medium-based (ie, of the virtual environment) and message-based (ie, of the narrative) elements [27].To understand how VR might enhance affective empathy, we explore specific psychological mechanisms: perceived self-location and narrative transportation.

Perceived self-location refers to the sensation of "being there" in a virtual environment, enhancing users' immersion and empathy [24,25], which we hypothesize will lead to greater empathy by allowing users to more deeply understand and share the feelings of virtual characters. Narrative transportation is the cognitive and emotional absorption into a story, where individuals lose awareness of their physical surroundings and form intense connections with the narrative and characters [28,29]. This deep absorption can lead to significant shifts in attitudes and empathy toward others [30]. Research has found that VR narratives lead to higher levels of transportation compared with traditional media, which in turn enhance empathetic responses [31]. Therefore, we hypothesize that VR will enhance narrative transportation, significantly impacting affective empathy.

A noteworthy characteristic of VR-based empathy training programs is their ability to feature multiple characters, thus allowing users to experience a narrative from varying character perspectives [32,33] The perspective-taking aspect significantly impacts users' emotional engagement with the character [34]. Research has found that users are more likely to experience emotional engagement toward characters that are portrayed positively [33] and are similar to them [35]. Therefore, we hypothesize that character perspective will impact emotional engagement, with higher engagement when viewing from the clinician's perspective compared with the parents' perspective due to character-user similarity, subsequently influencing affective empathy.

Moreover, we propose that perceived self-location, narrative transportation, and emotional engagement will form a sequential mediation model to account for the effect of VR-based training. Previous research on video games has found that the feeling of "being there" in the game is a predictor of flow, an experience similar to narrative transportation [36,37]. A recent study obtained similar findings: spatial presence predicted narrative transportation in a VR storytelling experience [38]. Furthermore, existing work suggests that narrative transportation is associated with an increase in emotional responses [39,40]. Hence, we predict that perceived self-location, enhanced by VR, will foster narrative transportation, which will, in turn, promote emotional engagement. Emotional engagement will then lead to affective empathy.

Hypotheses and Research Questions

Based on the literature review, the following hypotheses were proposed to explore the interconnected roles of perceived self-location, narrative transportation, and emotional engagement in enhancing affective empathy through VR interventions:

1. Hypothesis 1: VR conditions will enhance (1) perceived self-location and (2) narrative transportation compared with non-VR conditions.
2. Hypothesis 2: the participants will experience higher emotional engagement when viewing the narrative from a clinician's perspective compared with a parent's perspective.
3. Hypothesis 3: perceived self-location will positively predict narrative transportation in VR-based training programs.
4. Hypothesis 4: narrative transportation will positively predict emotional engagement within VR-based experiences.
5. Hypothesis 5: emotional engagement will positively predict affective empathy.

In addition to these hypotheses, we proposed the following research questions to explore potential interaction and indirect effects:

1. Research question 1: Are there interaction effects between the media platform (VR vs non-VR) and character perspective (clinician vs parents) on psychological factors such as perceived self-location, narrative transportation, and emotional engagement?
2. Research question 2: Do perceived self-location, narrative transportation, and emotional engagement mediate the relationship between the media platform and affective empathy?

Methods

Design

The proposed conceptual framework of this study is illustrated in Figure 1. This study used a 2×2 between-subjects experimental design to investigate the effects of the platform (VR vs non-VR) and perspective (parents vs clinicians) on nursing undergraduates' empathy levels. Participants were randomly assigned to 1 of 4 conditions: VR parents, VR

XSL•FO
RenderX

clinicians, non-VR parents, or non-VR clinicians. Figure 2 shows the 4 conditions of the study.

Figure 1. The proposed conceptual framework.

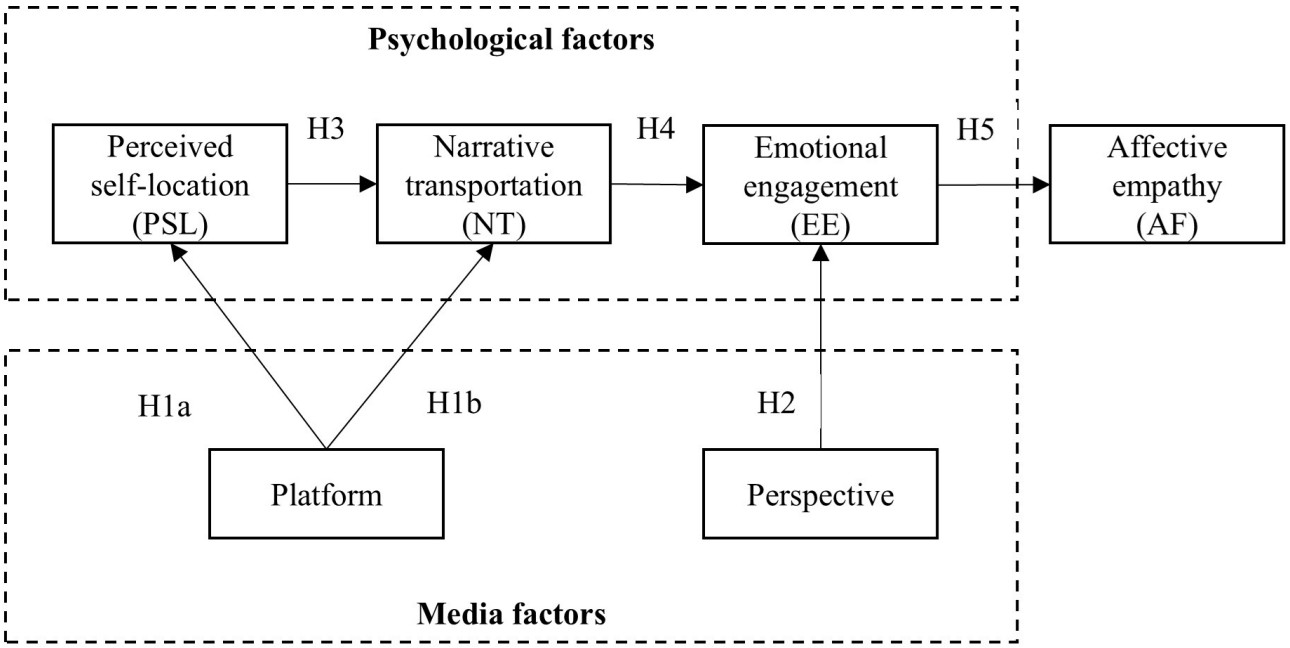
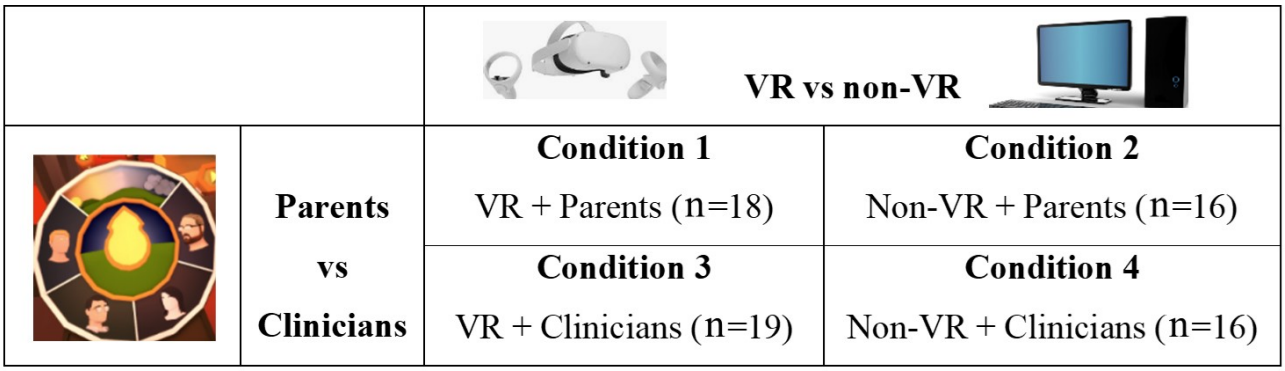


Figure 2. Four conditions of the study.



Participants

A total of 69 nursing undergraduates from two Midwest universities participated in the study, predominantly female (57/69, 82.6%) and White (57/69, 82.6%), mostly juniors (39/69, 56.5%; sophomores: 24/69, 34.8%; seniors: 6/69, 8.7%). The average age was 22.13 (SD 3.70) years. The participants were recruited via university mailing lists and classroom announcements, with inclusion criteria requiring them to be nursing undergraduates aged 18 years or older. Most participants (54/69, 77.6%) reported they were not familiar with VR.

Before participation, all were screened for visual impairments and susceptibility to motion sickness; those with significant issues were excluded to avoid adverse effects during the VR experience.

Randomization was done through a drawing process, assigning participants to 1 of 4 conditions, ensuring balanced allocation between VR and non-VR platforms and the perspectives of parents or clinicians. Given the sample size and complexity of the 2×2 design, the study may be underpowered to detect small effects. No a priori power calculations were conducted; however, this exploratory research aims to investigate initial effects and generate hypotheses for future studies with larger samples.

Experimental Procedures and Stimulus

Upon arrival at the research lab, participants provided informed consent and completed a short pretest questionnaire. They were then assigned to 1 of the 4 gameplay conditions as per the randomization process described above. The participants were informed that they could stop or report at any time if they experienced motion sickness or visual discomfort.

The participants engaged with the seventh chapter of the narrative-focused video game That Dragon, Cancer, titled “I’m Sorry Guys, It’s Not Good.” That Dragon, Cancer is an interactive narrative game developed by Numinous Games (Mainframe Studios) [41] that tells the real-life story of a family’s experience with their son’s terminal cancer diagnosis. The game is designed to evoke emotional responses and foster empathy through immersive storytelling [42]. This chapter was selected based on prior research demonstrating its efficacy in increasing empathy among medical students [42]. The gameplay allowed participants to experience a pivotal moment when Joel’s parents were informed of his terminal cancer diagnosis,

XSL•FO
RenderX

navigating the scene from 4 unique perspectives: dad, mom, doctor, and nurse. The game is designed as a point-and-click adventure, which allows participants to trigger conversations and access a selected character's inner thoughts.

The participants in the VR conditions used Oculus Go headsets, seated in a quiet room to minimize distractions. The headsets provided a 360-degree immersive experience with built-in headphones for audio. For the non-VR conditions, the participants used Dell laptops or iPads (Apple, Inc) with over-ear headphones, seated at a desk in the same room to ensure consistent environmental conditions. The game was presented on a standard screen, and the participants interacted using a mouse or touchscreen, replicating typical non-VR gameplay settings. The gameplay lasted approximately 10 minutes. The selection of this exposure time was based on previous studies indicating that brief VR experiences can effectively elicit emotional and empathetic responses [43]. Immediately after the gameplay, participants completed a posttest questionnaire assessing their empathy and gaming experience. This immediate administration was intended to capture their reactions and reduce potential recall bias.

Instrument

Several validated scales were used to measure the constructs of interest. These measures are specifically applicable to our study's context in nursing education and VR-based empathy training. First, the Spatial Presence Experience Scale, developed by Hartmann et al [44], was used to evaluate self-location and assess nursing students' immersion in the simulated clinical scenarios. This validated scale, widely used in diverse media environments, measures 2 facets of the spatial presence experience: perceived self-location and potential actions, while also considering key factors that influence spatial presence. The study's transportation scale was an adaptation from Green and Brock [28], which measures students' absorption of patient stories that may foster empathy. In addition, the Emotional Engagement scale used in the study was sourced from Knol and Van Linge [45], which captures students' emotional connection with virtual characters, vital for developing affective empathy. Affective empathy was assessed with 3 items from the validated empathy scale by Batson et al [46]. The participants responded to items on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). The results of these items were averaged to produce a composite score for data analysis. Covariates such as participants' gender, age, race, and university affiliation were included in the following statistical analyses. Descriptive statistics, the items used, and reliability estimates for all scales are presented in Table 1.

Table . List of items used in the study and the descriptive statistics (N=69).

Variable	Statistics	Items
Perceived self-location [44]	• Mean 5.19 (SD 1.12) • Cronbach (α=0.88)	• I felt like I was actually there in the environment of the presentation. • It was as though my true location had shifted into the environment in the presentation. • I felt as though I was physically present in the environment of the presentation. • It seemed as though I actually took part in the action of the presentation
Transportation [28]	• Mean 5.77 (SD 0.93) • Cronbach α=0.83	• I could picture myself in the scene of the events described in the story. • I was mentally involved in the story while watching it. • I wanted to learn how the story ended.
Emotional engagement [45]	• Mean 5.83 (SD 0.98) • Cronbach α=0.83	• The story affected me emotionally. • During the media experience, when the characters suffered in some way, I felt sad. • I felt sorry for some of the characters in the story.
Affective empathy [46]	• Mean 5.65 (SD 1.24) • Cronbach α=0.93	• Did watching/playing this video make you feel __ • softhearted • sympathetic • compassionate

Statistical Analysis

Data analysis was performed using SPSS 29 statistical software. Hypotheses H1, H2, and RQ1 were assessed through a series of analyses of covariance (ANCOVA), controlling for covariates such as participants' gender, age, race, and university affiliation. Covariates such as participants' gender, age, race, and university affiliation were included in the ANCOVA because these demographic factors have been shown to influence empathy levels and responses to VR experiences [23]. Including these covariates helps control for potential confounding variables, ensuring that the effects observed are attributable to the

XSL•FO
RenderX

experimental manipulations rather than demographic differences. Hypotheses H3-H5 and RQ2 were tested with mediation analyses using the PROCESS macro, following a bootstrap estimation approach with 5000 samples, based on Hayes' Process Model 6 [47]. Control variables were also included in these analyses to control for potential confounding influences.

Ethical Considerations

This study was approved by the Institutional Review Board at Ball State University (approval number 1386023‐1). All participants provided informed consent before data collection. To acknowledge their participation, they received extra course credits as compensation. Confidentiality and privacy were maintained, and participants had the right to withdraw at any time without consequences.

Results

Descriptive Statistics and Correlation Analysis

The descriptive statistics for each condition, including mean and SD, are presented in Table 2. In addition, skewness and kurtosis values were assessed to check the normality of the data distribution, and the results were within the acceptable range confirming the appropriateness of the data for further analysis. A correlation analysis was conducted to identify any potential multicollinearity issues among the variables. The results indicated that while variables were correlated, they did not exceed the threshold that would suggest multicollinearity, thus ensuring the independence of predictors. The means for all variables were above the midpoint of the scale, indicating generally high levels of perceived self-location, narrative transportation, emotional engagement, and affective empathy among participants. The SD values ranged from 0.83 to 1.24, suggesting moderate variability in responses.

Table . Means and SEs by experimental conditions (N=69). Participants' race, gender, school year, and university affiliation were controlled.

Platform and perspective	Perceived self-location	Narrative transportation	Emotional engagement	Affective empathy
VR, mean (SE)				
Parents (n=18)	5.58 (1.07)	6 (0.87)	5.85 (0.72)	5.76 (1.43)
Clinicians (n=19)	5.66 (0.98)	5.81 (1.07)	6.02 (1.04)	5.84 (1.15)
Non-VR, mean (SE)				
Parents (n=16)	4.50 (1.34)	5.27 (1.16)	5.31 (1.26)	5.23 (1.41)
Clinicians (n=16)	4.89 (0.65)	5.96 (0.59)	6.13 (0.7)	5.75 (0.9)

Analysis of Covariance

The analysis of covariance (ANCOVA) included checks for assumptions such as homogeneity of variances, assessed using the Levene test. The results of these tests confirmed that the assumptions of ANCOVA were met across all variables of interest. Specifically, Levene test results for homogeneity of variances were $F_{3,65}$=2.146, P=.10 for self-location; $F_{3,65}$=1.566, P=.20 for narrative transportation; $F_{3,65}$=0.904, P=.44 for emotional engagement; and $F_{3,65}$=0.620, P=.60 for affective empathy. These results suggest that the variances of the residuals were not significantly different from each other across groups for each variable, thus fulfilling one of the key assumptions for conducting ANCOVA and lending validity to the subsequent analyses.

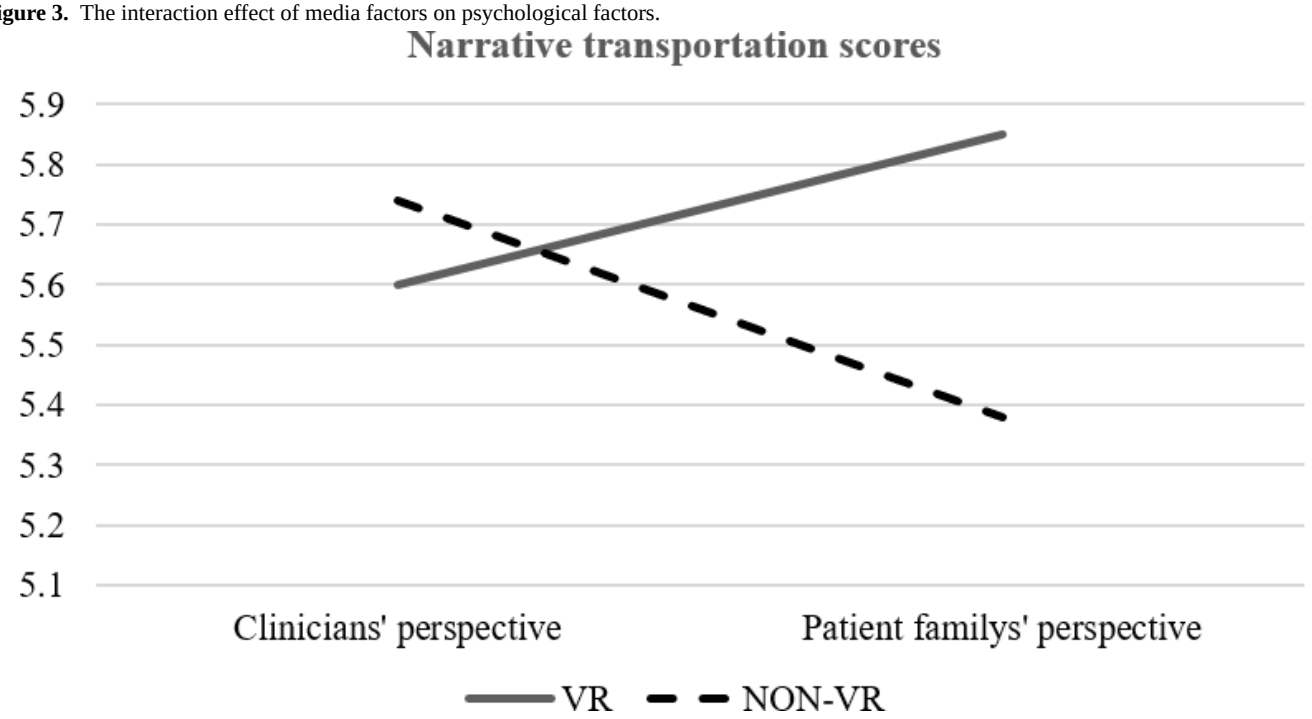
Based on our findings (Table S1 in Multimedia Appendix 1), the platform of the game had a significant influence on perceived self-location ($F_{1,61}$=6.60, P=.01, partial η^2=0.098), indicating that VR has a stronger influence on perceived self-location compared with non-VR environments. This substantial difference in perceived self-location depending on the platform used provided support for hypothesis H1a, suggesting VR's unique capacity to enhance users' sense of presence within the virtual environment. We also examined the effect of the adopted perspective on participants' emotional engagement. The results revealed a significant effect ($F_{1,61}$=4.76, P=.03, partial η^2=0.072), indicating that participants who assumed the clinician's perspective exhibited greater emotional engagement compared with those adopting a patient's perspective. This finding supported hypothesis H2, highlighting the importance of perspective in influencing emotional responses in VR settings. However, the effect of VR on narrative transportation did not yield significant results ($F_{1,61}$=0.90, P=.35, partial η^2=0.014), thereby not supporting hypothesis H1b.

To answer the first research question, we also test whether there is an interaction effect of media factors on psychological factors. The results revealed that the platform and perspective had an interaction effect on narrative transportation ($F_{1,61}$=4.68, P=.04, partial η^2=0.070). The post hoc analysis indicated that participants experiencing the game from the perspective of patients' families in a non-VR platform exhibited the lowest level of narrative transportation. This nuanced finding sheds light on how different combinations of platform and perspective can uniquely affect the immersive experience of users. These interaction effects are further elucidated in Figure 3, providing

XSL•FO
RenderX

a visual representation of these dynamics. Specifically, the narrative transportation scores for clinicians and patient families under VR (solid line) and non-VR (dashed line) conditions showed diverging trends between the two perspectives.

Figure 3. The interaction effect of media factors on psychological factors.



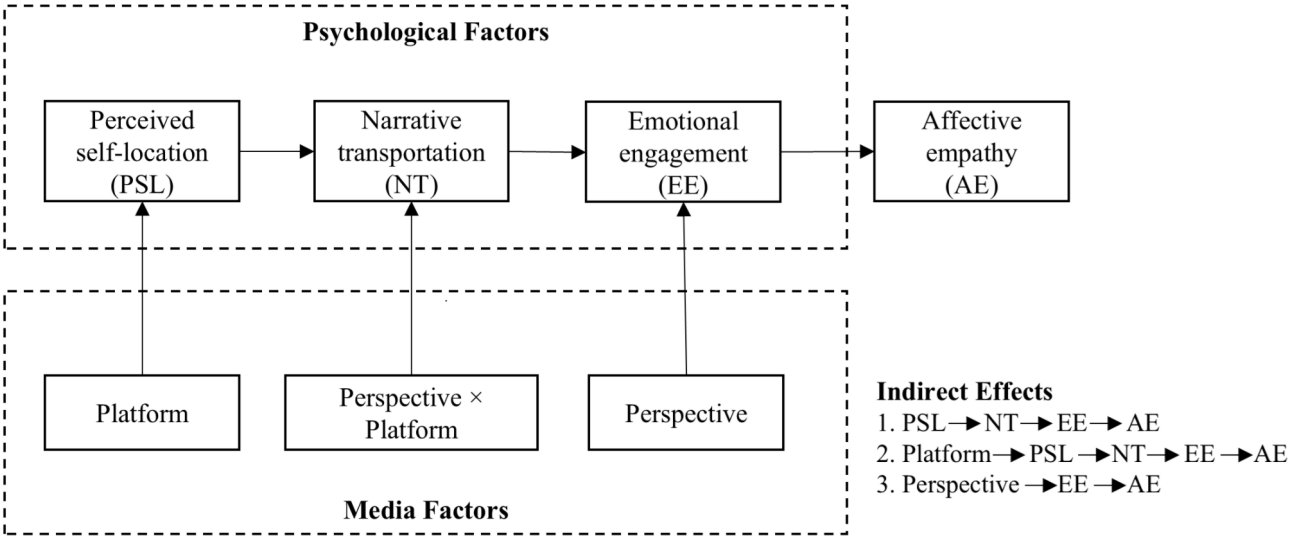
Mediation Analysis

The mediation analyses using Hayes' PROCESS model 6 revealed several direct and indirect effects. Regarding direct effects, self-location was found to have a significant positive effect on transportation (b=0.54, P<.001). Transportation had a positive effect on emotional engagement (b=0.74, P<.001). Emotional engagement was found to have a strong positive effect on affective empathy (b=0.84, P<.001). These findings, supporting H3-H5, illustrate the psychological process from self-location in VR, through narrative transportation and emotional engagement, to the ultimate development of affective empathy. We used 5000 bootstrap samples to generate bias-corrected confidence intervals for the indirect effects. The significance of the mediation pathways was determined by examining whether the confidence intervals excluded zero. All variables were included in the model simultaneously, and control variables were accounted for in the analysis. The model fit was assessed, and all pathways were found to be significant, confirming the validity of the mediation pathways.

The results also uncovered three sets of indirect effects, which answered our second research question. First, perceived self-location exerted an indirect effect on affective empathy through narrative transportation and emotional engagement (indirect effect=0.25, 95% CI 0.0610-0.5048). Second, the platform had an indirect effect on affective empathy through perceived self-location, narrative transportation, and emotional engagement (indirect effect=0.25, 95% CI 0.0610-0.5048). Finally, the perspective adopted by the participants was found to have an indirect effect on affective empathy through emotional engagement (indirect effect=0.27, 95% CI 0.0059-0.6153). These indirect effects, along with the direct effects, are represented in the revised conceptual framework presented in Figure 4.

XSL•FO
RenderX

Figure 4. The revised conceptual framework.



Discussion

Summary and Interpretations of the Findings

The study investigates the impact of media platforms and players' perspectives on perceived self-location, narrative transportation, emotional engagement, and affective empathy within medical education, which capture key psychological processes essential for developing empathy in clinical practice. The results indicate that all psychological factors were influenced by media elements, albeit through different mechanisms.

Our hypothesis H1a was supported, demonstrating that playing the game in VR (vs non-VR) significantly increased perceived self-location. This finding is consistent with prior studies [24,25]. However, contrary to our hypothesis H1b, VR (vs non-VR) did not differ in narrative transportation, which echoes findings from a recent meta-analysis [27]. This nonsignificant finding suggests that while VR enhances the sense of presence or self-location, it may not necessarily increase narrative transportation compared with non-VR platforms. One possible explanation is that narrative transportation is more strongly influenced by the quality of the narrative itself rather than the medium through which it is delivered [28]. It is possible that the narrative content was equally engaging in both VR and non-VR formats, resulting in similar levels of transportation. Future research could explore how different narrative structures or content types interact with VR to affect narrative transportation.

Regarding our hypothesis H2 adopting a clinician's perspective during the VR experience significantly influenced emotional engagement, the result is consistent with previous research [32,33]. This insight emphasizes the value of learning from other clinicians in strengthening emotional ties with patients, thereby fostering affective empathy. The findings showed the potential of VR in medical education by enhancing perceived self-location, which can improve empathy toward patients. The significant influence of perspective adopted by players reiterates its role in enhancing emotional engagement, crucial for fostering empathy in health care professionals.

The proposed direct relationships between psychological factors and affective empathy (H3-H5) were all supported. The findings demonstrate that a sense of perceived self-location in VR can enhance narrative transportation, leading to increased emotional engagement. In turn, this fosters affective empathy, a critical skill for health care professionals to understand and respond to patients' emotional experiences effectively, a critical skill for health care professionals to understand and respond to patients' emotional experiences effectively. This sequential process confirms the potential of VR in facilitating immersive, emotionally engaging learning experiences in medical training, promoting the development of affective empathy.

The exploration of the research questions (RQ1 and RQ2) yielded compelling insights. The first research question investigates potential interaction effects. Our findings indicate a significant interaction effect among these conditions, specifically revealing that scores on narrative transportation were significantly lower in the condition using non-VR with a patient's perspective compared with the other 3 conditions. This interaction effect suggests that the combination of platform and perspective plays a crucial role in influencing narrative transportation. One possible explanation is that adopting the patient's perspective in a non-VR environment may not provide sufficient immersion or sensory cues to facilitate deep engagement with the narrative. In contrast, VR may compensate for the less immersive perspective by enhancing sensory immersion, while adopting a clinician's perspective may align more closely with the students' professional identity, facilitating engagement even in non-VR settings. This finding indicates that the effectiveness of narrative transportation may depend on the congruence between the medium, the perspective adopted, and the user's own identity and experiences. Future studies could explore how personal relevance and role identification influence narrative engagement across different platforms.

For the second research question, the study uncovered 3 indirect effects: perceived self-location impacted affective empathy via narrative transportation and emotional engagement, the platform influenced affective empathy through self-location, narrative transportation, and emotional engagement, and the perspective affected affective empathy through emotional engagement.

XSL•FO
RenderX

Theoretical Contributions and Practical Implications

This study makes significant theoretical contributions to the fields of empathy research and narrative communication. It demonstrates how VR can influence affective empathy through mechanisms such as perceived self-location, narrative transportation, and emotional engagement, thereby deepening our understanding of the integral role immersive technologies play in fostering critical emotional competencies. By identifying the sequential mediation of these psychological factors, our findings extend existing theories on empathy development and immersive media, providing empirical evidence within the context of nursing education.

These findings add valuable empirical evidence to empathy research and highlight the importance of immersive, technology-enabled experiences in shaping affective responses. Specifically, our study fills a gap in the literature by focusing on affective empathy rather than cognitive empathy, which has been less examined in VR research. Furthermore, by exploring the interaction between character perspective and affective empathy within VR environments, the study offers a novel perspective on empathy development and enriches narrative communication research. This contributes to a more nuanced understanding of how perspective-taking in VR can differentially impact emotional engagement and empathy outcomes, which has practical implications for designing effective educational interventions.

From a practical standpoint, the findings offer actionable insights for integrating VR into nursing education. We propose that nursing programs should incorporate VR experiences that emphasize perspective-taking from a clinician's viewpoint to enhance emotional engagement and affective empathy among students. To address potential barriers such as cost, accessibility, and technological limitations, nursing programs could start by incorporating affordable VR solutions like stand-alone VR headsets or 360-degree video experiences, which are more feasible than high-end VR systems or simulation stations. For accessibility, it is important to ensure that VR experiences are also designed for students in the classroom settings. Technological limitations, such as a lack of technical expertise among faculty, can be mitigated through training workshops and technical support services. In addition, curricula should be designed to seamlessly integrate VR experiences into existing courses, perhaps starting with pilot programs to evaluate effectiveness before broader implementation. By proactively addressing these barriers, educators can more effectively leverage VR technology to enhance empathy training in nursing education. By implementing these recommendations, educators and institutions can leverage VR technology to significantly enhance the quality of medical education and training, especially in the domain of empathy development.

Limitations and Future Research Directions

This study has several limitations. First, while the sample size was sufficient to yield statistical power, it was relatively small. The participants were primarily female, white nursing undergraduates from two Midwestern universities. The small sample size may have also limited our ability to detect smaller effect sizes. The homogeneity in gender and race may have influenced the results, as previous research suggests that empathy levels and responses to VR experiences can vary across different demographic groups. For instance, gender differences have been observed in emotional processing and empathetic responses, which could affect how participants engage with VR-based empathy training. Therefore, our sample may limit the generalizability of the results.

Future studies should prioritize recruitment strategies that enhance demographic diversity to ensure the broader applicability of findings. One approach is expanding outreach to institutions with more diverse student populations, such as historically Black colleges and universities, Hispanic-serving institutions, and community colleges. Establishing partnerships with nursing programs in urban and rural areas can also help reach a wider range of participants with different socioeconomic and educational backgrounds. In addition, leveraging professional nursing associations, student organizations, and social media platforms can improve the recruitment of participants from underrepresented groups. Providing flexible participation options, such as internet-based study components or varied scheduling, may further increase accessibility and encourage participation from nontraditional students, working professionals, or those with caregiving responsibilities. Implementing these strategies can enhance inclusivity and variability, ultimately strengthening the generalizability of future research.

Second, the research design offered only a brief, single-session exposure to VR and non-VR platforms and varying character perspectives. This short exposure duration may not have allowed sufficient time for participants to fully immerse themselves in the VR experience or for the effects on empathy to fully manifest. This limited interaction may not fully capture the potential effects of extended VR-based training on empathy. Longer or repeated exposures could provide a more accurate assessment of VR's impact on empathy development. Therefore, longitudinal studies are recommended to investigate the long-term effects of VR on empathy, providing insights into the sustainability and potential long-term integration of VR into medical education.

Third, while we used self-reported scales that have been validated and widely used, these measures may introduce a certain level of response bias. Self-reported data can be influenced by social desirability or participants' subjective interpretations of the questions, which may affect the accuracy of the results. Future investigations could stand to gain significantly from incorporating more objective evaluative methods, such as physiological and behavioral observations, that would serve to substantiate the self-reported data. For example, using biometric measures like heart rate variability or skin conductance could provide objective insights into emotional engagement and empathy responses. In addition, behavioral assessments during simulated interactions could offer tangible evidence of empathy development. Implementing mixed-method approaches would mitigate reliance on self-reports and enhance the validity of future research findings.

Finally, although the usability of the VR interface is important, our study did not assess usability using standard questionnaires

XSL•FO
RenderX

such as the System Usability Scale [48] or the Usefulness, Satisfaction, and Ease of use [49] questionnaire. Future research could incorporate these usability measures to provide a more comprehensive evaluation of VR interfaces in educational settings. Usability testing plays a crucial role in ensuring that VR interventions in nursing education are intuitive, user-friendly, and meet the needs of learners. A system that is difficult to navigate or provides a poor user experience can disrupt engagement and limit the effectiveness of the intervention. Future research should integrate standardized usability assessments, such as the System Usability Scale or Usefulness, Satisfaction, and Ease of use questionnaire, to systematically evaluate the user experience. These measures will help identify areas for improvement, offering insights into how the VR interface aligns with learning goals and how usability affects student engagement and comprehension. By incorporating such evaluations, design improvements can be made—whether refining the interface, enhancing interaction features, or adjusting the VR experience to better suit diverse learning styles. Ultimately, addressing usability issues can improve the practical application of VR in nursing education, ensuring that immersive learning experiences are both accessible and impactful for real-world clinical practice.

Conclusion

In this study, we explored the interactive roles of media platforms and perspective-taking in shaping key psychological factors, including perceived self-location, narrative transportation, emotional engagement, and affective empathy, in nursing education settings. The initial findings provide empirical evidence for the potential of immersive technologies as applicable pedagogical tools, particularly in teaching and training future health care providers. The capacity of virtual reality to facilitate the feeling of presence, build emotional engagement, and foster empathy among nursing students shows the potential to foster a greater degree of patient-centered care. Therefore, this study advocates for a wider consideration of integrating technologies into health care education curriculum design and development. The potential benefits and financial viability of VR technologies could enrich pedagogical experiences and pave the way for the emergence of competent health care professionals, well-equipped to navigate different scenarios in health care delivery.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Analysis of covariance (ANCOVA) results for multiple outcome measures.
[DOCX File, 17 KB - mededu_v11i1e59083_app1.docx]

References

1. Gupta A, Cecil J, Pirela-Cruz M, Ramanathan P. A virtual reality enhanced cyber-human framework for orthopedic surgical training. IEEE Systems Journal 2019;13(3):3501-3512. [doi: 10.1109/JSYST.2019.2896061]
2. Jiang H, Vimalesvaran S, Wang JK, Lim KB, Mogali SR, Car LT. Virtual reality in medical students' education: scoping review. JMIR Med Educ 2022 Feb 2;8(1):e34860. [doi: 10.2196/34860] [Medline: 35107421]
3. González Izard S, Juanes Méndez JA, Ruisoto Palomera P, García-Peñalvo FJ. Applications of virtual and augmented reality in biomedical imaging. J Med Syst 2019 Mar 14;43(4):1-5. [doi: 10.1007/s10916-019-1239-z] [Medline: 30874965]
4. Manuel JK, Purcell N, Abadjian L, et al. Virtual worlds technology to enhance training for primary care providers in assessment and management of posttraumatic stress disorder using motivational interviewing: pilot randomized controlled trial. JMIR Med Educ 2023 Aug 28;9:e42862. [doi: 10.2196/42862] [Medline: 37639299]
5. Dyer E, Swartzlander BJ, Gugliucci MR. Using virtual reality in medical education to teach empathy. J Med Libr Assoc 2018 Oct;106(4):498-500. [doi: 10.5195/jmla.2018.518] [Medline: 30271295]
6. Helle N, Vikman MD, Dahl-Michelsen T, Lie SS. Health care and social work students' experiences with a virtual reality simulation learning activity: qualitative study. JMIR Med Educ 2023 Sep 20;9:e49372. [doi: 10.2196/49372] [Medline: 37728988]
7. Cummings JJ, Cahill TJ, Wertz E, Zhong Q. Psychological predictors of consumer-level virtual reality technology adoption and usage. Virtual Real 2023;27(2):1357-1379. [doi: 10.1007/s10055-022-00736-1] [Medline: 36597421]
8. Brown L, Ilhan E, Pacey V, Hau W, Van Der Kooi V, Dale M. The effect of high-fidelity simulation–based learning in acute cardiorespiratory physical therapy—a mixed-methods systematic review. JOPTE 2021;35(2):146-158. [doi: 10.1097/JTE.0000000000000183]
9. Liu JYW, Yin YH, Kor PPK, et al. The effects of immersive virtual reality applications on enhancing the learning outcomes of undergraduate health care students: systematic review with meta-synthesis. J Med Internet Res 2023 Mar 6;25:e39989. [doi: 10.2196/39989] [Medline: 36877550]
10. Saab MM, Hegarty J, Murphy D, Landers M. Incorporating virtual reality in nurse education: a qualitative study of nursing students' perspectives. Nurse Educ Today 2021 Oct;105:105045. [doi: 10.1016/j.nedt.2021.105045] [Medline: 34245956]
11. Plotzky C, Lindwedel U, Sorber M, et al. Virtual reality simulations in nurse education: a systematic mapping review. Nurse Educ Today 2021 Jun;101:104868. [doi: 10.1016/j.nedt.2021.104868] [Medline: 33798987]

XSL•FO
RenderX

12. Dean S, Halpern J, McAllister M, Lazenby M. Nursing education, virtual reality and empathy? Nurs Open 2020 Nov;7(6):2056-2059. [doi: 10.1002/nop2.551] [Medline: 33072391]
13. Wu D, Lowry PB, Zhang D, Tao Y. Patient trust in physicians matters-understanding the role of a mobile patient education system and patient-physician communication in improving patient adherence behavior: field study. J Med Internet Res 2022 Dec 20;24(12):e42941. [doi: 10.2196/42941] [Medline: 36538351]
14. Wu DC, Zhao X, Wu J. Online physician-patient interaction and patient satisfaction: empirical study of the internet hospital service. J Med Internet Res 2023 Aug 24;25:e39089. [doi: 10.2196/39089] [Medline: 37616031]
15. Yu J, Parsons GS, Lancastle D, Tonkin ET, Ganesh S. “Walking in Their Shoes”: the effects of an immersive digital story intervention on empathy in nursing students. Nurs Open 2021 Sep;8(5):2813-2823. [doi: 10.1002/nop2.860] [Medline: 33743185]
16. Calvert J, Abadia R. Impact of immersing university and high school students in educational linear narratives using virtual reality technology. Comput Educ 2020 Dec;159:104005. [doi: 10.1016/j.compedu.2020.104005]
17. Yeo JY, Nam H, Park JI, Han SY. Multidisciplinary design-based multimodal virtual reality simulation in nursing education: mixed methods study. JMIR Med Educ 2024 Jul 26;10:e53106. [doi: 10.2196/53106] [Medline: 39058550]
18. Chan K, Kor PPK, Liu JYW, Cheung K, Lai T, Kwan RYC. The use of immersive virtual reality training for developing nontechnical skills among nursing students: multimethods study. Asian Pac Isl Nurs J 2024 Jul 10;8:e58818. [doi: 10.2196/58818] [Medline: 38986130]
19. Son H, Ross A, Mendoza-Tirado E, Lee LJ. Virtual reality in clinical practice and research: viewpoint on novel applications for nursing. JMIR Nurs 2022 Mar 16;5(1):e34036. [doi: 10.2196/34036] [Medline: 35293870]
20. Buchanan C, Howitt ML, Wilson R, Booth RG, Risling T, Bamford M. Predicted influences of artificial intelligence on nursing education: scoping review. JMIR Nurs 2021;4(1):e23933. [doi: 10.2196/23933] [Medline: 34345794]
21. Lange AK, Koch J, Beck A, et al. Learning with virtual reality in nursing education: qualitative interview study among nursing students using the unified theory of acceptance and use of technology model. JMIR Nurs 2020;3(1):e20249. [doi: 10.2196/20249] [Medline: 34345791]
22. Ma Z, Huang KT, Yao L. Feasibility of a computer role-playing game to promote empathy in nursing students: the role of immersiveness and perspective. Cyberpsychol Behav Soc Netw 2021 Nov;24(11):750-755. [doi: 10.1089/cyber.2020.0371] [Medline: 33989057]
23. Ventura S, Badenes-Ribera L, Herrero R, Cebolla A, Galiana L, Baños R. Virtual reality as a medium to elicit empathy: a meta-analysis. Cyberpsychol Behav Soc Netw 2020 Oct;23(10):667-676. [doi: 10.1089/cyber.2019.0681] [Medline: 32757952]
24. Cummings JJ, Tsay-Vogel M, Cahill TJ, Zhang L. Effects of immersive storytelling on affective, cognitive, and associative empathy: The mediating role of presence. New Media Soc 2022 Sep;24(9):2003-2026. [doi: 10.1177/1461444820986816]
25. Sundar SS, Kang J, Oprean D. Being there in the midst of the story: how immersive journalism affects our perceptions and cognitions. Cyberpsychol Behav Soc Netw 2017 Nov;20(11):672-682. [doi: 10.1089/cyber.2017.0271] [Medline: 29125787]
26. Herrera F, Bailenson J, Weisz E, Ogle E, Zaki J. Building long-term empathy: A large-scale comparison of traditional and virtual reality perspective-taking. PLoS One 2018;13(10):e0204494. [doi: 10.1371/journal.pone.0204494] [Medline: 30332407]
27. Ma Z, Ma R, Chen M, Walter N. Present, empathetic, and persuaded: a meta-analytic comparison of storytelling in high versus low immersive mediated environments. Hum Commun Res 2023 Dec 26;50(1):27-38. [doi: 10.1093/hcr/hqad030]
28. Green MC, Brock TC. The role of transportation in the persuasiveness of public narratives. J Pers Soc Psychol 2000 Nov;79(5):701-721. [doi: 10.1037//0022-3514.79.5.701] [Medline: 11079236]
29. Ma Z, Yang G. Show me a photo of the character: exploring the interaction between text and visuals in narrative persuasion. J Health Commun 2022 Feb 1;27(2):125-133. [doi: 10.1080/10810730.2022.2065387] [Medline: 35422202]
30. Green MC, Tesser A, Wood JV, Stapel DA. Transportation into narrative worlds: implications for the self. In: On Building, Defending and Regulating the Self: A Psychological Perspective: Psychology Press; 2005:53-75. [doi: 10.4324/9780203998052]
31. Ahn SJ(, Bailenson JN, Park D. Short- and long-term effects of embodied experiences in immersive virtual environments on environmental locus of control and behavior. Comput Human Behav 2014 Oct;39:235-245. [doi: 10.1016/j.chb.2014.07.025]
32. Kors MJ, van der Spek ED, Ferri G, Schouten BA. You; the observer, partaker or victim. delineating three perspectives to empathic engagement in persuasive games using immersive technologies. Presented at: CHI PLAY ’18 Extended Abstracts: Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in PLAY Companion Extended Abstracts; Oct 28-31, 2018; Melbourne, VIC, Australia. [doi: 10.1145/3270316.3271547]
33. Ma Z, Zytko D. Designing immersive stories for health: choosing character perspective based on the viewer’s modality. International Journal of Human–Computer Interaction 2021 Sep 14;37(15):1423-1435. [doi: 10.1080/10447318.2021.1886486]
34. Beverly E, Rigot B, Love C, Love M. Perspectives of 360-degree cinematic virtual reality: interview study among health care professionals. JMIR Med Educ 2022 Apr 29;8(2):e32657. [doi: 10.2196/32657] [Medline: 35486427]
35. Hoeken H, Kolthoff M, Sanders J. Story perspective and character similarity as drivers of identification and narrative persuasion. Hum Commun Res 2016 Apr;42(2):292-311 [FREE Full text] [doi: 10.1111/hcre.12076]

XSL•FO
RenderX

36. Jin SAA. “I feel present. therefore, I experience flow:” a structural equation modeling approach to flow and presence in video games. J Broadcast Electron Media 2011 Feb 25;55(1):114-136. [doi: 10.1080/08838151.2011.546248]
37. Weibel D, Wissmath B. Immersion in computer games: the role of spatial presence and flow. International Journal of Computer Games Technology 2011;2011:1-14. [doi: 10.1155/2011/282345]
38. Ma Z. Effects of immersive stories on prosocial attitudes and willingness to help: testing psychological mechanisms. Media Psychol 2020 Nov 1;23(6):865-890. [doi: 10.1080/15213269.2019.1651655]
39. Liu S, Yang JZ. Incorporating message framing into narrative persuasion to curb e-cigarette use among college students. Risk Anal 2020 Aug;40(8):1677-1690. [doi: 10.1111/risa.13502] [Medline: 32390210]
40. Murphy ST, Frank LB, Moran MB, Patnoe-Woodley P. Involved, transported, or emotional? exploring the determinants of change in knowledge, attitudes, and behavior in entertainment-education. J Commun 2011 Jun;61(3):407-431. [doi: 10.1111/j.1460-2466.2011.01554.x]
41. That dragon, cancer. Numinous Games. 2016. URL: https://www.thatdragoncancer.com/ [accessed 2025-03-18]
42. Chen A, Hanna JJ, Manohar A, Tobia A. Teaching empathy: the implementation of a video game into a psychiatry clerkship curriculum. Acad Psychiatry 2018 Jun;42(3):362-365. [doi: 10.1007/s40596-017-0862-6] [Medline: 29204755]
43. Martingano AJ, Hererra F, Konrath S. Virtual reality improves emotional but not cognitive empathy: A meta-analysis. Technol Mind Behav 2021;2(1). [doi: 10.1037/tmb0000034]
44. Hartmann T, Wirth W, Schramm H, et al. The spatial presence experience scale (SPES). J Media Psychol 2016 Jan;28(1):1-15. [doi: 10.1027/1864-1105/a000137]
45. Knol J, van Linge R. Innovative behaviour: the effect of structural and psychological empowerment on nurses. J Adv Nurs 2009 Feb;65(2):359-370. [doi: 10.1111/j.1365-2648.2008.04876.x] [Medline: 19191936]
46. Batson CD, Fultz J, Schoenrade PA. Distress and empathy: two qualitatively distinct vicarious emotions with different motivational consequences. J Pers 1987 Mar;55(1):19-39. [doi: 10.1111/j.1467-6494.1987.tb00426.x] [Medline: 3572705]
47. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach: Guilford Publications; 2017.
48. Vlachogianni P, Tselios N. Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review. J Res Technol Educ 2022 May 27;54(3):392-409. [doi: 10.1080/15391523.2020.1867938]
49. Gao M, Kortum P, Oswald F. Psychometric evaluation of the USE (usefulness, satisfaction, and ease of USE) questionnaire for reliability and validity. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2018 Sep;62(1):1414-1418. [doi: 10.1177/1541931218621322]

Abbreviations

ANCOVA: analysis of covariance
VR: virtual reality

Edited by A Hasan Sapci; submitted 02.04.24; peer-reviewed by B Garrett, CV Gomez, S Benbelkacem; revised version received 07.11.24; accepted 02.01.25; published 31.03.25.

Please cite as:
Huang KT, Ma Z, Yao L
Media-Induced and Psychological Factors That Foster Empathy Through Virtual Reality in Nursing Education: 2×2 Between-Subjects Experimental Study
JMIR Med Educ 2025;11:e59083
URL: https://mededu.jmir.org/2025/1/e59083
doi:10.2196/59083

© Kuo-Ting Huang, Zexin Ma, Lan Yao. Originally published in JMIR Medical Education (https://mededu.jmir.org), 31.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Case-Based Virtual Reality Simulation for Severe Pelvic Trauma Clinical Skill Training in Medical Students: Design and Pilot Study

Peng Teng[1], MM; Youran Xu[2], BS; Kaoliang Qian[3], BS; Ming Lu[4*], MD; Jun Hu[3*], MD

[1]Department of Teaching Resources Management, Teaching Management Office of Nanjing Medical University, Nanjing, China
[2]School of Stomatology, Nanjing Medical University, Nanjing, China
[3]Department of Orthopedics, First Affiliated Hospital of Nanjing Medical University, Nanjing, China
[4]Department of Pharmacology & Jiangsu Key Laboratory of Neurodegeneration, Nanjing Medical University, Nanjing, China
[*]these authors contributed equally

Corresponding Author:
Jun Hu, MD
Department of Orthopedics
First Affiliated Hospital of Nanjing Medical University
Guang Zhou Road 300
Nanjing,
China
Phone: 86 02568303196
Email: junhu89@vip.sina.com

Abstract

Background: Teaching severe pelvic trauma poses a significant challenge in orthopedic surgery education due to the necessity of both clinical reasoning and procedural operational skills for mastery. Traditional methods of instruction, including theoretical teaching and mannequin practice, face limitations due to the complexity, the unpredictability of treatment scenarios, the scarcity of typical cases, and the abstract nature of traditional teaching, all of which impede students' knowledge acquisition.

Objective: This study aims to introduce a novel experimental teaching methodology for severe pelvic trauma, integrating virtual reality (VR) technology as a potent adjunct to existing teaching practices. It evaluates the acceptability, perceived ease of use, and perceived usefulness among users and investigates its impact on knowledge, skills, and confidence in managing severe pelvic trauma before and after engaging with the software.

Methods: A self-designed questionnaire was distributed to 40 students, and qualitative interviews were conducted with 10 teachers to assess the applicability and acceptability. A 1-group pretest-posttest design was used to evaluate learning outcomes across various domains, including diagnosis and treatment, preliminary diagnosis, disease treatment sequencing, emergency management of hemorrhagic shock, and external fixation of pelvic fractures.

Results: A total of 40 students underwent training, with 95% (n=38) affirming that the software effectively simulated real-patient scenarios. All participants (n=40, 100%) reported that completing the simulation necessitated making the same decisions as doctors in real life and found the VR simulation interesting and useful. Teacher interviews revealed that 90% (9/10) recognized the VR simulation's ability to replicate complex clinical cases, resulting in enhanced training effectiveness. Notably, there was a significant improvement in the overall scores for managing hemorrhagic shock (t_{39}=37.6; 95% CI 43.6-48.6; P<.001) and performing external fixation of pelvic fractures (t_{39}=24.1; 95% CI 53.4-63.3; P<.001) from pre- to postsimulation.

Conclusions: The introduced case-based VR simulation of skill-training methodology positively influences medical students' clinical reasoning, operative skills, and self-confidence. It offers an efficient strategy for conserving resources while providing quality education for both educators and learners.

(JMIR Med Educ 2025;11:e59850) doi:10.2196/59850

KEYWORDS

case-based learning; virtual reality; pelvic fracture; severe pelvic trauma; hemodynamic instability; clinical skill training; VR; pelvic trauma; medical student; pilot study; orthopedic surgery; theoretical teaching; acceptability

XSL•FO
RenderX

Introduction

Severe pelvic trauma is characterized by unstable pelvic fractures resulting from high-energy impacts, typically accompanied by complications such as fatal massive bleeding and organ injuries. The mortality rate in China is as high as 10%-30% [1-3]. Managing pelvic fractures with hemodynamic instability poses a significant challenge within the orthopedic surgery discipline [4,5]. The current gap that exists in the field of the effectiveness of severe pelvic trauma clinical skill training is multifaceted. It includes the inaccessibility of clinical teaching in hospitals and the constraints of traditional classroom instruction. In addition, owing to the complexity, unpredictability of treatment locations, the rarity of typical cases, reluctance to cooperate, and ethical concerns surrounding clinical teaching, traditional methods of clinical skill training in diagnosing and treating severe pelvic trauma are limited to theoretical instruction and mannequin-based simulation. These methods face limitations, including disproportionate teacher-student ratios, high model consumption, insufficient training spaces, the absence of comprehensive and immediate feedback, and inadequate training overall. Specifically, on-site mentoring can be challenging to achieve effectively and efficiently. Therefore, a teaching model for severe pelvic trauma needs to integrate new teaching strategies and computer technology to address these issues. This study addresses gaps in the current training methods and tools for managing severe pelvic trauma by designing an innovative virtual reality (VR)–based simulation platform.

In recent years, VR technology has gained widespread acceptance in orthopedic surgery education because of its multisensory immersive experience, the convenience of real-time interaction, and a psychologically secure experimental environment [6,7]. These technologies have mitigated the limitations of time, space, and teaching resources in orthopedic surgery education and addressed issues such as simulated patients and the difficulty of repeated practice. To some extent, this has enhanced the efficiency and quality of clinical practice. Digital patient simulators have become valuable tools in medical education, offering a standardized method of patient simulation [8,9]. However, virtual patients typically only exhibit clinical symptoms and signs, with minimal explanation of the underlying fundamental medical knowledge of primary symptoms and positive or negative signs. Furthermore, there is a scarcity of research on virtual simulation teaching models that integrate basic and clinical medicine for severe pelvic trauma. To mimic the real teaching scenario of the disease and capture the sudden and variable condition of patients in clinical settings, we developed an electronic standardized patient (ESP) for severe pelvic trauma, capable of replicating both macroscopic changes, such as monitor display data, and microscopic changes, including alterations in blood circulation, organs, tissues, and cells. The ESP model facilitates the integration and application of clinical and basic knowledge. The original design and technology of the ESP were conceived by Professor Xing Ya Gao's team at Nanjing Medical University, Department of Physiology [10]. An ESP is grounded in contemporary theories of human systems physiology and incorporates relevant clinical literature and data through analog circuits, physics, and other methods to formulate a mathematical model. Artificial intelligence and data analytics are used to refine and adjust the simulation data. In summary, the ESP represents a web-based intelligent standardized patient, enabling students to interact with it in a virtual hospital setting from a physician's perspective. To date, the ESP has not been fully used in the clinical skills education for severe pelvic trauma.

Case-based learning (CBL) is recognized as an effective teaching strategy in clinical skills training [11,12]. CBL is particularly valued in orthopedic surgery education for its ability to enhance learner participation, foster active learning, and develop critical thinking and problem-solving skills [13]. CBL is an educational method designed to analyze medical records, recreate real clinical scenarios, and engage learners in addressing actual clinical challenges, thereby stimulating their curiosity and promoting active learning [14].

CBL, combined with VR simulation technology, has been successfully applied in midwifery laboratory courses, with its effectiveness widely acknowledged by students [15]. However, the application of these combined methodologies in teaching clinical skills for severe pelvic trauma has yet to be explored. Therefore, in this study, we introduced a case-based digital skill training program for severe pelvic trauma and conducted mixed methods to evaluate its acceptance among users. Additionally, we implemented a pretest-posttest design to investigate the potential impact of this clinical skills training on undergraduate and graduate students. We hypothesize that the case-based VR simulation teaching method will effectively complement current training practices for severe pelvic trauma, enhancing knowledge, procedural skills, and confidence, while also improving instructional efficiency and effectiveness for educators.

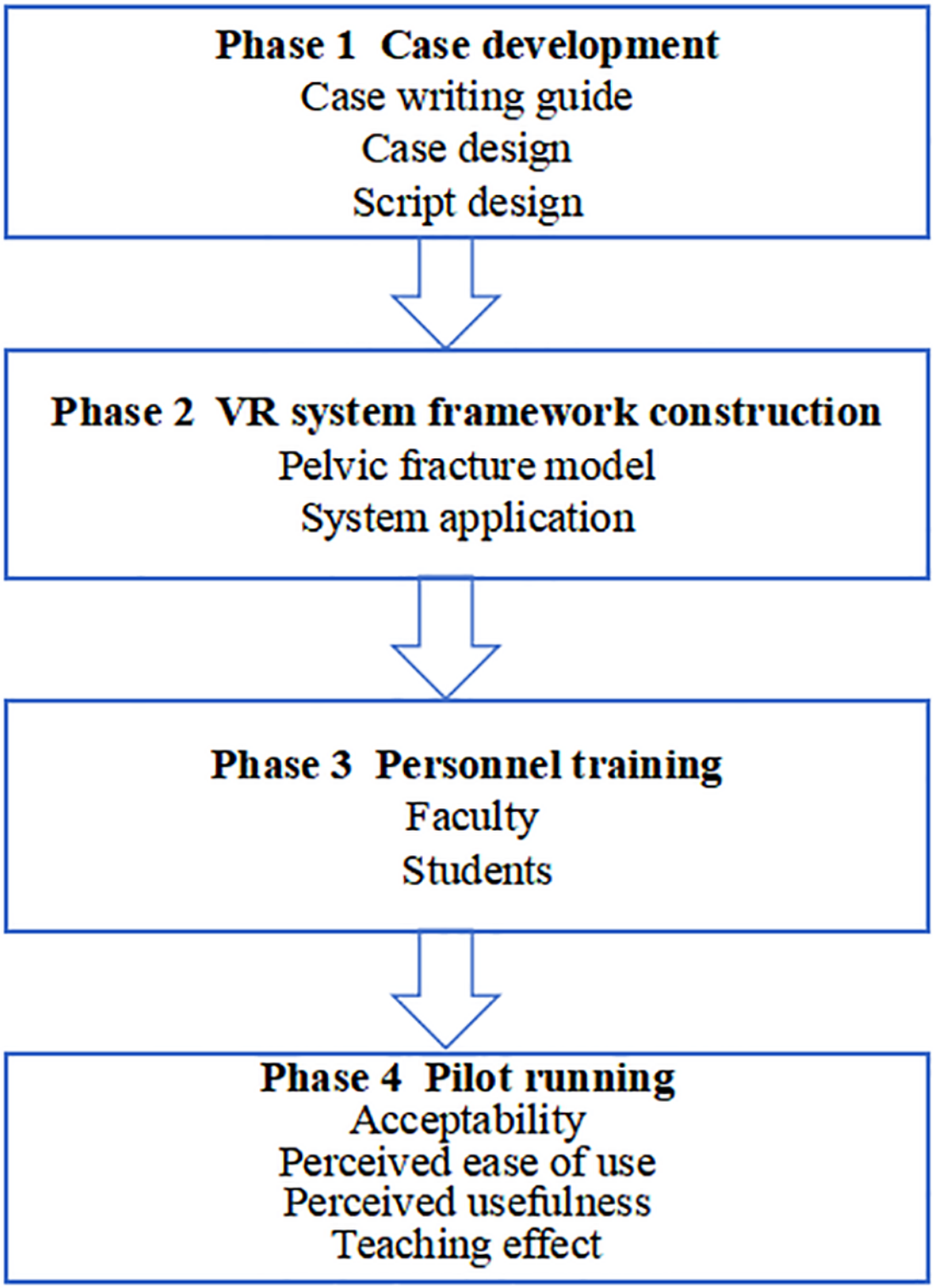
Methods

Study Design

We conducted a 4-phase study (Figure 1). First, we created the simulated teaching case base and adapted typical clinical cases based on the case writing guide [16,17]; subsequently, the case script was developed. Next, we established the framework of the VR system, comprising 3 ports and 3 system modules. To facilitate the effective implementation of the teaching system, we formulated a comprehensive training plan for department administrators, course instructors, and medical students. A pilot test of the system was conducted on a limited scale, followed by a 1-group pretest-posttest design to assess its acceptability, potential application, and existing limitations.

XSL•FO
RenderX

Figure 1. A flow diagram illustrating the steps involved in developing a case-based VR simulation for severe pelvic trauma clinical skill training. VR: virtual reality.



Phase 1: Case Development

Case Design

Specialists in basic and clinical medicine, drawing on a literature review, real patient cases, and the course syllabus, designed the cases. A representative case involved a worker who fell from a high platform and was sequentially evaluated in the emergency department, admitted to the intensive care unit, and taken to the operating room as his condition worsened. The diagnostic and treatment processes for this condition were standardized. Learners were guided through different scenarios to acquire both declarative and procedural knowledge for managing patients with severe pelvic trauma. We developed 3 initial cases, each representing a different stage of trauma and treatment approach (Table 1). Learning paths for diagnosis and treatment were outlined in a flowchart (Figure 2), based on clinical

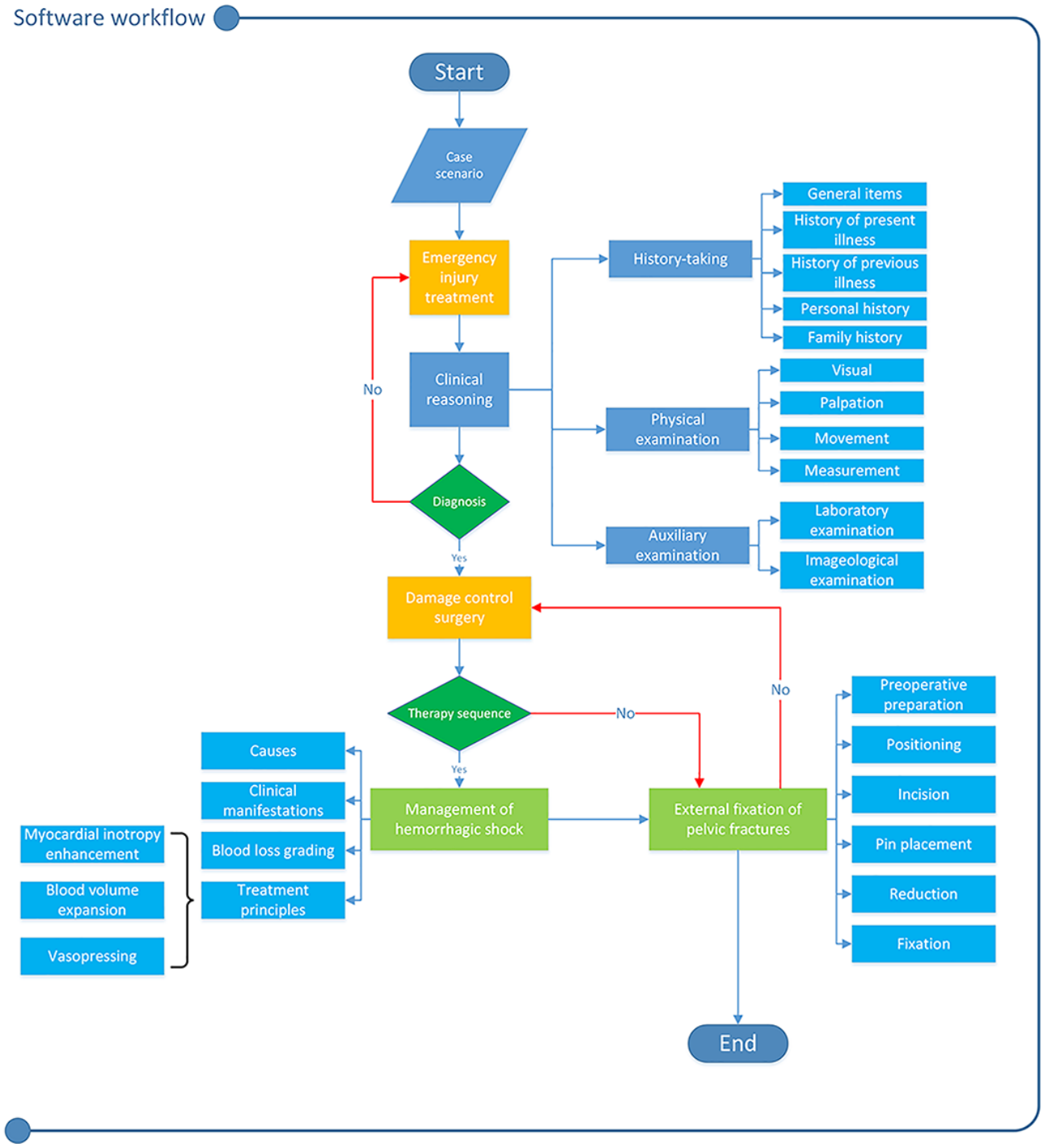
XSL•FO
RenderX

treatment guidelines [18,19] to serve as a framework for evaluating student performance. We collected clinical manifestations, relevant imaging, and laboratory examination results of real patients.

Table 1. Three cases with patients diagnosed with pelvic trauma.

	Case 1	Case 2	Case 3
Diagnosis	Pelvic fracture	Pelvic fracture with mild hemorrhagic shock	Pelvic fracture with severe hemorrhagic shock
Injury evaluation	Mild	Moderate	Severe
Diagnostics	Standard	Plus computed tomography of the pelvis	Plus computed tomography of the pelvis and abdomen
Therapy	Surgery	Antihemorrhagic shock therapy, surgery	Antihemorrhagic shock therapy, surgery

Figure 2. An optimal template for the diagnosis and treatment process of severe pelvic trauma (blue arrows for the correct path, red arrows for the wrong path).



XSL•FO
RenderX

The 3 cases varied in the severity of pelvic trauma, necessitating different diagnostic and treatment strategies.

Script Design

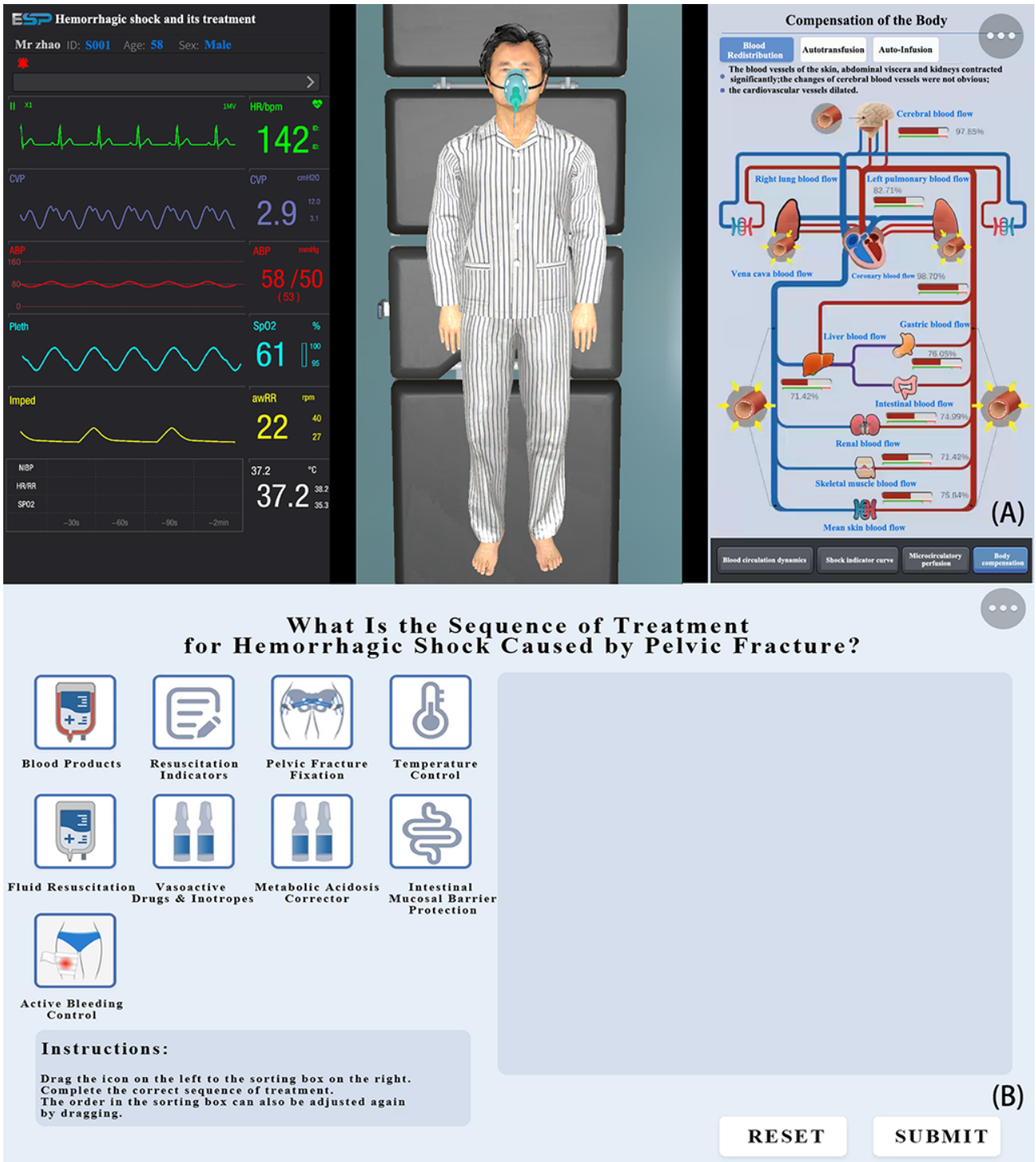
Scripts focused on scenarios in emergency rooms, intensive care units, and operating rooms. In a virtual emergency room, users could interact with the ESP, using various diagnostic and treatment methods, such as history taking, physical examination, and auxiliary examination, to formulate a preliminary diagnosis based on evidence gathered during the process. Learners engaged with the ESP through text input or voice chat during the history-taking session. They then entered pertinent information into the system's modules for general items, history of present illness, personal history, marital history, and family history. The system evaluated 4 aspects: completeness of the inquiry framework, logical order of inquiries, communication skills, and awareness of humanistic care, assigning scores based on the accuracy of the information collected. The system also recorded the total duration of inquiries and the time spent on each module for later analysis. Physical examination, a fundamental skill for diagnosing diseases, involves comparing and identifying positive signs. The system assessed the patient examination position, completeness and sequence of operational steps, standardization of procedures, comprehensiveness of examination content, accurate reporting of results, and basic professional quality. Auxiliary examination, crucial for diagnosis and treatment planning, involves evaluating laboratory tests and computed tomography images. Upon completing these steps, students made a preliminary diagnosis and a prioritized list of differential diagnoses. Correct diagnoses led to immediate treatment initiation; otherwise, students continued trial and error in this module (Figure 3).

Figure 3. Screenshot of the emergency room. (A) The user acts as a doctor in the emergency room while interacting with the ESP by selecting different diagnoses and treatment icons. (B) Trainees can choose the present history module in history taking to communicate with the ESP. (C) During the physical examination session, students can use a tape measure to measure the distance between the bilateral anterior superior iliac spine and the xiphoid process to determine whether the pelvis was displaced. (D) The students order a computed tomography scan to evaluate the location and severity of the ESP's injury. ESP: electronic standardized patient. Please note that a higher resolution version of this image can be found in Multimedia Appendix 1.

Upon initiating appropriate immediate treatment measures, the ESP was transferred to the intensive care unit. Vital sign monitors reflected real-time changes based on the condition and treatment progression. In this module, users learned about the causes, clinical manifestations, severity classifications, and management of hemorrhagic shock (Figure 4).

XSL•FO
RenderX

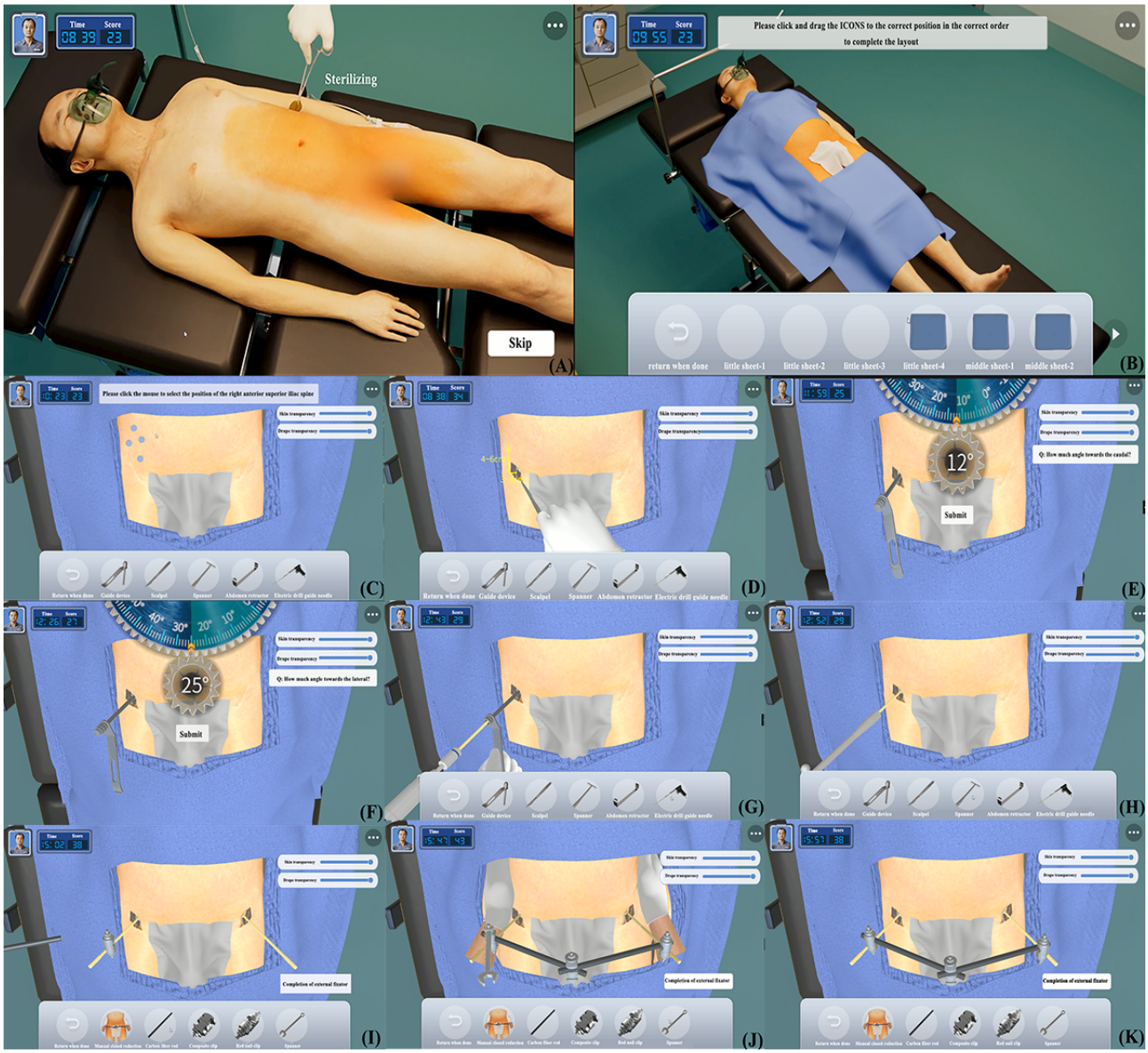
Figure 4. Screenshot of hemorrhagic shock and its treatment. (A) Monitors in the intensive care unit show decreased blood pressure, increased heart rate, and subnormal central venous pressure and arterial oxygen saturation in the ESP. By users' observation, the ESP at this time showed pallor, decreased urine output, and cold sweats. Users can also learn about the compensatory mechanisms of ESP in terms of blood redistribution, autotransfusion, and auto-infusion at the system, organ, and tissue levels after hemorrhagic shock. Screenshot of treatment principles for severe pelvic trauma. (B) In this module, the user should drag the corresponding icons from all the treatment measures given on the left to the right blank according to the treatment principle. When the dragged content matches the built-in answer of the system, the next section will be entered. ESP: electronic standardized patient. Please note that a higher resolution version of this image can be found in Multimedia Appendix 1.



During the surgical procedure for external fixation of a pelvic fracture, users first completed preparatory tasks such as handwashing and donning surgical attire. Then, the ESP underwent anesthesia, sterilization, positioning, drilling, and bracket installation according to surgical standards (Figure 5).

XSL•FO
RenderX

Figure 5. Screenshot of the surgical procedure of external fixation of pelvic fractures. (A) According to the sequence of operation, the users need to complete disinfection, (B) laying, (C) positioning, (D) incising, (E,F) positioning and protection of the guide, (G) electric insertion of screw, (H) manual insertion of screw, (I) use of rod-nail clamp and composite clamps, and (J,K) adjustment of the external fixator after manual reduction. Please note that a higher resolution version of this image can be found in Multimedia Appendix 1.



Phase 2: Framework Construction

Construction of Pelvic Fracture Model

To provide learners with a comprehensive understanding of pelvic anatomy and fracture morphology, we collaborated with software engineers to develop the relevant content. The development process involved several steps. Initially, 3D modeling data of pelvic fractures were extracted from real clinical cases, and a preliminary geometry was constructed using the Maya (Autodesk) development tool to complete the foundational model of pelvic fractures. Subsequently, the ZBrush (Pixologic) tool was used to sculpt and refine the fracture site shape, fracture line position, and bone surface structure of the basic pelvic fracture model, minimizing the detailed morphology of the pelvic fracture. Next, we used the support model subdivision level adjustment feature to enhance the pelvic fracture model's subdivision level, showcasing the complex structure of the fracture site. The addition of materials and textures in Maya, such as bone texture and skin color and texture, further improved the visual realism. The established skeletal system was then integrated into the pelvic fracture model, and animations were created based on real pelvic fracture case data, including the degree of fracture displacement and the relative position of fracture blocks. Finally, the pelvic fracture model was exported from Maya to Unity (Unity Technologies) for digital and real-time rendering, allowing learners to interact with a pelvic fracture model in a virtual environment to simulate actual surgical procedures. The development and construction process of the pelvic fracture model is elaborated in Multimedia Appendix 2.

Construction of System Application

The digital simulation system was segmented into 3 ports: the department administrator, course leader, and user ports, constructed on a website platform. The department's primary role is to create the appropriate courses in the system, either

XSL•FO
RenderX

compulsory or optional. Once the courses are established, they should be linked to the digital simulation skills training system, and relevant instructors assigned to teach all enrolled students for the semester. Course instructors are responsible for preparing preview materials, learning videos, self-assessment questions, and postclass surveys. They also review experimental outcomes and evaluate their performance within the system. Users can engage with the digital simulation system for severe pelvic trauma by enrolling in or attending a scheduled course. Upon system login, the interface offers 2 modes: training and assessment. Due to the digital simulation experiment teaching integration with the virtual simulation experiment teaching sharing platform of Nanjing Medical University, campus users are not required to register or authenticate before use, thanks to unified identity verification and experimental data integration.

Phase 3: Personnel Training

Faculty Training

Faculty training targeted course directors and instructors from various disciplines within basic and clinical medicine, aiming to familiarize them with the digital simulation experimental teaching system. They received training on case materials, experimental teaching objectives, principles, teaching processes and methods, steps, outcomes, and conclusions. Additionally, they learned to address technical system issues, respond to student inquiries, and interact with students on the platform during experiments.

Student Training

Before software utilization, students were informed about the operating system and hardware configuration requirements. The application runs on a Windows 7 64-bit or higher PC, equipped with a 3.60 GHz Intel i5 processor, 8 GB RAM, NVidia GTX 2060 graphic card, and a 1920×1080 display resolution. The application supports various browsers on different operating systems, such as Google Chrome, the 360 browser, and Firefox. After accessing a specific URL, users must install MengooLauncher, requiring less than 100 MB of plug-in capacity, as indicated. Before proceeding to the autonomous training or assessment interface, users familiarize themselves with the experimental teaching objectives and principles through introductory and instructional videos.

Phase 4: Pilot Running Evaluation of Digital Simulation Software

Design

A self-controlled teaching comparison study was conducted at Nanjing Medical University, China, from October 2023 to January 2024, to examine the impact on knowledge, skills, and confidence before and after using virtual simulation experimental teaching software. All participants underwent a knowledge assessment of equal difficulty before and after system engagement.

In the initial design phase, the teaching and research teams sought input from the software development team and feedback from various users through internal reviews. A case-based VR simulation of severe pelvic trauma was tested by students majoring in clinical medicine, clinical teachers, and basic medicine instructors. We distributed the ESP digital simulation teaching web link via WeChat (Tencent) to pertinent users, soliciting face-to-face or written feedback on case and script design and software development, including aspects related to clinical and basic medicine education. Additionally, the teaching research group reviewed classical cases of severe pelvic trauma and questionnaire responses.

Sampling and Recruitment

The recruitment criteria were as follows: (1) students must have completed courses in diagnostics, internal medicine, and surgery; (2) they needed to have a laptop for the study; (3) they should not have participated in any form of digital simulation software training for clinical skills prior; and (4) they agreed to participate in the pilot study and signed an informed consent form. We invited a purposive sample of 20 fourth-year undergraduates and 20 first-year graduate students majoring in clinical medicine from the First Clinical Medical College of Nanjing Medical University to test the case-based VR simulation software. Based on sample size requirements previously reported in the literature for evaluating data collection materials, a minimum of 10 samples is necessary to ensure the adequacy and validity of the assessment instrument [20]. The evaluation sought feedback from a diverse group of users, including undergraduate and graduate students, as well as teachers with various professional titles. The qualitative study used a representative population most familiar with the study topic, comprising 5 orthopedic teaching teachers and 5 basic medicine teachers.

Data Collection

Participants were required to complete the training and assessment using the digital simulation software for severe pelvic trauma treatment. Our data collection involved a repeated measurement approach to assess knowledge test scores before and after the simulation. Feedback on the simulation teaching tools was collected through a single questionnaire. Participation in the survey was entirely voluntary, and students were informed that their decision to participate would not impact their academic standing. The survey distribution was conducted independently, with no direct or known ties between the distributors and the students, ensuring an unbiased and pressure-free environment for participants. Participants were recruited and invited to complete the survey through the WeChat platform using the Wenjuanxing applet. The survey was administered anonymously to encourage honest feedback on their experience with the VR simulation software.

Outcome Assessment

Although evaluation questionnaires are commonly used to compare learning tools, there is a dearth of validated tools for assessing the ESP digital simulation software as a learning instrument. Consequently, the questionnaire was adapted from a validated assessment tool in educational literature, offering a resource for future research on the perception in clinical medical professional education. The questionnaire comprised 15 Likert-scale statements (1=strongly disagree to 5=strongly agree), assessing accessibility and usability. The teaching team and subject experts reviewed the questionnaire. Cronbach α for the questionnaire was 0.85 (n=15). The questionnaire was

XSL•FO
RenderX

adapted from a validated assessment tool widely used in educational literature for clinical medical professional education [21]. The instrument used for assessing system acceptability was based on a modified version of the Technology Acceptance Model questionnaire, with validation provided by Balki et al [22]. The Cronbach α coefficient, which reflects internal consistency, was calculated jointly for both the usability and acceptability surveys, yielding a consistent value for both aspects.

Data Analysis

Descriptive analysis was applied to the quantitative data obtained from the Likert scale. For qualitative data, which included responses to 7 open-ended questions, we used a validated content analysis method as described by Elo and Kyngäs [23]. All participant comments were transcribed and imported into Excel (Microsoft Corp) for coding. The content analysis process consists of several steps: (1) familiarizing oneself with the data and the hermeneutic spiral, (2) dividing up the text into meaning units and subsequently condensing those meaning units, (3) formulating codes, and (4) developing categories and themes [24]. Initially, the primary investigator analyzed the content, and the research team subsequently reviewed and discussed the codes to achieve consensus. In cases of disagreement, group discussions were held, and, if necessary, a third-party opinion was sought to ensure triangulation and enhance reliability. The 7 open-ended survey questions are provided in Multimedia Appendix 3 for reference.

For user acceptance analysis of the ESP platform among undergraduates, graduates, and tutors, descriptive statistics were used. To compare the mean rating scales of each survey item between groups, an independent 2-tailed *t* test was performed with a significance threshold of $P<.05$. Prior to conducting parametric tests, we verified the assumption of normality using the Shapiro-Wilk test, confirming that the data met the requirements for a parametric approach. This method was chosen over nonparametric tests due to the normal distribution of the data, making it suitable for our sample size and study design.

Ethical Considerations

The Nanjing Medical University ethics committee approved this study (2023418). During the informed consent process, participants were made aware that no incentives were provided for participation in the survey. All methods were implemented in accordance with the Helsinki declaration. All participants were voluntary in the study.

Results

Demographic Results

Of the 56 students enrolled in the optional course on the integrated case of severe pelvic trauma in October 2023, 40 students consented to participate in the pilot study. Among these participants, 50% (n=20) of the students were senior-year undergraduates, 50% (n=20) of the students were first-year graduates (n=20), 45% (n=18) of students were men, and 55% (n=22) of the students were women. The mean age was 22.9 (SD 1.3) years. Among the undergraduate participants (n=20), there were 10 male and 10 female students, with a mean age of 21.9 (SD 0.9) years. For the graduate participants (n=20), there were 8 male and 12 female students, with a mean age of 24.0 (SD 0.8) years. Ten faculty members with at least 5 years of teaching experience in orthopedic surgery or basic medicine were also invited to participate. Neither the students nor the faculty had prior experience with this type of digital simulation platform. All participants were required to complete the questionnaire shortly after finishing the training tasks.

Questionnaire Results

A 5-point Likert scale assessed perceptions of the acceptability, effectiveness, and applicability, summarized in Table 2.

XSL•FO
RenderX

Table 2. Average rating scores of survey questions (Q1-Q15) by all students.

Survey questions	Average rating scores, mean (SD)	
	Undergraduate students	Graduate students
(1) The digital software provides a simulation of a real patient	4.40 (0.66)[a]	4.80 (0.40)
(2) During the simulation, I felt like a doctor taking care of this patient	3.90 (0.99)	3.90 (1.14)
(3) When I finished the simulation, I felt I had to make the same decisions as doctors in real life	4.55 (0.50)	4.80 (0.40)
(4) The VR[b] simulation is interesting and useful	4.85 (0.36)	4.70 (0.46)
(5) The difficulty of the VR simulation is appropriate to my own level of knowledge and skills	4.10 (0.94)[a]	4.60 (0.49)
(6) The feedback from the system adequately reflected my actual performance	4.80 (0.40)	4.60 (0.66)
(7) The goals of scenario simulation are clear and easy to understand	4.55 (0.59)	4.80 (0.40)
(8) I can access the system anytime and anywhere for simulation training	4.75 (0.54)	4.90 (0.30)
(9) The VR simulation can help me to use basic medical knowledge to explain clinical manifestations of clinical reasoning skills	4.80 (0.40)[c]	4.20 (0.87)
(10) The ESP[d] simulator can help me develop clinical operation skills	4.30 (0.78)	4.60 (0.49)
(11) I feel more confident about working with hospital colleagues	4.40 (0.80)	4.70 (0.46)
(12) The VR simulation increased my confidence as a practicing physician	4.40 (0.66)	4.60 (0.49)
(13) The VR simulation can support courses and exams	4.60 (0.49)	4.75 (0.43)
(14) Compared with traditional teaching practice training methods, VR simulation can reduce my training cost and risk	4.90 (0.30)[c]	4.50 (0.50)
(15) In general, this VR simulation training should enhance my learning	4.80 (0.40)	4.80 (0.40)

[a]$P<.05$ compared to the graduate student group.

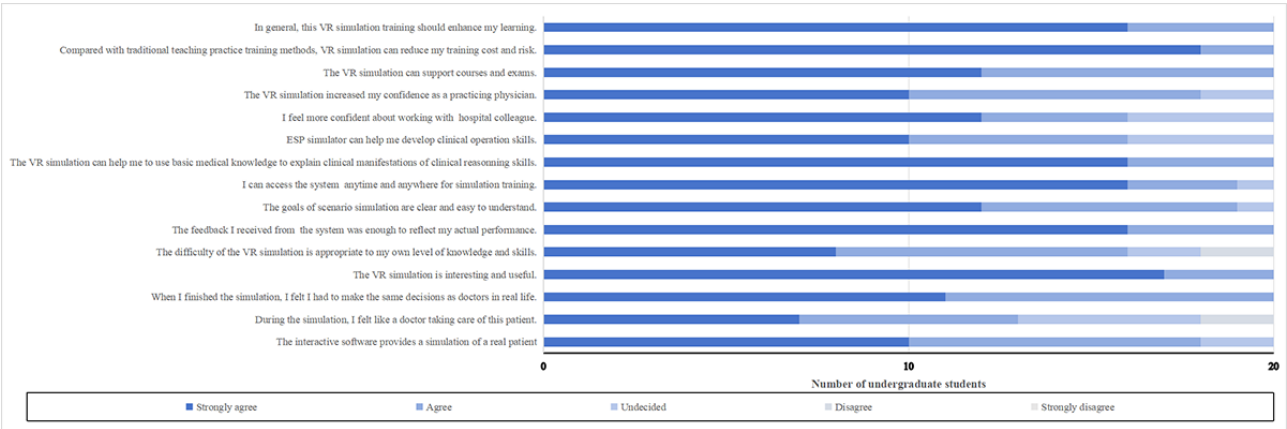
[b]VR: virtual reality.

[c]$P<.01$ compared with the graduate student group.

[d]ESP: electronic standardized patient.

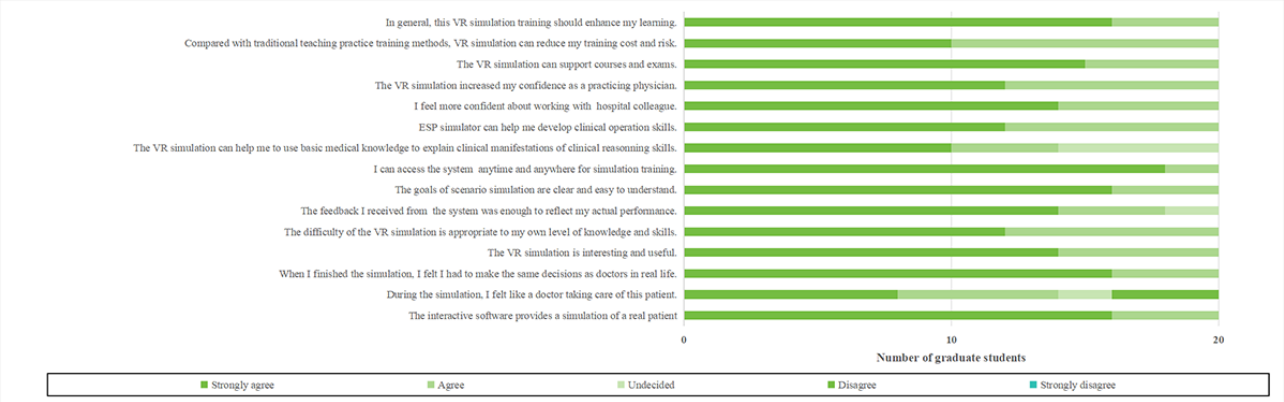
The respondents showed strong agreement; 95% (n=38) agreed or strongly agreed that the interactive software simulated a real patient scenario (Q1 in Figures 6 and 7). However, 68% (n=27) agreed or strongly agreed that "During the simulation, I felt like a doctor caring for this patient" (Q2), with 18% (n=7) neutral and 15% (n=6) disagreeing. All students (n=40, 100%) felt they had to make real-life doctor decisions by the end of the simulation (Q3) and found the VR simulation interesting and useful (Q4).

Figure 6. Acceptability, effectiveness, and applicability of the case-based VR software by undergraduate students. ESP: electronic standardized patient; VR: virtual reality. Please note that a higher resolution version of this image can be found in Multimedia Appendix 1.



XSL•FO
RenderX

Figure 7. Acceptability, effectiveness, and applicability of the case-based VR software by graduate students. ESP: electronic standardized patient; VR: virtual reality. Please note that a higher resolution version of this image can be found in Multimedia Appendix 1.



Additionally, 90% (n=36) believed the VR simulation's difficulty was appropriate for their knowledge and skills (Q5), while 5% (n=2) disagreed. Moreover, 95% (n=38) reported that the feedback from the system sufficiently reflected their performance (Q6). Most students (95%, n=38) understood the goals of the scenario simulation clearly (Q7). Nearly all students (98%, n=39) could access the system anytime for training (Q8), and 98% (n=39) agreed that "The VR simulation can help apply basic medical knowledge to clinical reasoning skills" (Q9).

When inquired if the ESP simulator aided in developing clinical operational skills (Q10), 85% (n=34) agreed. Regarding confidence in collaborating with hospital colleagues (Q11) and functioning as practicing physicians (Q12), 90% (n=36) agreed or strongly agreed. All students (n=40) concurred that the VR simulation supports courses and exams (Q13), is cost-effective compared to traditional training (Q14), and enhances learning overall (Q15).

Impact of Training Level on Questionnaire Answers

Finally, a 2-tailed *t* test was used to compare the average rating scales between undergraduate and graduate students. This analysis aimed to determine if the responses varied according to their academic level. Specifically, for Questions 1 and 5, undergraduate students exhibited significantly stronger disagreement than their graduate counterparts, as indicated by the *P* values (Q1, *P*=.03; Q5, *P*=.047). Moreover, when compared to graduate students, a larger proportion of undergraduates believed that VR simulation could enhance their clinical reasoning abilities (Q9, *P*=.009) and decrease their training costs and associated risks (Q14, *P*=.004). Additionally, all participants indicated a low level of agreement with the statement "During the simulation, I felt like a doctor taking care of this patient" (Q2, *P*=.99). No significant differences were observed in responses to the remaining questions when analyzed based on academic level.

Qualitative Analysis

Overview

The members of the research team conducted a 1-to-1 structured interview with the participating teachers around the interview outline of 7 open questions formulated in advance, as shown in Table 3.

XSL•FO
RenderX

Table 3. Themes of teacher groups' interview on the application and research of digital simulation teaching curriculum system.

Theme	Teacher
Perceived benefits of systematic teaching	"As one of the most complex and urgent diseases in orthopedics, severe pelvic trauma often fails to receive on-site teaching from teachers in the tense emergency treatment site. In addition, due to the long treatment period of this disease, it takes a long time to fully learn the diagnosis and treatment process of this disease. However, in real life, learners only spend limited time rotating with one department and cannot follow through the entire disease process and treatment course. By creating typical cases of severe pelvic trauma and constructing a virtual clinical diagnosis and treatment environment based on ESP, this system shortens the learning cycle of students, enables learners to experience different treatment settings, allows a large scale concurrent online participation breaking through the limitations of traditional teaching in time and space and improving the efficiency of teaching organizations."
The appropriateness of case application subjects	"It helps me conduct classified teaching according to the basic knowledge level of undergraduates and postgraduates. The basic medical and clinical medical knowledge involved in the disease set in the system is suitable for students at different undergraduate and postgraduate levels to learn. At the same time, the extensibility of the system enriches the flexibility and innovation of students' training and assessment."
The extendibility of course application	"The teaching design of this case is very suitable for the objective structured clinical examination scenario, which is closer to reality than traditional simulation training scenario. In addition, it introduces ESP, without the need for on-site re-placement of exam environments and standardized patient training."
The limitations of systematic research	"In the early stage of communication and interaction with the ESP speech inquiry, it was found that the ESP lacked a large sample of language training model library, so it could not recognize the semantics of the trainer. In addition, the ESP has not achieved the language style characteristics of different types of characters at this stage."
Recommendations for enhancing systematic research	"To enhance the virtual ESP simulation system, five improvements were suggested: enlarging the ESP case database, incorporating a feature for automatic and manual responses to technical queries, broadening the range of disease diagnosis and treatment simulations, expanding the ESP history collection database, and advancing the ESP's artificial intelligence for inquiry processing."

Theme 1: Perceived Benefits of Systematic Teaching

Participants noted that the digital simulation experimental teaching system for severe pelvic trauma significantly improved teaching efficiency and effectiveness, overcoming the traditional teaching constraints related to time and space.

Theme 2: The Appropriateness of Case Application Subjects

Most participants pointed out the variable difficulty of teaching cases within the system for different learning groups, highlighting the advanced design. The freedom for students to interact with the ESP in the system underscores its innovative development. The immersive simulation for diagnosis, treatment training, and assessment allowed students to thoroughly apply their theoretical knowledge and skills, presenting a notable challenge.

Theme 3: The Extendibility of Course Application

The majority of participants regarded case-based digital simulation systems as potent educational tools for both undergraduate and graduate training. A substantial number of participants viewed the system as suitable for integration into an objective structured clinical examination.

Theme 4: The Limitations of Systematic Research

Some limitations of the digital simulation software were reported by participants, particularly issues with the ESP not always accurately recognizing the semantics and tone of the inquiries.

Theme 5: Recommendations for Enhancing Systematic Research

A large case base, different training paths, and smarter ESP interaction can enhance the freshness, challenge, and realism of the ESP experience for the trainers.

Theoretical Knowledge Level of Severe Pelvic Trauma

A comparison of theoretical examination scores before and after participants used the digital simulation software for severe pelvic trauma showed significant improvements in their overall scores for diagnosing and treating the condition, making preliminary diagnoses, the sequence of disease treatment, emergency management of hemorrhagic shock, and performing external fixation of pelvic fractures (Table 4). The IQR box plots for the theoretical knowledge levels of severe pelvic trauma, both pretest and posttest, are provided in Multimedia Appendix 4.

XSL•FO
RenderX

Table 4. Mean scores at presimulation and postsimulation for the 5 uncoached assignments.

Severe pelvic trauma clinical skill training	Presimulation score, % (SEM)	Postsimulation score, % (SEM)	Mean difference[a] (95% CI)	*t* test (df=39)	Cohen *d*	*P* value[b]
Order of diagnosis and treatment	49.9 (2.0)	85.5 (1.4)	35.5 (32.7-38.3)	25.9	3.4	.001
Make a preliminary diagnosis	45.1 (1.4)	89.4 (1.0)	44.4 (42.2-46.5)	41.8	6.0	.001
Order of disease treatment	69.4 (1.8)	95.2 (0.5)	25.8 (22.2-29.5)	14.3	3.2	.001
Emergency treatment of hemorrhagic shock	39.7 (0.9)	85.8 (0.8)	46.1 (43.6-48.6)	37.6	8.5	.001
External fixation operation of pelvic fracture	32.7 (2.3)	91.1 (0.7)	58.4 (53.4-63.3)	24.1	5.3	.001

[a]The analysis included only paired data. The mean difference is the difference in mean presimulation score and mean postsimulation score.

[b]*P* value obtained from a paired 2-tailed *t* test.

Discussion

Principal Findings

This study yielded 3 primary findings. First, we developed a case-based digital simulation teaching system for severe pelvic trauma, incorporating principles of basic and clinical medicine. In contrast to traditional training methods, the VR system allows students to engage in repeated practice at their own pace, providing immediate and standardized feedback after each interaction. This feature overcomes challenges like high teacher-student ratios and insufficient feedback, which are common in traditional training environments. Furthermore, the use of a computer model to demonstrate physiological hemodynamic changes has been shown to aid in understanding the connection between clinical phenomena and underlying knowledge. Second, the software's acceptability, perceived ease of use, and perceived usefulness were highly regarded by users. Finally, the application of this digital simulation teaching system resulted in a significant improvement in all participating knowledge and skill scores. These findings contribute to the innovation in severe pelvic trauma skills training and may offer guidance for the development of enhanced training strategies and the revision of orthopedic surgery training standards.

Comparison to Prior Work

Although severe pelvic trauma is relatively rare in China, our hospital, being an orthopedic center of excellence, sees a higher incidence, treating over 100 patients annually and performing more than 20 external pelvic fixation procedures. The design of our case-based VR simulation curriculum for severe pelvic trauma draws from real cases, expert consensus, and literature reviews. Although previous research has demonstrated the efficacy of integrated learning [25], simulation training [26,27], traditional CBL [13], online learning [28], and digital patient simulator-assisted learning [29] in orthopedic clinical skills training, few studies have combined these methodologies. To our knowledge, this research is the inaugural study to amalgamate these proven effective training methods to enhance severe pelvic trauma clinical skill training, using a hybrid approach to assess the digital simulation efficacy.

Participants' acceptance of this clinical skills training was evident in several areas. Most participants felt the simulation training provided a compelling immersion experience, was accessible at any time and location, and had clear and understandable case scenario goals. Previous studies indicate that digital simulation software can significantly impact learning success [30]. Moreover, the degree of immersion is crucial in VR software, as identification with a digital character directly influences learning motivation and effectiveness [31]. Concerning the utility, all students concurred that the self-directed exploration learning method facilitated a deeper understanding of the knowledge and skills necessary for treating severe pelvic trauma and bolstered their confidence in handling similar conditions in real-life scenarios.

However, acceptance of clinical skill training was lower in certain aspects. A minority of trainees felt that the digital simulation technology's construction of the clinical environment and ESP allowed for experimentation within a safe psychological space. Studies suggest that digital simulators are effective for training doctor-patient communication skills [32]. Nevertheless, the discrepancy between virtual scenarios and real-life situations led to challenges in caring for actual patients. In our pilot study, most learners reported difficulty in direct communication with the ESP in the virtual environment, including eye and body language, and in discerning the nuances of the real language environment (such as tone and intonation); hence, they did not fully practice effective doctor-patient communication skills. Future educational efforts in all hospital departments should prioritize teaching doctor-patient communication skills to students.

Differences in the efficacy of case-based VR in pelvic trauma clinical skills training were observed between undergraduate and graduate students. Undergraduate respondents felt that after undergoing training with the digital simulation system, they solidified their basic medical knowledge and mastered the diagnostic and treatment processes for severe pelvic trauma; however, they expressed a lack of confidence in performing pelvic fracture external fixation. Graduate respondents believed that systematic training deepened their understanding of the diagnosis, treatment, and operational procedures for antihemorrhagic shock therapy and pelvic fracture external fixation. These variances are attributable to their respective stages of learning: undergraduates possess a stronger foundation in basic medical knowledge, while graduates have more opportunities to apply clinical knowledge in practice. Furthermore, undergraduate students outperformed graduate students in retaining disease-related knowledge due to their firmer grasp of basic medical principles. Conversely, their skills

XSL•FO
RenderX

were slightly inferior to those of graduate students, a disparity linked to the latter's greater internship experience and the number of surgical procedures conducted in the hospital. Thus, the training focus should be tailored to each need during severe pelvic trauma clinical skill training.

Quantitative outcome analyses revealed an overall improvement in pass rates at crucial assessment points posttraining with the digital simulation system, with external fixation of pelvic fractures displaying the most significant enhancement. Participants identified the realism and interactivity of the pelvic fracture model within the virtual environment as pivotal in elevating their learning experiences and assessment scores. Despite this, scores for external fixation of pelvic fractures were not the highest due to the inherent complexity and the necessity for interns to rotate through the orthopedic department to gain familiarity with patient care and the surgical technique [33]. Although the scores for diagnosing and treating the disease were the lowest presimulation, participants' scores in these 2 categories were the highest.

Personal interviews confirmed that teaching software facilitates large-scale online student learning in terms of ease and effectiveness. Tailoring cases to different learner groups introduced high levels of order, innovation, and challenge. The experiment also addressed challenges such as prolonged real-world teaching durations, access to actual patients, and teaching environment constraints. However, 1 instructor cautioned that this innovative skill training should complement, rather than replace, traditional teaching methods, a sentiment echoed by other educational research [21,34].

Implications

The digital simulation software for severe pelvic trauma provides undergraduates with an immersive learning experience that bridges theoretical knowledge and practical skills. For graduate students, it offers targeted preclinical training, preparing them for real-world trauma care. This approach enhances skill acquisition and promotes standardized training, potentially improving patient outcomes in severe trauma cases.

Limitations

Limitations include the inability to directly compare the digital simulation teaching system for severe pelvic trauma with traditional teaching models. Moreover, being self-controlled, participants' preexisting knowledge about the digital simulation system may have biased the outcomes. Although the participant count was sufficient for statistical analysis in this pilot study, the sample size remains limited.

Conclusions

Case-based VR simulation of skill training is an effective educational approach for medical students learning about severe pelvic trauma. It presents a potentially resource-efficient approach to delivering high-quality education for both educators and learners.

Acknowledgments

We express our gratitude to the 2021 first-class undergraduate curriculum project at Nanjing Medical University for its financial support. Additionally, we thank the undergraduate and graduate students of the First Clinical Medical College of Nanjing Medical University for their involvement, as well as the teaching development teams of the First Clinical Medical College and the Basic Medical College for their role in instructional design, software development, and production. The corresponding author may be contacted via email for further details or to request permission for educational tool use. Our appreciation also extends to Editage for their assistance with English language editing.

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

All the authors (PT, YX, KQ, ML, and JH) were involved in the study design and methods. PT conducted the investigation and formal analysis of the data and was responsible for writing the first draft of the manuscript. ML and JH supervised the study. All the authors contributed to and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Higher Resolution versions of Figures 3-7.
[DOCX File , 32564 KB - mededu_v11i1e59850_app1.docx]

Multimedia Appendix 2
Software development and construction process of pelvic fracture model.
[DOC File , 2820 KB - mededu_v11i1e59850_app2.doc]

XSL•FO
RenderX

Multimedia Appendix 3
An interview outline of 7 open-ended questions.
[DOC File , 32 KB - mededu_v11i1e59850_app3.doc]

Multimedia Appendix 4
Box plot with the median scores and IQR for pre- and posttest.
[DOC File , 148 KB - mededu_v11i1e59850_app4.doc]

References

1. Zong ZW, Chen SX, Qin H, Liang HP, Yang L, Zhao YF, Representing the Youth Committee on Traumatology branch of the Chinese Medical Association, PLA Professional Committee and Youth Committee on Disaster Medicine, Traumatology branch of the China Medical Rescue Association, Disaster Medicine branch of the Chongqing Association of Integrative Medicine. Chinese expert consensus on echelons treatment of pelvic fractures in modern war. Mil Med Res 2018 Jun 30;5(1):21 [FREE Full text] [doi: 10.1186/s40779-018-0168-3] [Medline: 29970166]
2. Heetveld MJ, Harris I, Schlaphoff G, Balogh Z, D'Amours SK, Sugrue M. Hemodynamically unstable pelvic fractures: recent care and new guidelines. World J Surg 2004 Sep;28(9):904-909. [doi: 10.1007/s00268-004-7357-9] [Medline: 15593465]
3. Coccolini F, Stahel PF, Montori G, Biffl W, Horer TM, Catena F, et al. Pelvic trauma: WSES classification and guidelines. World J Emerg Surg 2017;12:5 [FREE Full text] [doi: 10.1186/s13017-017-0117-6] [Medline: 28115984]
4. Marmor M, El Naga AN, Barker J, Matz J, Stergiadou S, Miclau T. Management of pelvic ring injury patients with hemodynamic instability. Front Surg 2020;7:588845 [FREE Full text] [doi: 10.3389/fsurg.2020.588845] [Medline: 33282907]
5. Bagaria D. The past, present, and future management of hemodynamic instability in patients with unstable pelvic ring injuries. Injury 2022 Mar;53(3):1294. [doi: 10.1016/j.injury.2021.11.061] [Medline: 34865818]
6. Cevallos N, Zukotynski B, Greig D, Silva M, Thompson RM. The utility of virtual reality in orthopedic surgical training. J Surg Educ 2022;79(6):1516-1525 [FREE Full text] [doi: 10.1016/j.jsurg.2022.06.007] [Medline: 35821110]
7. Keith K, Hansen D, Johannessen M. Perceived value of a skills laboratory with virtual reality simulator training in arthroscopy: a survey of orthopedic surgery residents. J Am Osteopath Assoc 2018 Oct 01;118(10):667-672 [FREE Full text] [doi: 10.7556/jaoa.2018.146] [Medline: 30264141]
8. Salem J, Fukuta J, Coombs A, Morgan J. Virtual patient journey: a novel learning resource. Clin Teach 2020 Jun;17(3):315-319. [doi: 10.1111/tct.13101] [Medline: 31680422]
9. Kononowicz AA, Woodham LA, Edelbring S, Stathakarou N, Davies D, Saxena N, et al. Virtual patient simulations in health professions education: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Jul 02;21(7):e14676 [FREE Full text] [doi: 10.2196/14676] [Medline: 31267981]
10. Yuan YB, Wang JJ, Lin MH, Gao XY. Building virtual simulation teaching platform based on electronic standardized patient. Sheng Li Xue Bao 2020 Dec 25;72(6):730-736 [FREE Full text] [Medline: 33349830]
11. Peñuela-Epalza M, De la Hoz K. Incorporation and evaluation of serial concept maps for vertical integration and clinical reasoning in case-based learning tutorials: perspectives of students beginning clinical medicine. Med Teach 2019 Apr;41(4):433-440. [doi: 10.1080/0142159X.2018.1487046] [Medline: 30091645]
12. Wei F, Sun Q, Qin Z, Zhuang H, Jiang G, Wu X. Application and practice of a step-by-step method combined with case-based learning in Chinese otoendoscopy education. BMC Med Educ 2021 Feb 04;21(1):89 [FREE Full text] [doi: 10.1186/s12909-021-02513-1] [Medline: 33541330]
13. Demetri L, Donnelley CA, MacKechnie MC, Toogood P. Comparison of case-based learning and traditional lectures in an orthopedic residency anatomy course. J Surg Educ 2021;78(2):679-685. [doi: 10.1016/j.jsurg.2020.08.026] [Medline: 32888846]
14. Gartmeier M, Pfurtscheller T, Hapfelmeier A, Grünewald M, Häusler J, Seidel T, et al. Teacher questions and student responses in case-based learning: outcomes of a video study in medical education. BMC Med Educ 2019 Dec 05;19(1):455 [FREE Full text] [doi: 10.1186/s12909-019-1895-1] [Medline: 31805913]
15. Zhao L, Dai X, Chen S. Effect of the case-based learning method combined with virtual reality simulation technology on midwifery laboratory courses: a quasi-experimental study. Int J Nurs Sci 2024 Jan;11(1):76-82 [FREE Full text] [doi: 10.1016/j.ijnss.2023.12.009] [Medline: 38352279]
16. Wang D, Kim L, Gronberg M, Stambaugh C, AAPM Medical Physics Leadership Academy (MPLA) Cases Subcommittee. A brief guide to writing a medical physics leadership case. J Appl Clin Med Phys 2021 Mar;22(3):285-286 [FREE Full text] [doi: 10.1002/acm2.13186] [Medline: 33739581]
17. ten Cate O, Custers EJFM, Durning SJ. Principles and Practice of Case-based Clinical Reasoning Education: A Method for Preclinical Students. Cham: Springer; 2018:95-108.
18. Encinas-Ullán CA, Martínez-Diez JM, Rodríguez-Merchán EC. The use of external fixation in the emergency department: applications, common errors, complications and their treatment. EFORT Open Rev 2020 Apr;5(4):204-214 [FREE Full text] [doi: 10.1302/2058-5241.5.190029] [Medline: 32377388]

XSL•FO
RenderX

19. Mi M, Kanakaris NK, Wu X, Giannoudis PV. Management and outcomes of open pelvic fractures: an update. Injury 2021 Oct;52(10):2738-2745. [doi: 10.1016/j.injury.2020.02.096] [Medline: 32139131]
20. Hertzog MA. Considerations in determining sample size for pilot studies. Res Nurs Health 2008 Apr;31(2):180-191. [doi: 10.1002/nur.20247] [Medline: 18183564]
21. Mahling M, Wunderlich R, Steiner D, Gorgati E, Festl-Wietek T, Herrmann-Werner A. Virtual reality for emergency medicine training in medical school: prospective, large-cohort implementation study. J Med Internet Res 2023 Mar 03;25:e43649 [FREE Full text] [doi: 10.2196/43649] [Medline: 36867440]
22. Balki E, Holland C, Hayes N. Use and acceptance of digital communication technology by older adults for social connectedness during the COVID-19 pandemic: mixed methods study. J Med Internet Res 2023 Aug 02;25:e41535 [FREE Full text] [doi: 10.2196/41535] [Medline: 37531187]
23. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
24. Erlingsson C, Brysiewicz P. A hands-on guide to doing content analysis. Afr J Emerg Med 2017 Sep;7(3):93-99 [FREE Full text] [doi: 10.1016/j.afjem.2017.08.001] [Medline: 30456117]
25. Boller E, Courtman N, Chiavaroli N, Beck C. Design and delivery of the clinical integrative puzzle as a collaborative learning tool. J Vet Med Educ 2021 Apr;48(2):150-157. [doi: 10.3138/jvme.2019-0036] [Medline: 33861187]
26. Marchand LS, Sciadini MF. Simulation training in fracture surgery. J Am Acad Orthop Surg 2020 Nov 01;28(21):e939-e947. [doi: 10.5435/JAAOS-D-20-00076] [Medline: 32796368]
27. Tillander B, Ledin T, Nordqvist P, Skarman E, Wahlström O. A virtual reality trauma simulator. Med Teach 2004 Mar;26(2):189-191. [doi: 10.1080/0142159042000192037] [Medline: 15203531]
28. Brennan JN, Hall AJ, Baird EJ. Surgeon 2023 Oct;21(5):e263-e270. [doi: 10.1016/j.surge.2023.02.005] [Medline: 36914519]
29. Huber M, Katzky U, Müller K, Blätzinger M, Goetz W, Grechenig P, et al. Evaluation of a new virtual reality concept teaching k-wire drilling with force feedback simulated haptic in orthopedic skills training. J Hand Surg Am 2022 Dec;47(12):1225.e1-1225.e7. [doi: 10.1016/j.jhsa.2021.09.008] [Medline: 34857404]
30. Howard T, Iyengar KP, Vaishya R, Ahluwalia R. High-fidelity virtual reality simulation training in enhancing competency assessment in orthopaedic training. Br J Hosp Med (Lond) 2023 Sep 02;84(9):1-8 [FREE Full text] [doi: 10.12968/hmed.2022.0360] [Medline: 37769263]
31. Thompson J, White S, Chapman S. Interactive clinical avatar use in pharmacist preregistration training: design and review. J Med Internet Res 2020 Nov 06;22(11):e17146 [FREE Full text] [doi: 10.2196/17146] [Medline: 33155983]
32. Guetterman TC, Sakakibara R, Baireddy S, Kron FW, Scerbo MW, Cleary JF, et al. Medical students' experiences and outcomes using a virtual human simulation to improve communication skills: mixed methods study. J Med Internet Res 2019 Nov 27;21(11):e15459 [FREE Full text] [doi: 10.2196/15459] [Medline: 31774400]
33. Ferede B, Ayenew A, Belay W. Pelvic fractures and associated injuries in patients admitted to and treated at emergency department of Tibebe Ghion Specialized Hospital, Bahir Dar University, Ethiopia. Orthop Res Rev 2021;13:73-80 [FREE Full text] [doi: 10.2147/ORR.S311441] [Medline: 34140815]
34. Kleinert R, Heiermann N, Plum PS, Wahba R, Chang DH, Maus M, et al. Web-based immersive virtual patient simulators: positive effect on clinical reasoning in medical education. J Med Internet Res 2015 Nov 17;17(11):e263 [FREE Full text] [doi: 10.2196/jmir.5035] [Medline: 26577020]

Abbreviations

CBL: case-based learning
ESP: electronic standardized patient
VR: virtual reality

Edited by SR Mogali; submitted 24.04.24; peer-reviewed by Q Yan, N Jiang; comments to author 28.10.24; revised version received 11.11.24; accepted 15.12.24; published 17.01.25.

Please cite as:
Teng P, Xu Y, Qian K, Lu M, Hu J
Case-Based Virtual Reality Simulation for Severe Pelvic Trauma Clinical Skill Training in Medical Students: Design and Pilot Study
JMIR Med Educ 2025;11:e59850
URL: https://mededu.jmir.org/2025/1/e59850
doi:10.2196/59850
PMID:

XSL•FO
RenderX

©Peng Teng, Youran Xu, Kaoliang Qian, Ming Lu, Jun Hu. Originally published in JMIR Medical Education (https://mededu.jmir.org), 17.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Extended Reality–Enhanced Mental Health Consultation Training: Quantitative Evaluation Study

Katherine Hiley[1,2], BSc; Zanib Bi-Mohammad[3,4], PhD; Luke Taylor[1], BSc; Rebecca Burgess-Dawson[5], MSc; Dominic Patterson[5], MSc, MBChB; Devon Puttick-Whiteman[5], BA; Christopher Gay[5], MSc; Janette Hiscoe[5]; Chris Munsch[5], MB, ChM, FRCS; Sally Richardson[5]; Mark Knowles-Lee[6]; Celia Beecham[6], BA; Neil Ralph[5], DClinPsych; Arunangsu Chatterjee[7], PhD; Ryan Mathew[1,8], FRCS, PhD; Faisal Mushtaq[1,2], PhD

[1]Centre for Immersive Technologies, HELIX, University of Leeds, Leeds, United Kingdom
[2]School of Psychology, Faculty of Medicine & Health, University of Leeds, Leeds, United Kingdom
[3]School of Science, Technology and Health, York St John University, York, United Kingdom
[4]School of Healthcare, Faculty of Medicine & Health, University of Leeds, Leeds, United Kingdom
[5]NHS England, England, United Kingdom
[6]Fracture Reality, Brighton, United Kingdom
[7]School of Medicine, Faculty of Medicine & Health, University of Leeds, Leeds, United Kingdom
[8]Department of Neurosurgery, Leeds Teaching Hospitals NHS Trust, Leeds, United Kingdom

Corresponding Author:
Faisal Mushtaq, PhD
School of Psychology
Faculty of Medicine & Health
University of Leeds
Woodhouse
Leeds, LS2 9JT
United Kingdom
Phone: 44 07525418924
Email: f.mushtaq@leeds.ac.uk

Abstract

Background: The use of extended reality (XR) technologies in health care can potentially address some of the significant resource and time constraints related to delivering training for health care professionals. While substantial progress in realizing this potential has been made across several domains, including surgery, anatomy, and rehabilitation, the implementation of XR in mental health training, where nuanced humanistic interactions are central, has lagged.

Objective: Given the growing societal and health care service need for trained mental health and care workers, coupled with the heterogeneity of exposure during training and the shortage of placement opportunities, we explored the feasibility and utility of a novel XR tool for mental health consultation training. Specifically, we set out to evaluate a training simulation created through collaboration among software developers, clinicians, and learning technologists, in which users interact with a virtual patient, "Stacey," through a virtual reality or augmented reality head-mounted display. The tool was designed to provide trainee health care professionals with an immersive experience of a consultation with a patient presenting with perinatal mental health symptoms. Users verbally interacted with the patient, and a human instructor selected responses from a repository of prerecorded voice-acted clips.

Methods: In a pilot experiment, we confirmed the face validity and usability of this platform for perinatal and primary care training with subject-matter experts. In our follow-up experiment, we delivered personalized 1-hour training sessions to 123 participants, comprising mental health nursing trainees, general practitioner doctors in training, and students in psychology and medicine. This phase involved a comprehensive evaluation focusing on usability, validity, and both cognitive and affective learning outcomes.

Results: We found significant enhancements in learning metrics across all participant groups. Notably, there was a marked increase in understanding ($P<.001$) and motivation ($P<.001$), coupled with decreased anxiety related to mental health consultations ($P<.001$). There were also significant improvements to considerations toward careers in perinatal mental health ($P<.001$).

XSL•FO
RenderX

Conclusions: Our findings show, for the first time, that XR can be used to provide an effective, standardized, and reproducible tool for trainees to develop their mental health consultation skills. We suggest that XR could provide a solution to overcoming the current resource challenges associated with equipping current and future health care professionals, which are likely to be exacerbated by workforce expansion plans.

(JMIR Med Educ 2025;11:e64619) doi:10.2196/64619

KEYWORDS

mental health; training; consultation; extended reality; virtual reality; augmented reality

Introduction

As the demand for mental health services in health care systems continues to rise, the need for skilled professionals capable of providing effective mental health consultation and support also increases [1,2]. In the face of changing workforce training requirements (coupled with significant health care workforce expansion plans), there is a growing recognition that the effective implementation of emerging technologies could help overcome some of the logistical and resource-related barriers involved in education and training.

Mental health nursing, in particular, faces distinct challenges that necessitate specialized training solutions. Mental health nurses encounter unique stressors, including high levels of emotional exhaustion, moral distress, and exposure to patient-initiated violence, all of which contribute to job dissatisfaction and high turnover rates [3], which in turn negatively impact workforce stability, patient outcomes, and overall health care service quality [4]. Additionally, mental health nurses often report insufficient opportunities for continuing professional development and limited support from leadership, further compounding retention challenges. Addressing these issues through targeted and innovative training approaches is essential for fostering resilience, enhancing job satisfaction, and improving workforce retention.

Beyond specialist mental health settings, primary care physicians or general practitioners (GPs) also play key roles in managing mental health conditions, with more than a third of general practice consultations involving mental health issues [5]. Effective communication and therapeutic relationships have been shown to significantly influence outcomes, emphasizing the need for better training in interpersonal and empathetic skills for managing mental health conditions in primary care. However, variability in the ability of GPs to detect and manage mental health issues highlights gaps in current training models [6]. As communication forms a central part of mental health treatment, poorly trained clinicians may inadvertently block disclosure of emotional distress, potentially delaying critical interventions [5]. Therefore, innovative training approaches are crucial not only for mental health nurses but also for GPs and other health care professionals involved in mental health consultations.

Traditional training for health professionals in managing mental health problems typically relies on a combination of in-person placements, which employ observation-based learning, and actor-based simulations. While in-person placements provide valuable real-world experience, they often present challenges, such as unpredictable exposure to a diverse range of patient demographics, risks to both students and vulnerable service users, and limited opportunities for structured feedback. Actor-based simulations, on the other hand, are difficult to scale and standardize due to variability in actors' interpretations of scripts and inconsistencies in their familiarity with specific case studies. These limitations make it challenging to provide health care professionals with the comprehensive training necessary to handle the complexities of mental health consultations. Effective and compassionate mental health consultations require more than procedural knowledge. They demand the ability to empathize, engage in therapeutic communication, and establish a strong patient-provider relationship. To address these needs, training must focus on promoting empathy and compassion while preparing health care professionals to navigate the diverse backgrounds and emotional experiences of patients. However, traditional training methods often struggle to meet these goals due to ethical concerns around exposing students to sensitive cases and the inherent difficulty in replicating the unpredictable dynamics of real-life mental health scenarios.

Advances in a suite of new immersive technologies that go under the banner of extended reality (XR) and include virtual reality (VR) and augmented reality (AR) could be particularly well-suited to address these challenges by providing interactive, standardized, repeatable learning experiences that bridge the gap between theory and practice. VR presents users with a computer-generated environment that immerses them in a fully digitally simulated environment, while AR overlays virtually generated elements onto the real world. The value of XR for health care training has already been demonstrated across various domains, such as surgery [7], physical rehabilitation [8], anatomy [9], and the training of practical skills in nurses [10]. However, the implementation of XR in the training of mental health professionals has lagged.

Given the importance and complexity of training for mental health consultations, coupled with the increasing workload pressure on GPs and mental health nurses to meet the population's mental health support needs [7], we set out to test whether XR technology could be used to create a training environment to support the development of mental health consultation skills. We reasoned that the ability to deliver standardized repeatable experiences of varied patient encounters (including more rare presentations) in a safe and controlled environment could provide a learning experience that nurtures confidence and competence in consultation skills that augment traditional training.

To assess the potential efficacy of XR in mental health consultation training, we focused on perinatal mental health

XSL•FO
RenderX

training, a subspecialty supporting women navigating mental health challenges during pregnancy or the initial postpartum year. This is an area of the mental health service with an urgent training need. The recent report of the Royal College of Psychiatrists [8] highlighted a critical need for comprehensive perinatal training programs across both specialized and general health care services. There is also a notable lack of confidence among perinatal mental health nurses in their capacity to deliver care to women with perinatal mental health challenges, with only a quarter feeling well-equipped to support these women [9]. Trainees also have relatively limited opportunities to train, with a shortage of placement opportunities. A recent review of perinatal mental health education across 32 UK medical schools [10] found that perinatal mental health was not considered a core curriculum topic. Instead, it was typically incorporated as a subtopic within broader topic areas, such as lectures on depression. Given the shortage of staff and limited placements in perinatal mental health, a new training tool that could support the development of the next generation of health care staff could have an immediate impact.

Here, we report on the validation and evaluation of a novel XR training tool developed through a collaboration between software developers and health care staff, including nurses specializing in perinatal mental health and GPs. The simulation presents an interactive virtual patient ("Stacey") with severe perinatal mental health problems. Stacey is a mother of 2 children, with her youngest child only 4 weeks old, and has a record of mild postnatal depression following her first birth. Low mood, suicidal ideation, and episodes of psychosis add complex layers to her clinical presentation. Users interact with Stacey verbally, and her responses are selected by a human instructor from a range of prerecorded voice-acted clips in an audio repository. We explore the utility of this tool for supporting social and emotional interactions with the simulation, investigate the ease of use for trainers and trainees, and evaluate the impact on cognitive and affective learning.

Methods

Overall Approach

We undertook a 2-stage evaluation process that included a pilot study exploring feasibility and a subsequent evaluation of the perinatal mental health XR training experience in terms of learning outcomes and perceptions. In this section, we introduce the simulation platform and training experience and subsequently detail the methods and procedures common to and distinct for each phase of the experiment. It should be noted that the authors involved in developing the content played no role in the evaluation. The analysis was carried out independently by the authors KH, LT, and FM.

This study was not designed as a head-to-head comparison with traditional training approaches. Some participants, particularly those in mental health nursing, had previously received standard training methods (eg, classroom-based teaching, in-person clinical placements, or actor-based role plays), but these forms of training were not systematically assessed here. Instead, the primary aim was to evaluate the feasibility and potential impact of XR as a supplementary training tool. We have thus included the details of traditional training experiences for context but did not incorporate a direct comparative arm in this work.

XR Simulation

The simulation was built on a platform ("JoinXR") created by the software developer Fracture Reality. The JoinXR platform was designed to enable multi-user simulation training environments over a range of head-mounted VR or AR displays. In the evaluation, we used the Meta Quest 2 headset (Meta Platforms, Inc) for the VR version of the platform and Microsoft HoloLens 2 (Microsoft Corp) for the AR version.

The human-computer interface within the JoinXR platform was central to facilitating immersive lifelike interactions with the virtual patient. The interface allowed learners to engage through natural voice-based dialogue, processed in real time using instructor-guided responses. The integration of audio feedback, gesture recognition (HoloLens 2), and hand controllers (Meta Quest 2) enabled users to navigate the virtual space intuitively. Learners interacted directly with Stacey, the patient avatar, whose responses, including eye contact, subtle emotional cues, and body movements, were programmed to mimic real-world patient behavior, creating a realistic and contextually relevant learning experience. Users engaged with the simulation through natural voice-based dialogue and through VR controllers (Meta Quest 2) or hand-tracking gestures (HoloLens 2). Nonverbal communication, such as the avatar's facial expression, body language, and spatial audio, enhanced the simulation's realism.

The JoinXR platform was designed to be a conversational engine enabling "human to digital avatar" interactions in a multi-user, real-time environment. In this way, it could facilitate remote participation by learners, instructors, and observers, supporting the practice and refinement of nonroutine clinical skills. The learner-instructor dynamic was a crucial component of the simulation, incorporating both real-time guidance and postsimulation feedback. Instructors played an active role during the interaction by interpreting learner inputs and controlling Stacey's responses using a soundboard system (Figure 1B). This allowed for dynamic adaptations, where learners could engage organically with the avatar and explore different conversational pathways. After the simulation, instructors conducted debrief sessions using performance analytics that tracked response accuracy, emotional sensitivity, and decision-making, providing learners with targeted feedback to refine their clinical competencies.

XSL•FO
RenderX

Figure 1. Development of the learning platform. (A) Wireframe of the patient avatar, Stacey; (B) Soundboard for instructors to control Stacey's responses; (C) Setting for the consultation, showing the learner (blue) and the patient avatar.

During the simulation, learners interacted with Stacey by asking questions that were either processed by conversational artificial intelligence (AI) or directly controlled by the instructor for tailored responses. Stacey's reactions were designed to simulate real-world patient behaviors, including nuanced emotional expressions and gestures. Figure 2B illustrates an example consultation scenario in which the learner uses voice input to ask about Stacey's symptoms, prompting verbal and nonverbal responses (eg, maintaining eye contact and gesturing to emphasize a point).

Learners using VR devices (Meta Quest 2) could navigate the virtual consultation room using handheld controllers to manipulate objects, such as a clipboard or a stethoscope, or to adjust their position relative to Stacey. In contrast, AR users (HoloLens 2) experienced a blended environment where Stacey's avatar appeared within a real-world room.

The clinical simulations were developed through collaboration between Fracture Reality and a panel of subject-matter experts from the National Health Service (NHS), including mental health clinicians, GPs, and psychologists specializing in perinatal mental health. These experts supported the design of all aspects of the simulations, from character development and storyline construction to ensuring the accurate portrayal of medical conditions. Prior to the present evaluation, the development process included an iterative feedback process involving clinicians with primary care and perinatal mental health experience, software developers, and intended end users.

The specific focus of our evaluation is the first clinical simulation scenario developed using this new platform (Figure 1). The simulation is centered around a female patient avatar ("Stacey"). The aforementioned clinical experts contributed to the development of her patient history and personal attributes. Digital reference photos were then gathered to build a montage of the patient. A base model was built by taking a full body scan of a human model and was modified using a combination of 3D modeling software. The 3D models were created using a combination of Maya (Autodesk) and Blender (Blender Foundation). Clothing was designed, and then, the digital model was dressed. Bespoke custom lighting and skin rendering pipelines were developed to deliver realistic digital human features that could be rendered on headsets using low-powered graphics processing units.

Multiple script iterations were recorded, and the dialogue was reviewed and refined for clinical authenticity by the Fracture Reality team in consultation with the aforementioned subject-matter experts. Auditions were held to select actors. Studio sessions and spatial audio engineering rebalanced vocals to realistically imitate the patient avatar. Animations combined motion capture, hand animation, and lip-syncing for seamless responses. A custom Unity system facilitated quick and accurate lip-syncing to facial expressions and body poses (Figure 2). Reference photos from NHS facilities were used, and lighting was tailored for realistic environments, focusing on meaningful prop placement.

Two scenarios were designed, each tailored to address the needs of 2 primary but distinct target groups: mental health nursing students and primary care trainees (postgraduate doctor in GP training). While both scenarios feature a patient named Stacey presenting with a similar mental health condition, contextual variations were introduced to align more closely with the necessary professional capabilities of the respective trainee groups.

In the mental health nursing scenario, Stacey Morris is introduced as an emergency referral from her GP for a comprehensive assessment. Stacey, a 32-year-old mother of 2 children, with a 4-week-old newborn, has a history of postpartum depression following the birth of her first child. The primary objective for the student in this scenario is to conduct an initial mental health examination of Stacey.

In the primary care scenario, following a telephone conversation with her husband, Josh, who expressed concerns about her behavior, the postgraduate doctor in GP training agrees to meet Stacey in her home. Stacey in this scenario has a similar profile as in the mental health nursing scenario. She is a 32-year-old mother of 2 children, with her youngest child being 4 weeks old. In this context, the role of the postgraduate doctor in GP training centers on conducting a comprehensive mental health assessment with Stacey.

For each context, specific learning outcomes were defined by subject-matter experts. For the mental health nursing scenario, learners were expected to (1) understand and reflect on the lived experience of assessing the mental health of a patient with perinatal mental health problems; (2) identify signs and symptoms of perinatal mental ill health in acute assessment presentation; (3) apply the skills, knowledge, and abilities relevant to one's own profession in the assessment of mental health; and (4) have an appropriate reflected and evaluated performance of the task in a supported reflection. For the primary care setting, learners were expected to be able to (1) take history from a patient presenting with an acute psychotic illness; (2) ascertain and evaluate information relating to safeguarding; and (3) assess suicide and homicide risk.

XSL•FO
RenderX

Figure 2. Development of the verbal storyboard and voice integration for patient avatar interactions. (A) Demonstration of the process of integrating voice actors' performances into the patient avatar through a custom Unity-based lip-syncing system. Multiple iterations of dialogue scripts were recorded, and voice actors were selected via auditions, with audio engineering applied to simulate realistic patient speech patterns. Motion capture, hand animation, and spatial audio balancing enhanced the avatar's authenticity. (B) Demonstration of the verbal storyboard for the virtual patient, displaying categorized responses covering key clinical themes such as psychosis, self-harm, and family concerns. The storyboard guided the avatar's realistic conversational flow, informed by clinical expertise and reference material from National Health Service facilities to create immersive training scenarios.

General Methods

Following study advertisement, interested participants were screened for physical conditions that would exclude them from participation, including physical and auditory impairments and epilepsy. Included participants met with the instructor for a one-to-one session in a quiet room located on the university campus or at a local NHS hospital. Multimedia Appendix 1 outlines the study procedure. At the beginning of the session, participants had the opportunity to read the information sheet

XSL•FO
RenderX

and ask questions related to the study. Participants provided their consent to the study once they had been informed of their right to withdraw.

After consenting, participants were asked to complete a baseline questionnaire, capturing demographic and attitudinal data. Participants were then randomly allocated to 1 of 2 immersive environments: VR (Meta Quest 2) or mixed reality (HoloLens 2). Participants were subsequently exposed to either "primary care" Stacey (targeted at medical students and postgraduate doctors in GP training) or "mental health" Stacey (for mental health nursing or psychology students), contingent upon their current training program. These simulations share identical features, with the sole distinction lying in the introductory context of the consultation process. Both simulations were configured to align with the familiar protocols of health care trainees, specifically in terms of patient reception. Importantly, the responses of Stacey and the trajectory of the consultation remained consistent across the 2 scenarios.

Trainees verbally interacted with Stacey, who was in turn controlled by an instructor through the navigation of a soundboard, which triggered prerecorded audio clips from Stacey (see Figure 1B). The conversation journey would typically begin with general introductions; discussion of Stacey's relationships with her daughter, son, and husband; discussion of her birthday; discussion of experiences of psychosis and self-harm; and finally, formulation of a care plan. If the instructor felt the student was unable to lead the conversation or the student expressed having difficulty conversing with the avatar, prompts could be provided within the simulation. Multimedia Appendix 2 shows the prompts available for the early, mid, and late stages of the conversation that could be made visible to the students by the instructor.

Following the experience, the instructor carried out a postexperience debrief session with the trainee, including a critical discussion of the experience and the participant's performance. Following this, participants completed a postexperience survey measuring attitudinal domains and career considerations alongside measures of usability, presence, discomfort, and preference. The total session lasted approximately 1 hour (Multimedia Appendix 1).

Pilot Study

Participants

In the pilot study, we recruited 9 subject-matter experts from primary care and mental health disciplines. This included a consultant perinatal psychiatrist, a GP, 4 mental health nurses, a specialist perinatal mental health nurse, a psychiatry trainee (ST4), and a mental health nursing lecturer. All had more than 5 years of experience in their respective fields, with 8 having 10 or more years of experience. The purpose of this study was to formally test face and content validity and usability, and to support the latter, we included 5 undergraduate university students (mean age 22.4 years, SD 0.8 years).

Participants followed the procedure outlined in Multimedia Appendix 1. We evaluated face and content validity, usability, and utility as reported by a group of nonnursing or medical students and subject matter experts in the postexperience questionnaire. Face validity was assessed using a scale applied previously to expert evaluations of VR health care training [11]. This original 13-item scale was adapted to this study, and 11 items analyzed the ease of use, effectiveness, and immersion of the XR simulation on a 4-point Likert scale (strongly agree to strongly disagree).

In addition, the Lawshe method [12] also known as the content validity ratio (CVR) method was used. This is a method used to assess the content validity of a measurement instrument or a test, especially in the context of psychological, educational, or health care research, using expert opinion. Here, experts rate each item on a 3-point scale: (1) Essential: if the item is crucial and necessary for measuring the construct; (2) Useful but not essential: if the item is relevant but not critical for measuring the construct; and (3) Not necessary: if the item is irrelevant or not needed for measuring the construct. The CVR is calculated using the equation:

where *Ne* represents the count of experts who have deemed the item as "essential," and *N* denotes the total number of experts who have participated in the rating process. The CVR is a numerical value that quantifies the consensus among experts regarding the essential nature of the items under consideration. The critical value is a benchmark used to assess the appropriateness of items included in a content validity assessment. If the number of experts who agree on the relevance of an item meets or exceeds the critical value, the item is deemed valid; otherwise, it may be considered for revision or removal from the assessment. According to the values calculated previously [13] with a panel of 8 subject matter experts, this study's critical value was 0.75. Thus, constructs must surpass a CVR of 0.75 to be deemed essential to the procedure.

We also captured usability through the 10-item System Usability Scale (SUS) [14] as it has widely been used to evaluate XR as a tool for health care training [15-17]. Scores of more than 80 indicate excellence, between 70 and 80 are considered good, and less than 50 are not acceptable [18].

We assessed user discomfort through the Virtual Reality Sickness Questionnaire (VRSQ) [19]. As a more context appropriate adaptation of the validated Simulator Sickness Questionnaire (SSQ) [20], the VRSQ was designed to minimize burden on participants. The VRSQ sums the scores of oculomotor and disorientation discomfort items to generate an overall total. While there are no widely agreed bounds of acceptability for the VRSQ, we set out to compare the scores of the Meta Quest 2 and HoloLens 2 to assess relative differences in physical discomfort between the 2 devices.

Experiment

Following the demonstration of the feasibility of the use of XR in consultation training, we undertook a larger-scale evaluation. Here, we continued to collect measures of usability and supplemented them with surveys exploring cognitive and affective learning, training preference, presence, and career considerations.

XSL•FO
RenderX

Participants

The experiment involved 123 participants (mean age 24.3 years, SD 7.86 years; 97 female participants, 22 male participants, 3 nonbinary or third gender participants, and 1 who did not disclose gender). No participants had known health conditions, such as epilepsy, or visual, auditory, or cognitive disorders that would prevent participation in XR-based activities. They were drawn from a range of health care disciplines, including postgraduate doctors in GP training (n=18; mean age 38.2 years, SD 6.38 years), mental health nursing students (n=30; mean age 25.9 years, SD 7.6 years) from the Universities of Leeds and Huddersfield, and undergraduate medical students (n=28; mean age 19.8 years, SD 3.12 years) and psychology students (n=47; mean age 19.8 years, SD 3.2 years) recruited from the University of Leeds.

Participants were approached via their institutions, through the distribution of emails including information sheets. Participants were offered a monetary voucher incentive where appropriate (ie, for registered students), while clinical staff were asked to undertake the study voluntarily with no remuneration. Participants were randomly allocated to the AR (n=63, 51.2%) and VR training groups (n=60, 48.8%).

Measures

In phase 2, the VRSQ and usability continued to be evaluated as described in the pilot study. The experiment extended the evaluation to capture attitudes, cognitive and affective learning, career aspirations, presence, and conversation fluency. These self-reported measures were implemented to provide insights into the user's social and emotional interactions with the simulation, as well as any reported enhancements in knowledge, understanding, motivation, learning satisfaction, and learning confidence, further assessing the effectiveness of XR mental health consultations.

Attitudes

Cognitive Learning

Success and confidence in practical situations are often predicted by possessing knowledge, familiarity, and understanding of the themes and techniques embedded in a course curriculum [21]. Conversely, a deficiency in such familiarity may hinder the ability to apply theoretical knowledge in practice [22].

To capture this, a 14-item Perinatal Mental Health Familiarity and Awareness Scale (PMHAFS) (Multimedia Appendix 3) was developed by the study team with subject-matter experts. Participants were asked to evaluate their knowledge with, awareness of, and understanding of the perinatal mental health assessment conditions and care on a 5-point Likert scale (strongly disagree to strongly agree).

Affective Learning

Intrinsic and extrinsic motivation for learning was assessed through a 6-item scale [23], developed based on the Motivated Strategies for Learning Questionnaire Manual [24]. Evaluating intrinsic and extrinsic constructs provides a holistic examination of the influences of learner engagement from within the learner and from the learning environment [25]. Higher scores on each item suggest a greater motivation for learning.

To assess self-confidence and learning satisfaction, a 12-item variant [26] of the original Student Satisfaction and Self-Confidence in Learning Scale was used [27]. This instrument has been shown to be highly reliable, with a Cronbach α of .92 for the presence of features and .96 for their importance. Each item on the Likert scale was coded from 1 to 5 (strongly disagree to strongly agree), with 5 items reverse coded to prevent acquiescent responding. Higher scores on the scale indicate greater satisfaction and self-confidence with learning [28].

Career Attitudes

To assess students' considerations of health care specialization, we assessed 9 items across 3 affective domains of motivation, preparedness, and sense of support toward perinatal mental health specialization. Higher scores on each 5-point Likert scale indicate a greater desire to consider perinatal mental health upon graduation.

Presence

The construct of presence is regularly evaluated in studies involving virtual environments. Defined as the subjective experience of being in one place or environment, even when physically in another [29], there has been an active debate on its contribution to learning [30-32]. High levels of presence are speculated to be associated with deeper cognitive engagement, a cornerstone for effective learning [29], increasing intrinsic motivation and creating an environment where learners are more likely to integrate and retain new information [32]. A high degree of presence may help to minimize the impact of real-world distractions, allowing learners to fully immerse themselves in the task at hand [33]. Presence has also been proposed to be instrumental for the transfer of skills from the virtual to the real world [31]. We sought to measure presence through the previously validated iGroup Presence Questionnaire (IPQ) [34,35], a 14-item scale capturing spatial presence, realness, and involvement.

Ethical Considerations

Approval for the study was granted by the School of Psychology Ethics Committee (approval number: PSYC-615; date of approval: November 13, 2022). Consent was obtained from participants at the beginning of the session.

Statistical Analysis

ANOVA tests were performed to examine the effect of the XR training tool on the ratings of improvement in cognitive learning of conditions, assessment, and care. This same technique was applied to attitude changes in career motivation, support and preparedness, learning confidence, and learning satisfaction. Where appropriate, a between-subjects variable was introduced in the ANOVA when comparing population groups: GP postgraduate doctor in training, mental health nursing student, psychology student, or medical student.

For presence, specific data items related to presence were filtered and selected to include measures, such as "general," "spatial," "involvement," and "realism," as defined in the IPQ. The presence scores were reported across different devices and groups, examining how users experienced each of these presence

XSL•FO
RenderX

measures. An ANOVA assessed differences in presence scores between devices and measures (device [AR vs VR] × iGroup construct [general vs spatial vs involvement vs realism]). Post-hoc tests were applied to decompose interaction effects for VR and AR where appropriate.

For each family of tests (per construct), *P* values were corrected for multiple comparisons using the Bonferroni method. Corrected *P* values below an α threshold of .05 were considered statistically significant. All data analyses were performed in R (version 4.2.2) using RStudio (version 2022.12.0.353; Posit).

Results

Pilot Study

All experts, across both VR and AR systems (n=9, 100%), felt actively involved and in charge of the situation. The simulation software responded adequately and did not lag according to 8 of the experts, while all 9 experts reported that it was easy to learn how to interact with the software. Notably, all were interested in the progress of events throughout the simulation, suggesting high engagement. Additionally, all stated that it was easy to move around in the virtual environment, and the same amount of people reported that the controller buttons responded adequately.

Using the Lawshe method, we calculated the CVR for each step of the simulation process. These steps were: briefing instructions, medical notes, in-simulation prompts, instructor prompts, and postsimulation debrief. Briefing instructions provided the user with the necessary context for the forthcoming consultation. Medical notes, collaboratively developed with subject-matter experts, provided a comprehensive medical history for the virtual character, Stacey, to enhance the contextual richness of the consultation. In-simulation text prompts, illustrated in Multimedia Appendix 2, could be administered within the XR environment by the instructor, without verbal disruption to the ongoing consultation. In contrast, instructor prompts denoted verbal interventions made by the instructor at any time during the simulation. The postsimulation debrief is an opportunity for the user to reflect and for both the user and instructor to critically evaluate the consultation. The critical value in our study for the content validity of a construct and the component part of the simulation was 0.75. The obtained CVR scores for simulation outcomes, briefing instructions, and postexperience debrief were all 1, indicating that these processes were all rated as essential by all experts.

Some parts of the procedure, including previewing medical notes and using prompts during sessions, were considered optional by design. Our evaluation revealed that all content experts rated it as either essential or useful. In the case of in-simulation and instructor prompts, the majority found them essential or useful, but some considered them “not necessary,” as indicated by a score of 0.75.

The SUS scores were 78.75 for VR and 73.75 for AR, indicating good usability for both systems. The VRSQ scores were 0 for VR and 4.17 for AR, suggesting a negligible amount of discomfort for participants.

Pilot Study Summary

Participants provided positive feedback, reporting high usability levels for both VR and AR systems and minimal discomfort. Subject-matter experts rated the XR simulation highly in terms of engagement, involvement, and simulation quality, particularly within the context of perinatal training. They found the content and procedures valid, aligning with their expectations for an effective training session. These results suggest that the XR simulation has the potential to serve as a learner-centered training tool and provide the basis for conducting a larger-scale evaluation with health care trainees.

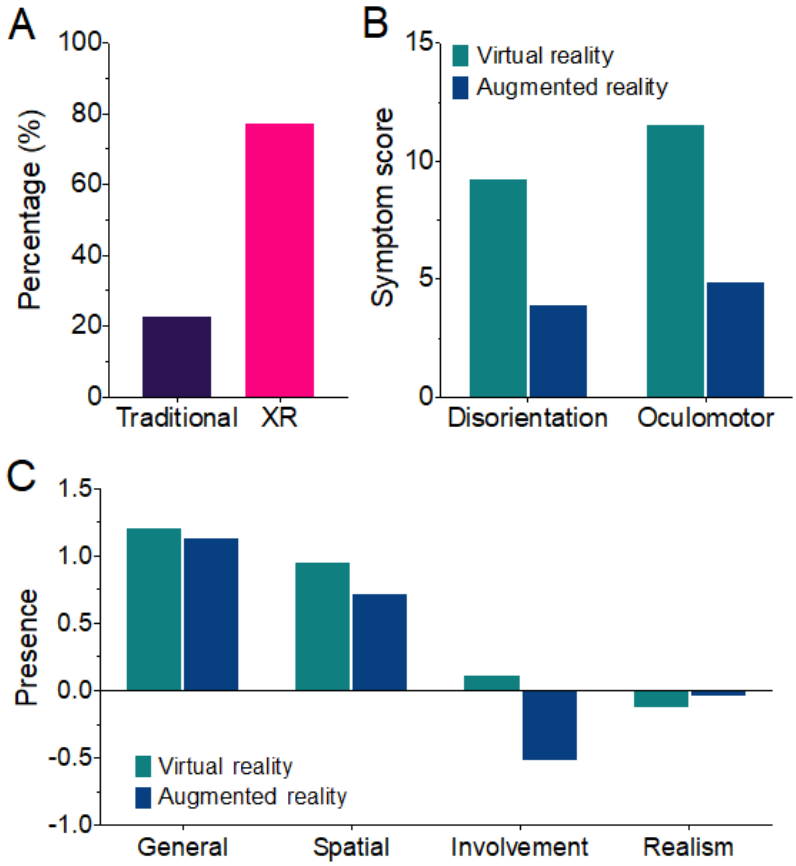
Experiment

Preference

Participants were asked whether they preferred the XR simulation or traditional approach to training that they had been exposed to. Overall, 77.2% (95/123) of participants preferred XR over traditional training methods (28/123, 22.8%) (Figure 3A).

XSL•FO
RenderX

Figure 3. Usability and preference. (A) User preference toward traditional learning for perinatal mental health or the integration of extended reality (XR) to augment perinatal mental health learning. (B) Symptom scores for the disorientation and oculomotor domains of the Virtual Reality Sickness Questionnaire for virtual reality (VR) and augmented reality (AR). (C) Self-reported experience of presence, illustrating that participants felt less involved in AR relative to VR. Error bars represent ±1 SEM.



Feasibility

The overall SUS score was 81.6 (SD 11.1), with no difference (t_{73}=0.75; P=.45) between the scores for AR (mean 82.3, SD 10.9) and VR (mean 80.3, SD 12.8), which translates to an excellent usability rating for both systems.

Simulator Sickness

In an analysis designed to understand the impact of different devices on simulator sickness, a 2-way ANOVA revealed a significant interaction between device and symptom ($F_{3,312}$=6.41; P<.001). There were greater sickness scores in the disorientation domain in VR (mean 9.22, SD 1.06) than in AR (mean 3.92, SD 0.81) (t_{208}=3.47; P<.001) and greater scores in the oculomotor domain in VR (mean 11.53, SD 1.33) than in AR (mean 4.89, SD 1.01) (t_{208}=4.34; P<.001). The analysis suggests that VR is more likely to cause symptoms of disorientation and oculomotor discomfort than AR.

Presence

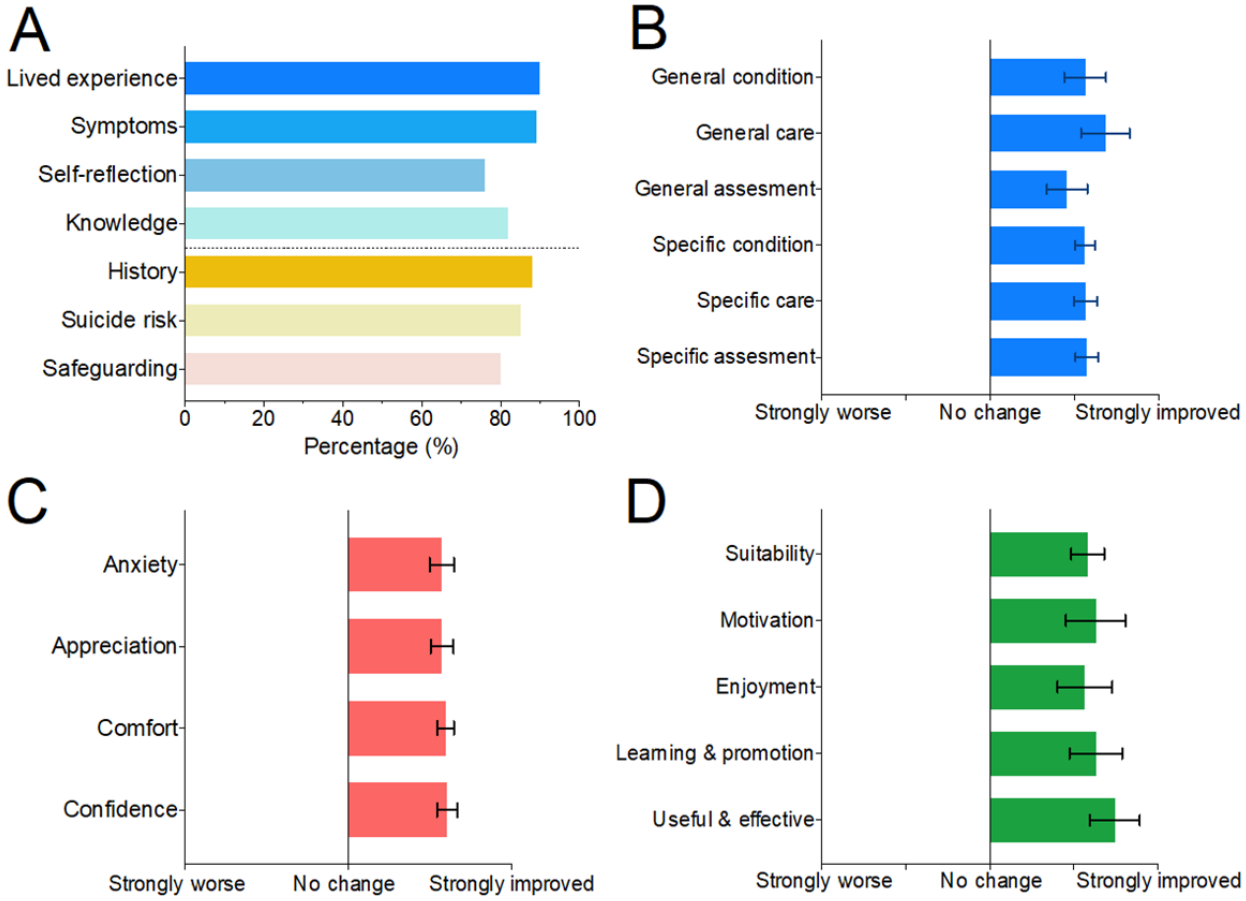
IPQ scores were compared between VR and AR. There was a statistically significant interaction between presence measure and device ($F_{3,327}$=5.78; P=.02; η^2_G=0.025). Post-hoc analysis revealed a statistically improved sense of involvement for those using VR relative to AR (t_{484}=3.18; P=.002). There were no significant differences between the systems in general (t_{484}=0.13; P=.90) and in the spatial (t_{484}=–1.11; P=.27) and experienced realism (t_{484}=1.17; P=.24) domains of presence.

Learning Outcomes

For the 2 simulations evaluated in this study, specific learning outcomes were defined by subject matter experts from perinatal mental health and primary care. We report these separately for each group. Figure 4A shows the percentage of sessions in which the learning outcome was achieved across all groups, as reported by the instructor.

XSL•FO
RenderX

Figure 4. Learning outcomes, cognitive and affective changes, and learning satisfaction. (A) Percentage of achievement of learning objectives for perinatal mental health and primary care simulations within sessions across all participants. (B) Improvements in understanding across perinatal conditions, assessment, and care domains following the simulation for general practitioner (GP) trainees, and improvements in understanding across perinatal conditions, assessment, and care domains following the simulation for mental health nursing students. (C) Improvements in the affective domains of confidence, comfort, appreciation for the challenges in providing support to perinatal cases, and reduction in anxiety among perinatal cases for GP trainees. (D) Improvements in the utility and feasibility domains for mental health consultation training following XR simulations compared with the current training approach. Error bars represent the SEM for domain change responses.



In the primary care simulation, instructors rated that they were able to achieve learning objective 1 (able to take history from a patient presenting with an acute psychotic illness) in 100% of sessions, learning objective 2 (able to ascertain and evaluate information relating to safeguarding) in 80% of sessions, and learning objective 3 (able to assess suicide and homicide risk) in 80% of sessions.

In the perinatal mental health simulation, instructors rated that they were able to achieve learning objective 1 (understand and reflect on the lived experience of assessing the mental health of a patient with perinatal mental health problems) in 100% of sessions, learning objective 2 (identify the signs and symptoms of perinatal mental ill health in acute assessment presentation) in 90% of sessions, learning objective 3 (apply the skills, knowledge, and abilities relevant to one's own profession in the assessment of mental health) in 89% of sessions, and learning objective 4 (have appropriate reflected and evaluated performance of the task in a supported reflection) in 80% of sessions.

Changes in Cognitive and Affective Attitudes

Primary Care

At baseline, 22% of participants stated that they had "no experience" with perinatal mental health cases, 61% expressed "little experience," and only 17% expressed "some experience." Understanding of complex mental health (general) and perinatal mental health (specific) was measured at baseline, revealing that 59% of GP trainees expressed an understanding of complex mental health at a general level and 58% expressed an understanding of perinatal mental health specifically.

Regarding affective constructs, 44% of trainees expressed anxiety around complex mental health cases and 50% expressed anxiety around perinatal mental health cases. Following the simulation, participants reported a statistically significant improvement in cognitive attitudes (mean 0.91, SD 0.86; t_{17}=4.47; P=.003; d=1.05).

Participants further reported a statistically significant improvement in affective attitudes following the simulation (mean 0.92, SD 0.74; t_{17}=5.27; P<.001; d=1.17). Across the affective domain, participants reported an improvement in confidence (mean 0.83, SD 1.04; t_{17}=3.39; P=.004; d=0.79),

XSL•FO
RenderX

comfort (mean 0.89, SD 0.76; t_{17}=4.97; P<.001), appreciation for the challenges of providing perinatal mental health support (mean 0.94, SD 1.00; t_{17}=4.01; P=.001; d=0.95), and reduced anxiety toward perinatal mental health cases (mean 1.00, SD 1.09; t_{17}=3.91; P=.001; d=0.92).

Medical Students

Following the simulation, medical students reported an improvement in cognitive attitudes (mean 1.38, SD 0.40; t_{27}=18.14; P<.001; d=3.42). This group also reported a statistically significant improvement in affective attitudes (mean 1.35, SD 0.46; t_{27}=10.01; P<.001; d=1.89). Across the affective domain, students reported an improvement in confidence (mean 1.64, SD 0.58; t_{27}=13.45; P<.001; d=2.54), comfort (mean 1.39, SD 0.57; t_{27}=13.00; P<.001; d=2.46), appreciation (mean 1.29, SD 0.90; t_{27}=7.59; P<.001; d=1.43), and reduced anxiety toward perinatal mental health cases (mean 1.25, SD 0.97; t_{27}=6.84; P<.001; d=1.29).

Mental Health Students and Psychology Students

Mental health and psychology students reported a significantly improved understanding of perinatal mental health conditions (mean 1.01, SD 0.62; t_{76}=14.26; P<.001; d=1.63), assessment (mean 1.21, SD 0.74; t_{76}=14.60; P<.001; d=1.62), and care (mean 1.09, SD 0.65; t_{26}=14.82; P<.001; d=1.69) following the simulation.

Within the mental health student group, improvements were seen following the simulation across the domains of perinatal mental health conditions (mean 0.86, SD 0.65; t_{29}=7.26; P<.001; d=1.32), assessment (mean 0.91, SD 0.67; t_{29}=7.41; P<.001; d=1.35), and care (mean 0.76, SD 0.60; t_{29}=6.88; P<.001; d=1.26).

Improvements were also seen in the psychology group across the domains of conditions (mean 1.11, SD 0.59; t_{46}=12.85; P<.001; d=1.87), assessment (mean 1.39, SD 0.73; t_{46}=13.10; P<.001; d=1.91), and care (mean 1.31, SD 0.59; t_{26}=15.32; P<.001; d=2.23).

Across all mental health and psychology students, we found a significant increase in learning confidence (mean 1.14, SD 0.49; t_{76}=20.32; P<.001; d =2.32). Students further reported a significant increase in learning satisfaction (mean 1.33, SD 0.69; t_{76}=16.51; P<.001; d=1.88). There was a similar finding within groups, as mental health students reported a significant increase in learning confidence following the simulation (mean 1.13, SD 0.57; t_{29}=10.83; P<.001; d=1.98). Psychology students also reported a significant increase in learning confidence following the simulation (mean 1.13, SD 0.43; t_{46}=15.32; P<.001; d=2.61).

For learning satisfaction, mental health students reported a significant increase following the simulation (mean 1.25, SD 0.82; t_{29}=8.33; P<.001; d=1.52), and psychology students also reported a significant increase following the simulation (mean 1.37, SD 0.62; t_{46}=15.12; P<.001; d=2.20).

Career Considerations

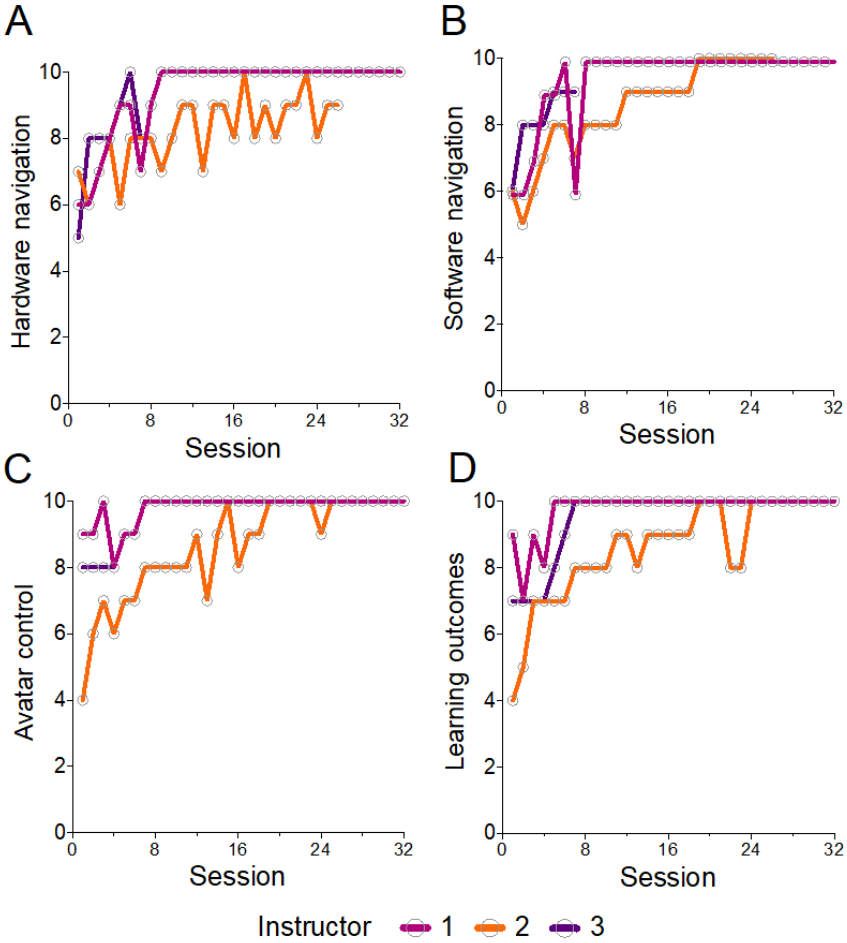
At baseline, 49% of mental health nursing students stated that they were motivated to pursue a career in perinatal mental health, while 30% agreed that they felt prepared to pursue a career in perinatal mental health and 24% felt supported to pursue a career in perinatal mental health. Only 25% of psychology students were considering a career in perinatal mental health. Following the simulation, mental health nursing students felt significantly more motivated (mean 0.73, SD 0.65; t_{29}=6.15; P<.001; d=1.12), prepared (mean 1.10, SD 0.52; t_{29}=11.61; P<.001; d=2.12), and supported (mean 0.74, SD 0.74; t_{29}=5.79; P<.001; d=1.06) to pursue a career in perinatal mental health. Similarly, psychology students also reported a significantly greater likelihood of considering a career in perinatal mental health following the simulation (t_{46}=7.04; P<.001; d=1.03).

Instructor Training

In addition to assessing the benefits for participants, to better understand how much time it would take to train staff without previous XR experience to become comfortable with navigating through this training platform, we asked our instructors to document their degree of confidence on a scale from 0 to 10 regarding four key dimensions: (1) hardware navigation, (2) software navigation, (3) avatar control, and (4) delivery of session learning outcomes. Following each session, instructors assigned ratings to these constructs, thereby creating a subjective trajectory of their session delivery proficiency (Figure 5). Notably, these ratings rose rapidly and plateaued after approximately 6 to 8 sessions across all key constructs, suggesting that it will take multiple training sessions before instructors feel that they can deliver reliably consistent training sessions. We also observed some variation from session to session, which may be accounted for by a combination of measurement errors and technical and logistical factors. While not amenable to formal statistical analysis, instructors reported lower scores when they experienced Wi-Fi dropouts or software crashes.

XSL•FO
RenderX

Figure 5. Instructor development over session delivery. Following each session, instructors self-reported the following on a scale of 1 to 10: (A) ease of hardware handling; (B) ease of software navigation; (C) confidence in controlling the avatar; and (D) comfort in achieving learning outcomes.



Discussion

Principal Findings

We explored the idea that XR technologies could support the delivery of mental health training through a simulated mental health consultation, in which a trainee interacts with a human-controlled virtual avatar. An initial feasibility pilot with subject-matter experts and students demonstrated potential efficacy worthy of further investigation. We subsequently followed this up with a comprehensive evaluation of its impact on trainees from across mental health nursing, medical doctors training to be GPs, and undergraduate psychology and medicine students. Our findings demonstrate the significant potential of XR as a pedagogical tool in supporting the development of mental consultation delivery skills.

We observed notable enhancements in cognitive and affective learning across all health care trainee groups. Instructors reported high rates of successful delivery of learning objectives, while participant groups reported increased knowledge in diverse perinatal domains, including the recognition of conditions, such as depression and anxiety, during pregnancy and in the postpartum period. Trainees demonstrated proficiency in systematic evaluation using diagnostic tools to assess severity. We also observed improvements in knowledge and confidence, specific to perinatal mental issues and broader issues of working with complex mental health challenges.

The integration of XR into mental health training represents a significant advancement, offering immersive, interactive, and repeatable learning environments that traditional methods often fail to replicate effectively. Conventional training typically relies on static case studies, peer-based role play, or interactions with real patients, each of which presents limitations. XR, however, combines high-quality instructional content with advanced technological features, including real-time feedback, iterative “fail and retry” opportunities, and high-fidelity simulations. This integration is not merely a shift in delivery medium but a holistic synthesis of content and technology, fostering experiential and contextually relevant learning.

The educational content needed to address complex and sensitive scenarios, such as perinatal mental health, is notably limited within mainstream mental health nursing curricula and GP training. Traditional training often emphasizes general psychiatric principles or common conditions, leaving significant gaps in specialized instructions for nuanced cases like acute postpartum psychosis and perinatal depression. This lack of exposure to high-risk sensitive clinical contexts underscores the need for innovative training solutions that can bridge this gap. XR-based simulations offer a tailored and immersive approach, allowing learners to engage with realistic perinatal

XSL•FO
RenderX

mental health cases and gain practical experience-driven insights beyond what conventional programs typically provide.

Although this study did not conduct a direct comparison with conventional approaches, XR's ability to standardize and replicate complex scenarios addresses many logistical and ethical challenges associated with actor- or patient-based training. By facilitating autonomous practice, enhancing critical competencies, and building learner confidence in a psychologically safe environment, XR provides a valuable environment for high-stake contexts such as mental health consultations. Effective training in this domain is essential, as errors can negatively impact both therapeutic relationships and patient outcomes [36]. XR's capacity to support the development of these therapeutic relationships is key to achieving improved health outcomes for individuals with mental illness [37]. This work also suggests that immersive educational technologies might be able to influence career planning and specialization. Our study found an increase in the reported interest among trainees considering a career in perinatal mental health. This positive shift in attitude toward perinatal mental health careers is particularly significant given the documented shortage in this specialty [9]. Such tools may extend beyond traditional educational outcomes to influence career aspirations and potentially bridge the gap between abstract career concepts and tangible professional identity formation.

Immersive educational technologies, exemplified by XR simulations, possess the potential to not only shape career preferences but also address significant concerns regarding the cultivation of empathetic connections and the practical application of theoretical knowledge during training. In mental health training, a crucial aspect involves nurturing the user's ability to establish therapeutic relationships. This necessitates engaging in specific scenarios and subsequent reflection to ensure nurses can comprehensively apply theoretical knowledge effectively [38]. We were concerned that the interaction with a virtual avatar may be a poor substitute for the development of this relationship and that it may be difficult to empathize with. However, our investigation into users' social and emotional interactions within the simulation revealed positive indicators, including general and spatial presence and improvements across cognitive and affective domains. These promising outcomes suggest that immersive technologies may not act as barriers but instead as facilitators in establishing effective therapeutic relationships.

Further grounds for our concerns about the feasibility of this tool in this context came from the "Uncanny Valley" [39] phenomenon, which describes the sense of unease or discomfort experienced when an artificial representation closely resembles a human but is not quite convincingly lifelike. Stacey had indeed been designed to be as realistic as possible (working within the graphical constraints of today's technology). Our outcomes indicate that the design quality and the method for interacting with the avatar were sufficient to circumvent this effect, allowing users to transcend potential unease and engage meaningfully with the simulation. Nevertheless, somewhat paradoxically, as the graphical capabilities of XR technology increase, this area will become increasingly more important to monitor in the design and implementation of patient avatars until they become indistinguishable from real humans. This necessitates a careful iterative approach in the design and implementation of patient avatars, one that is cognizant of these psychological effects. Future iterations of XR simulations must be not only technically advanced but also underpinned by a deep understanding of user psychology to ensure that they support rather than detract from the learning objectives [40].

In looking to the future, the rapid advances being made in generative AI provide an avenue for such training tools to become increasingly autonomous, which could significantly alleviate the workload of instructors while simultaneously enhancing the dynamic interactivity of training sessions through the development of bespoke patient avatars tailored to the needs of learners. AI analysis of utterance-response pairs could predict context-specific reactions, enabling intelligent and adaptive XR training tools. XR training tools could leverage this "generative" AI to create dynamic and realistic scenarios for training health care professionals in mental health consultations, thereby enhancing their ability to understand and respond to a wide range of patient interactions. The use of generative AI could also democratize access to high-quality training resources, making them available across different geographies and socioeconomic contexts, thereby potentially reducing disparities in mental health training quality globally. Instructors could personalize scenarios, offer real-time feedback, and adapt to unique learner needs. Such potential advances do, however, raise ethical concerns [41], including the risk of bias that would need to be tackled for effective, efficient, and inclusive training.

Limitations

It is important to note that this study does not suggest that XR learning can replace traditional placements or direct learning opportunities and experiences or that simulation avatars can fully replicate real patients. What it does show is that XR could be a valuable tool for providing standardized training experiences to mental health trainees across different institutions and professional domains. The simulation employed in this study serves as a potential solution for exposing trainees to complex and nonroutine patient presentations. Going a step further, we suggest that the tool could also offer an opportunity to explore underrepresented scenarios, including those involving minoritized populations, and could be a useful vehicle for promoting cultural competence and enhancing the overall diversity of training scenarios. We propose that by using XR technology, mental health training programs may be able to bridge gaps in exposure to various clinical scenarios and populations, contributing to a more comprehensive and inclusive approach to mental health training.

While this study demonstrated significant improvements in various aspects of trainee confidence and perceived competence, it is important to clarify that the study's primary aim was not to evaluate current educational provisions or compare XR training directly to traditional methods. Instead, the focus was on assessing the feasibility and potential benefits of an XR-based tool as a supplementary learning aid within existing training frameworks. The intention was to explore how XR could augment current educational experiences rather than to position it as a replacement for established training methods. Future

XSL•FO
RenderX

research should consider comparative studies that directly assess the effectiveness of XR against traditional pedagogical approaches to determine the conditions under which XR-based learning is most beneficial. Incorporating controlled trials and longitudinal assessments would further strengthen the understanding of XR's role in skill retention and clinical application. It is also important to note that our evaluation only involved a single session and an examination of changes immediately after the session. This has shown substantial promise and must be followed up with an examination of any longer-term changes, capturing skill retention and whether this knowledge and confidence can be translated to clinical practice. Equally, the implementation of this technology into the curriculum should not be a "one-shot" standalone affair. Instead, we propose that it should be integrated systematically across multiple sessions to reinforce and build upon the acquired knowledge and skills. Long-term evaluations, including follow-up assessments at intervals beyond the immediate postsession period, are imperative to gauge the durability and sustainability of the observed impacts. Additionally, future research endeavors should explore the application of XR technology in diverse clinical scenarios to assess its versatility and effectiveness across various health care contexts. The iterative and continuous integration of XR simulations into the curriculum, coupled with ongoing assessments, will contribute to a more comprehensive understanding of its benefits and practical applicability in real-world health care settings.

While we focused on evaluating one-to-one sessions, the platform also affords the delivery of one-to-many training sessions and the opportunity for group-led discussion. One-to-one sessions in XR offer personalized interactions where trainees can practice engaging with virtual patients in a safe controlled environment, receiving tailored feedback from instructors. This approach allows for intensive skill development, particularly in handling complex or sensitive mental health scenarios. On the other hand, one-to-many sessions leverage XR's multi-user capabilities to enable group training, where multiple participants can observe and interact within the same virtual environment. By leveraging its multi-user capabilities, XR training tools could be used to create an environment conducive to collaboration, group discussion, and the promotion of intra- and interprofessional discussions.

Furthermore, in a world where hybrid (or blended) learning has started to become a norm, XR provides a practical solution for overcoming resource and time constraints faced by training programs. The ability to access training sessions and share the same learning space from anywhere in the world could provide a practical solution to the resource and time constraints faced by training programs, promoting both inclusivity and efficiency in health care education. This flexibility ensures that trainees across diverse locations and professional domains can participate in standardized training experiences, contributing to equitable and scalable mental health education.

While immersive technologies present transformative opportunities as learning tools, accessibility for individuals with visual and auditory impairments remains a critical concern. XR environments heavily rely on visual and auditory inputs, which can exclude users with disabilities if not adequately addressed. For visual impairments, accessibility may involve features, such as screen reader compatibility, audio descriptions, and haptic feedback, to convey spatial or contextual information. For auditory impairments, captions, subtitles, and integration with assistive hearing devices, such as cochlear implants, are essential. To address these challenges, the software in this study integrates specific accessibility features. For users with hearing impairments, the JoinXR platform includes automatic captioning, enabling subtitles to appear beneath a user's avatar during interactions. For users with visual impairments, the design process emphasized hardware compatibility, recommending the HTC Vive Focus 3 for VR due to its adjustable lenses and focal settings and the Microsoft HoloLens 2 for AR, which allows users to keep their glasses on. These measures reflect a commitment to inclusivity, though further advancements are needed to fully overcome accessibility barriers in XR technologies.

Finally, important considerations for the implementation of XR training tools are the economic cost and the return on investment. There are significant start-up expenditures, including the procurement of XR hardware and software licenses. In addition, adopting XR technology requires appropriate technical infrastructure, such as accessible, reliable, and reasonably fast internet connectivity, along with a long-term strategy for sustainable implementation. The rapid pace of technological advancement poses the risk of hardware and software becoming quickly obsolete, compelling organizations to contemplate strategies for regular updates and maintenance to keep pace with technological innovations. On the other hand, XR technology affords numerous opportunities to enhance educational experiences, reducing training time and improving learning efficacy [42]. Additionally, the potential of XR to facilitate remote learning could reduce the necessity for travel and accommodation expenses, which are traditionally associated with centralized training programs [43]. A critical next step for advancing this field is the development of a return-on-investment framework. This framework should account for the wide spectrum of benefits as well as the initial and ongoing expenses. In this way, organizations will have clear insights into the viability and value of adopting tools, such as the one introduced here, as they address the escalating demands of health care workforce training.

Conclusions

The use of an XR-based simulated mental health consultation scenario, where trainees interacted with a human-controlled virtual avatar, showed promise in an initial feasibility pilot and was further substantiated by a comprehensive evaluation across various health care trainee groups. Our findings indicate significant enhancements in cognitive and affective learning, with high rates of successful delivery of learning objectives. These findings show, for the first time, that XR can be used to provide an effective, standardized, and reproducible tool for trainees to develop their mental health consultation skills. We suggest that XR could provide a solution to overcome the current resource challenges associated with equipping current and future health care professionals, which are likely to be exacerbated by workforce expansion plans.

XSL•FO
RenderX

Acknowledgments

This project was funded by Health Education England, which is now part of National Health Service (NHS) England. Authors RM and FM are supported in part by the National Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre (NIHR203331). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care. FM is further supported by the European Union's Horizon Research and Innovation Programme under grant agreement number 101070155 and UK Research and Innovation (UKRI) through the Horizon Europe Guarantee (#10039307). The authors would like to thank Dr Amy Micklethwaite, Dr James Bullock, Dr Mayur Vibhuti, Dr William Edney, Dr Matthew Bull, Dr Richard Elliott, Dr Giles Berrisford, and Dr Jelena Jankovic for their contributions as subject matter experts in the development of the simulation evaluated in this paper.

Authors' Contributions

Conceptualization: RBD, DP, DPW, CG, JH, MKL, CB
Data curation: KH, ZBM, LT, FM
Formal analysis: KH, ZBM, FM
Funding acquisition: RM, FM
Investigation: KH, ZBM, LT, FM
Methodology: KH, ZBM, LT, RBD, DP, DPW, RM, FM
Project administration: KH, ZBM, JH, SR, RM, FM
Resources: RBD, DP, MKL, CB
Software: MKL, CB
Supervision: ZBM, FM
Validation: FM
Visualization: KH, FM
Writing – original draft: KH, RBD, FM
Writing – review & editing: KH, ZBM, LT, RBD, DP, CG, CM, SR, MKL, CB, NR, AC, RM, FM

Conflicts of Interest

Authors from the University of Leeds (KH, ZBM, LT, RM, AC, and FM) declare no conflicts of interest relating to this study and undertook data collection and analysis independent of the rest of the authorship team. Authors MKL (Founder, Fracture Reality) and CB (Product Manager, Fracture Reality) led the development of the application. They were not involved in data collection or analysis and did not contribute to the discussion section of this manuscript. Co-authors RBD and DP, Health Education England (now NHS England) contributed to the development of the simulation scenarios created by Fracture Reality and to the development of the research project and were not involved in data collection or analysis. DPW, CG, JH, and SR were involved at the project supervisory level from Health Education England (now NHS England) and were not involved in data collection or analysis.

Multimedia Appendix 1
Process flow chart of the hour-long evaluation session.
[DOCX File , 37 KB - mededu_v11i1e64619_app1.docx]

Multimedia Appendix 2
In-simulation prompts available to the instructors to support users.
[DOCX File , 48 KB - mededu_v11i1e64619_app2.docx]

Multimedia Appendix 3
Perinatal Mental Health Familiarity and Awareness Scale (PMHAFS).
[DOCX File , 15 KB - mededu_v11i1e64619_app3.docx]

References

1. Committee of Public Accounts. Progress in improving NHS mental health services. UK Parliament. URL: https://publications.parliament.uk/pa/cm5803/cmselect/cmpubacc/1000/report.html [accessed 2025-02-24]
2. Mental health services: addressing the care deficit. NHS Providers. URL: https://nhsproviders.org/mental-health-services-addressing-the-care-deficit [accessed 2025-02-24]

XSL•FO
RenderX

3. Adams R, Ryan T, Wood E. Understanding the factors that affect retention within the mental health nursing workforce: a systematic review and thematic synthesis. Int J Ment Health Nurs 2021 Dec 28;30(6):1476-1497 [FREE Full text] [doi: 10.1111/inm.12904] [Medline: 34184394]
4. Oates J, Jones J, Drey N. Subjective well-being of mental health nurses in the United Kingdom: Results of an online survey. Int J Ment Health Nurs 2017 Aug 23;26(4):391-401. [doi: 10.1111/inm.12263] [Medline: 27878917]
5. Imrie R. Communication skills for consultations about mental health problems. InnovAiT: Education and Inspiration for General Practice 2015 Mar 03;8(4):246-251. [doi: 10.1177/1755738015570980]
6. Mitchell AJ, Rao S, Vaze A. International comparison of clinicians' ability to identify depression in primary care: meta-analysis and meta-regression of predictors. Br J Gen Pract 2011 Feb 01;61(583):e72-e80. [doi: 10.3399/bjgp11x556227]
7. Magnée T, de Beurs DP, de Bakker DH, Verhaak PF. Consultations in general practices with and without mental health nurses: an observational study from 2010 to 2014. BMJ Open 2016 Jul 18;6(7):e011579 [FREE Full text] [doi: 10.1136/bmjopen-2016-011579] [Medline: 27431902]
8. Perinatal mental health services: Recommendations for the provision of services for childbearing women. Royal College of Psychiatrists. URL: https://www.rcpsych.ac.uk/docs/default-source/improving-care/better-mh-policy/college-reports/college-report-cr232---perinatal-mental-heath-services.pdf [accessed 2025-02-24]
9. Noonan M, Galvin R, Jomeen J, Doody O. Public health nurses' perinatal mental health training needs: A cross sectional survey. J Adv Nurs 2019 Nov 27;75(11):2535-2547. [doi: 10.1111/jan.14013] [Medline: 30937923]
10. King JD, Crowley G, El-Maraghy M, Davis W, Jauhari A, Wilson-Jones C. Perinatal mental health in medical school curricula: a national scoping survey of British universities and student psychiatry societies. BJPsych Bull 2024 Feb 12;48(1):51-56 [FREE Full text] [doi: 10.1192/bjb.2022.91] [Medline: 36632805]
11. Sadeghi AH, Peek JJ, Max SA, Smit LL, Martina BG, Rosalia RA, et al. Virtual reality simulation training for cardiopulmonary resuscitation after cardiac surgery: face and content validity study. JMIR Serious Games 2022 Mar 02;10(1):e30456 [FREE Full text] [doi: 10.2196/30456] [Medline: 35234652]
12. Lawshe C. A quantitative approach to content validity. Personnel Psychology 2006 Dec 07;28(4):563-575. [doi: 10.1111/j.1744-6570.1975.tb01393.x]
13. Ayre C, Scally AJ. Critical values for Lawshe's Content Validity Ratio. Measurement and Evaluation in Counseling and Development 2017 Mar 08;47(1):79-86. [doi: 10.1177/0748175613513808]
14. Brooke J. SUS: A 'Quick and Dirty' Usability Scale. In: Usability Evaluation In Industry. Boca Raton, FL: CRC Press; 1996.
15. Taylor L, Dyer T, Al-Azzawi M, Smith C, Nzeako O, Shah Z. Extended reality anatomy undergraduate teaching: A literature review on an alternative method of learning. Ann Anat 2022 Jan;239:151817. [doi: 10.1016/j.aanat.2021.151817] [Medline: 34391910]
16. Arpaia P, De Benedetto E, De Paolis L, D'Errico G, Donato N, Duraccio L. Performance and usability evaluation of an extended reality platform to monitor patient's health during surgical procedures. Sensors (Basel) 2022 May 21;22(10):3908 [FREE Full text] [doi: 10.3390/s22103908] [Medline: 35632317]
17. Donekal Chandrashekar N, Manuel M, Park J, Greene A, Safford S, Gračanin D. An Extended Reality Simulator for Advanced Trauma Life Support Training. In: Chen JYC, Fragomeni G, editors. Virtual, Augmented and Mixed Reality: Applications in Education, Aviation and Industry. HCII 2022. Lecture Notes in Computer Science, vol 13318. Cham: Springer; 2022:31-44.
18. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. Journal of Usability Studies 2009;4(3):114-123 [FREE Full text] [doi: 10.5555/2835587.2835589]
19. Kim SK, Lee Y, Yoon H, Choi J. Adaptation of extended reality smart glasses for core nursing skill training among undergraduate nursing students: usability and feasibility study. J Med Internet Res 2021 Mar 02;23(3):e24313 [FREE Full text] [doi: 10.2196/24313] [Medline: 33650975]
20. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. The International Journal of Aviation Psychology 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
21. Tasdemir C, Gazo R. Integrating sustainability into higher education curriculum through a transdisciplinary perspective. Journal of Cleaner Production 2020 Aug;265:121759. [doi: 10.1016/j.jclepro.2020.121759]
22. Hansen J, Ramachandran R, Vockley J. Survey of health care provider understanding of gene therapy research for inherited metabolic disorders. Clin Ther 2022 Aug;44(8):1045-1056 [FREE Full text] [doi: 10.1016/j.clinthera.2022.07.002] [Medline: 35927093]
23. Wang L, Chen M. The effects of game strategy and preference‐matching on flow experience and programming performance in game‐based learning. Innovations in Education and Teaching International 2010 Feb;47(1):39-52. [doi: 10.1080/14703290903525838]
24. Pintrich PR, Smith D, Garcia T, McKeachie W. A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ). ERIC. URL: https://eric.ed.gov/?id=ED338122 [accessed 2025-02-24]

XSL•FO
RenderX

25. Ryan RM, Deci EL. Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. Contemporary Educational Psychology 2020 Apr;61:101860. [doi: 10.1016/j.cedpsych.2020.101860]
26. Unver V, Basak T, Watts P, Gaioso V, Moss J, Tastan S, et al. The reliability and validity of three questionnaires: The Student Satisfaction and Self-Confidence in Learning Scale, Simulation Design Scale, and Educational Practices Questionnaire. Contemp Nurse 2017 Feb 10;53(1):60-74. [doi: 10.1080/10376178.2017.1282319] [Medline: 28084900]
27. Jeffries PR, Rizzolo MA. Designing and Implementing Models for the Innovative Use of Simulation to Teach Nursing Care of Ill Adults and Children: A National, Multi-Site, Multi-Method Study. National League for Nursing. URL: https://www.nln.org/docs/default-source/uploadedfiles/professional-development-programs/read-the-nln-laerdal-project-summary-report-pdf.pdf [accessed 2025-02-24]
28. Franklin AE, Burns P, Lee CS. Psychometric testing on the NLN Student Satisfaction and Self-Confidence in Learning, Simulation Design Scale, and Educational Practices Questionnaire using a sample of pre-licensure novice nurses. Nurse Educ Today 2014 Oct;34(10):1298-1304. [doi: 10.1016/j.nedt.2014.06.011] [Medline: 25066650]
29. Witmer BG, Singer MJ. Measuring presence in virtual environments: a presence questionnaire. Presence 1998 Jun;7(3):225-240. [doi: 10.1162/105474698565686]
30. Berkman MI, Akan E. Presence and Immersion in Virtual Reality. In: Lee N, editor. Encyclopedia of Computer Graphics and Games. Cham: Springer; 2019:1-10.
31. Mania K, Chalmers A. The effects of levels of immersion on memory and presence in virtual environments: a reality centered approach. Cyberpsychol Behav 2001 Apr;4(2):247-264. [doi: 10.1089/109493101300117938] [Medline: 11710251]
32. Slater M, Neyret S, Johnston T, Iruretagoyena G, Crespo M, Alabèrnia-Segura M, et al. An experimental study of a virtual reality counselling paradigm using embodied self-dialogue. Sci Rep 2019 Jul 29;9(1):10903 [FREE Full text] [doi: 10.1038/s41598-019-46877-3] [Medline: 31358846]
33. Makransky G, Terkildsen TS, Mayer RE. Adding immersive virtual reality to a science lab simulation causes more presence but less learning. Learning and Instruction 2019 Apr;60:225-236. [doi: 10.1016/j.learninstruc.2017.12.007]
34. Schubert TW. The sense of presence in virtual environments:. Zeitschrift für Medienpsychologie 2003 Apr;15(2):69-71. [doi: 10.1026//1617-6383.15.2.69]
35. Schubert T, Friedmann F, Regenbrecht H. The experience of presence: factor analytic insights. Presence: Teleoperators & Virtual Environments 2001 Jun;10(3):266-281. [doi: 10.1162/105474601300343603]
36. Nyttingnes O, Ruud T, Rugkåsa J. 'It's unbelievably humiliating'-Patients' expressions of negative effects of coercion in mental health care. Int J Law Psychiatry 2016 Nov;49(Pt A):147-153. [doi: 10.1016/j.ijlp.2016.08.009] [Medline: 27726890]
37. Wolsing SK, Hjorth P, Løkke A, Hilberg O, Frølund J. Experiences of receiving a medical consultation - an interview study among hospitalized psychiatric patients. Nord J Psychiatry 2024 Oct 22;78(7):583-590 [FREE Full text] [doi: 10.1080/08039488.2024.2373251] [Medline: 39037071]
38. Skår R. Knowledge use in nursing practice: the importance of practical understanding and personal involvement. Nurse Educ Today 2010 Feb;30(2):132-136. [doi: 10.1016/j.nedt.2009.06.012] [Medline: 19631424]
39. Mori M. The Uncanny Valley: The Original Essay by Masahiro Mori. IEEE Spectrum. URL: http://umnikizdes.ru/aways/web.ics.purdue.edu/~drkelly/MoriTheUncannyValley1970.pdf [accessed 2025-02-24]
40. MacDorman KF, Chattopadhyay D. Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. Cognition 2016 Jan;146:190-205 [FREE Full text] [doi: 10.1016/j.cognition.2015.09.019] [Medline: 26435049]
41. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med 2022 Jan 20;28(1):31-38. [doi: 10.1038/s41591-021-01614-0] [Medline: 35058619]
42. Huang CL, Luo YF, Yang SC, Lu CM, Chen A. Influence of students' learning style, sense of presence, and cognitive load on learning outcomes in an immersive virtual reality learning environment. Journal of Educational Computing Research 2019 Aug 05;58(3):596-615. [doi: 10.1177/0735633119867422]
43. Cheng K, Tsai C. A case study of immersive virtual field trips in an elementary classroom: Students' learning experience and teacher-student interaction behaviors. Computers & Education 2019 Oct;140:103600. [doi: 10.1016/j.compedu.2019.103600]

Abbreviations

AI: artificial intelligence
AR: augmented reality
CVR: content validity ratio
GP: general practitioner
IPQ: iGroup Presence Questionnaire
NHS: National Health Service
SUS: System Usability Scale
VR: virtual reality

XSL•FO
RenderX

VRSQ: Virtual Reality Sickness Questionnaire
XR: extended reality

Edited by J Moen; submitted 22.07.24; peer-reviewed by S Smith, M Pritchard; comments to author 18.12.24; revised version received 07.02.25; accepted 11.02.25; published 02.04.25.

Please cite as:
Hiley K, Bi-Mohammad Z, Taylor L, Burgess-Dawson R, Patterson D, Puttick-Whiteman D, Gay C, Hiscoe J, Munsch C, Richardson S, Knowles-Lee M, Beecham C, Ralph N, Chatterjee A, Mathew R, Mushtaq F
Extended Reality–Enhanced Mental Health Consultation Training: Quantitative Evaluation Study
JMIR Med Educ 2025;11:e64619
URL: https://mededu.jmir.org/2025/1/e64619
doi:10.2196/64619
PMID:

©Katherine Hiley, Zanib Bi-Mohammad, Luke Taylor, Rebecca Burgess-Dawson, Dominic Patterson, Devon Puttick-Whiteman, Christopher Gay, Janette Hiscoe, Chris Munsch, Sally Richardson, Mark Knowles-Lee, Celia Beecham, Neil Ralph, Arunangsu Chatterjee, Ryan Mathew, Faisal Mushtaq. Originally published in JMIR Medical Education (https://mededu.jmir.org), 02.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring the Impact of the COVID-19 Pandemic on Learning Experience, Mental Health, Adaptability, and Resilience Among Health Informatics Master's Students: Focus Group Study

Nadia Davoody[1], MSc, PhD; Natalia Stathakarou[1], MSc; Cara Swain[1,2], MBChB; Stefano Bonacina[1], MSc, PhD

[1]Health Informatics Centre, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden
[2]Academic Department of Military Surgery & Trauma, Royal Centre for Defence Medicine, Birmingham, United Kingdom

Corresponding Author:
Nadia Davoody, MSc, PhD
Health Informatics Centre, Department of Learning, Informatics, Management and Ethics
Karolinska Institutet
Tomtebodavägen 18 A
Stockholm, S-17177 Stockholm
Sweden
Phone: 46 08 524 864 86
Email: nadia.davoody@ki.se

Abstract

Background: The shift to online education due to the COVID-19 pandemic posed significant challenges and opportunities for students, affecting their academic performance, mental well-being, and engagement.

Objective: This study aimed to explore the overall learning experience among health informatics master's students at Karolinska Institutet, Sweden, and the strategies they used to overcome learning challenges posed by the COVID-19 pandemic.

Methods: Through 3 structured focus groups, this study explored health informatics master's students' experiences of shifting learning environments for classes that started in 2019, 2020, and 2021. All focus group sessions were recorded and transcribed verbatim. Inductive content analysis was used to analyze the data.

Results: The results highlight the benefits of increased autonomy and flexibility and identify challenges such as technical difficulties, diminished social interactions, and psychological impacts. This study underscores the importance of effective online educational strategies, technological preparedness, and support systems to enhance student learning experiences during emergencies. The findings of this study highlight implications for educators, students, and higher education institutions to embrace adaptation and foster innovation. Implications for educators, students, and higher education institutions include the need for educators to stay current with the latest educational technologies and design teaching strategies and pedagogical approaches suited to both online and in-person settings to effectively foster student engagement. Students must be informed about the technological requirements for online learning and adequately prepared to meet them. Institutions play a critical role in ensuring equitable access to technology, guiding and supporting educators in adopting innovative tools and methods, and offering mental health resources to assist students in overcoming the challenges of evolving educational environments.

Conclusions: This research contributes to understanding the complexities of transitioning to online learning in urgent circumstances and offers insights for better preparing educational institutions for future pandemics.

(JMIR Med Educ 2025;11:e63708) doi:10.2196/63708

KEYWORDS

COVID-19 pandemic; eHealth; blended learning; health informatics; higher education adaptation

Introduction

Background

The COVID-19 pandemic disrupted education worldwide, and many universities were required to shift to online education despite most being unprepared for such a shift [1]. The education response during the early phase of the COVID-19 pandemic focused on implementing remote learning modalities as an emergency response and mostly on the online delivery of educational material. The result of these efforts was a substantive rise in e-learning, whereby teaching and learning

XSL•FO
RenderX

activities took place remotely via digital platforms. After the pandemic, the use of e-learning was expected to grow [1].

e-Learning is the use of internet technologies to enhance knowledge and performance. e-Learning technologies offer learners control over the content, learning sequence, pace of learning, time, and often media, allowing them to tailor their experiences to meet their learning objectives [2]. Within the context of the COVID-19 pandemic, e-learning and online learning refer to remote teaching strategies and methods that universities used to urgently respond to the requirements of health protocols and restrictions in mobility [3].

While e-learning holds significant potential, the rapid deployment of strategies during the pandemic showed mixed results. Most of the swift adoption focused on reaching all students and enhancing accessibility, but educators lacked the time to develop and implement pedagogical strategies that could enhance the learning experience [3]. Strategies used for short-term online education may not have been suitable for the prolonged disruption caused by the pandemic.

Student experiences with e-learning during the pandemic were mixed. Some studies reported a “new normal” that students viewed positively, citing benefits such as education continuity, increased accessibility, stronger learner-lecturer interactions, and greater confidence in expressing themselves in an online learning environment [1,4-7].

Within the field of medical education, one study reported that students responded positively to the transition to online learning methods, with a notable improvement in student satisfaction related to course structure [8]. Another study highlighted medical students’ generally positive attitudes toward e-learning during the pandemic. However, while e-learning was seen as a necessary and beneficial alternative, challenges were recognized, such as increased stress and anxiety, limited internet access, technical difficulties, and reduced hands-on clinical training [9]. Health profession education programs have reported issues in transitioning to online learning and maintaining continuity in education [10]. There have been difficulties related to adapting traditional teaching methods to online formats and the stress this placed on both students and educators [11]. Stress, unfamiliarity with online classrooms, uncertainty about academic futures, and the rapid shift to e-learning contributed to negative experiences for many students [12,13]. In addition, insufficient training, inadequate internet infrastructure in some countries, and lack of preparation led to poor student experiences, undermining the sustainability of e-learning [14].

The contradictory evidence on diverse experiences of e-learning during the pandemic underscores the need for further investigation within this area. To better prepare for future crises, higher education institutions need to understand the experiences of both educators and students and develop educational strategies for online learning during emergencies. Although many studies have investigated students’ learning experiences during the pandemic, there are limited insights into the specific challenges and strategies used to overcome them. Contextual factors such as internet access, cultural differences, and subject of study likely influenced the experiences and perceptions of both students and educators. Therefore, there is a need to further explore e-learning experiences and strategies across different contexts and subjects of study. Health informatics education uniquely integrates theoretical knowledge with applied technical and health care–related skills, which may have been significantly affected by the shift to online learning.

Study Aim and Research Question

This study aimed to explore the overall learning experience of health informatics master’s students at Karolinska Institutet (KI), Sweden, and the strategies they used to overcome the learning challenges posed by the COVID-19 pandemic. This study addressed the following research question: How did health informatics master’s students at KI experience learning during the COVID-19 pandemic, and what strategies did they use to overcome related challenges?

Methods

Study Design

In this study, we conducted 3 semistructured focus group interviews comprising a total of 16 registered students and alumni of the Master’s Programme in Health Informatics. The COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines [15] were used for reporting the results. The process was piloted, and questions for the participants were determined in advance by authors ND, SB, and NS adapted from previous studies on the impact of the COVID-19 pandemic on medical education [9] and on teaching and learning in health professional education [11].

In this explorative study, focus group interviews were chosen as a data collection technique as they allow for the generation of rich qualitative data through group interactions and dynamics, which is beneficial when addressing an exploratory question. Participants can build on each other’s ideas, leading to more nuanced insights [16,17]. During these focus groups, the students could share their experiences and perspectives as part of an open-ended discussion, leading to a deeper understanding of their learning experiences and the strategies used to overcome challenges during the COVID-19 pandemic.

Given the exploratory study design, focus groups are ideal as a data collection method as they facilitate open-ended discussion and are effective in collecting rich data that might not have been possible to collect through more structured data collection methods. The focus groups were held remotely using the Zoom platform (Zoom Video Communications) with CS in the role of interviewer.

Study Setting and Participants

The context of the study was the Master’s Programme in Health Informatics provided by the Department of Learning, Informatics, Management, and Ethics at KI. It is a 2-year global program run jointly with Stockholm University. The program is designed for students with an interest in IT and how it can be applied to the fields of medicine and health care. As such, the students may have either a technical or health care background.

Convenience sampling was used to recruit participants. Individuals who had participated in the master’s program during the COVID-19 pandemic were invited to contribute. Therefore,

XSL•FO
RenderX

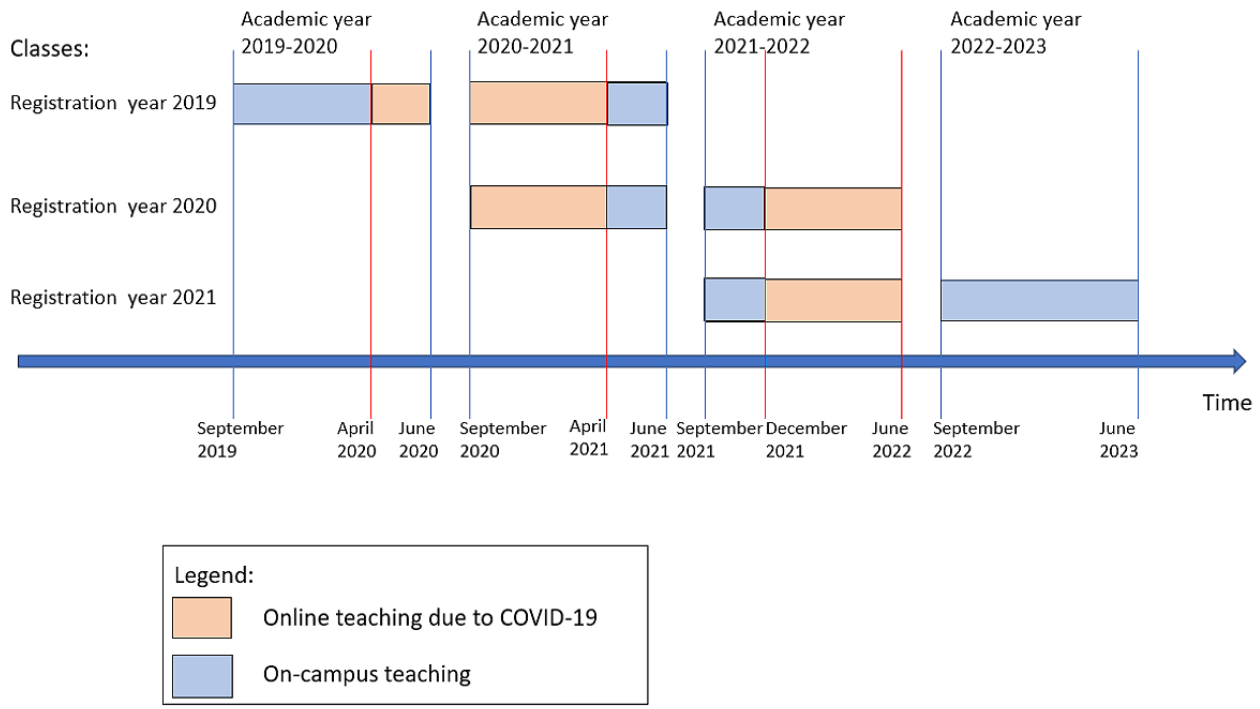
all participants were alumni or active registered students split into 3 cohorts: students who had registered in 2019 and graduated in 2021 (cohort 1), students who started in 2020 and graduated in 2022 (cohort 2), and students who started in 2021 and graduated in 2023 (cohort 3). Table 1 provides some characteristics of the participants. Most participants (13/16, 81%) were living in Sweden; however, as some of the students had relocated to their native countries or were working overseas, this informed the decision to conduct online focus groups.

Figure 1 shows the timeline of how the Master's Programme in Health Informatics from the academic year 2019-2020 to 2022-2023 was conducted. On-campus teaching and online teaching due to the COVID-19 pandemic are indicated using distinct colors. Students that started the program in September 2020 mostly received online teaching.

Table 1. Participant characteristics.

	Background	Experience with online learning
2019_Participant 1	Health care	No
2019_Participant 2	Health care	Yes
2019_Participant 3	Health care	No
2020_Participant 1	IT or technical	No
2020_Participant 2	Health care	Yes
2020_Participant 3	Health care	Yes
2020_Participant 4	Health care	Yes
2020_Participant 5	Health care	Yes
2020_Participant 6	Health care	No
2020_Participant 7	IT or technical	No
2020_Participant 8	Health care	Yes
2021_Participant 1	IT or technical	Yes
2021_Participant 2	IT or technical	Yes
2021_Participant 3	Health care	Yes
2021_Participant 4	Health care	Yes
2021_Participant 5	Health care	Yes

Figure 1. On-campus and online teaching during the COVID-19 pandemic.



XSL•FO
RenderX

Data Collection and Analysis

A total of 3 online focus group interviews were conducted during 2023 using the Zoom platform; the interviews were recorded and transcribed verbatim. To reduce familiarity bias in the study, CS, who had not been involved in teaching within the health informatics program, conducted the focus groups. The first focus group included 3 students from cohort 1, the second focus group included 8 students from cohort 2, and the third and final focus group included 5 students from cohort 3. The first focus group had a duration of 45 minutes, with the other 2 focus groups lasting approximately 1 hour and 30 minutes. The piloting process did not generate any data and was not included in the analysis.

Generated data were analyzed using inductive content analysis [18]. The inductive content analysis allowed for themes and patterns to be constructed directly from the data that were grounded in the students' actual experiences rather than imposing preconceived categories. A combination of coding methods was used: descriptive coding summarizes the main topics of the text, and pattern coding was used to condense meaning units into broader patterns to group the initial codes into broader themes. Pattern coding helps identify and understand the broader patterns and relationships within the data [19]. ND, NS, and SB conducted the initial coding. Each coder was instructed to familiarize themselves with the data, read through the entire dataset to gain an overall understanding before starting the coding process, identify meaning units, condense them, and assign codes that captured the essence of each unit. They independently reviewed all transcribed interviews, dividing the responsibilities for identifying relevant meaning units and conducting the initial coding. CS joined the analytical process once initial coding and subcategories had been generated. Upon identifying subcategories, a comparative analysis was conducted to reveal similarities and differences among student groups. Comparative analysis was conducted through peer debriefing sessions in which ND, NS, and SB compared their assigned codes and discussed differences in interpretations. All authors reviewed discrepancies collaboratively until a consensus was reached. Discrepancies were resolved by re-examining the raw data for the meaning unit in question, discussing interpretations considering the research aim, and refining the coding if needed. At this point, it was decided to collaboratively proceed with the categorization and identification of subthemes for all 3 student groups. This not only minimized the recurrence of redundant findings resulting from similarities among groups but also empowered us to emphasize the subcategories in which different student groups expressed distinct experiences or opinions.

Ethical Considerations

This research was carried out in Sweden. According to the Swedish Ethical Review Act, the research presented in our submitted manuscript did not require ethics approval as it did not handle sensitive personal information (as understood by the European General Data Protection Regulation). However, ethical requirements still apply, and written informed consent to take part in the study was obtained from all participants. The consent form outlined the study's purpose, potential risks and discomforts, the voluntary nature of participation, and the right to withdraw at any time. It also stated that no compensation would be provided for participation. Participants were assured that their confidentiality and privacy would be preserved.

Results

Overview of Data Collection and Data Analysis

Our analysis of the generated focus group data identified 1 main theme—*adapting to hybrid learning*—and three subthemes: (1) students' considerations of learning during the pandemic, (2) moving between learning environments, and (3) students' well-being and engagement in learning. Each subtheme included several categories relevant to all student groups, summarized in Table 2.

The overarching theme of *adapting to hybrid learning* highlights the challenges and experiences faced by participants navigating the transition from traditional on-campus education to online learning and vice versa. It includes the autonomy that students seek in shaping their learning experiences, the integration of learning into their daily lives, and the varying perceptions of online education as an obligation rather than a choice during the COVID-19 pandemic. It further explores the changed aspects of social interaction, the struggle with engagement in online learning, the technical challenges faced, and the diverse technological readiness levels for different learning environments. In addition, this theme addresses the psychological impact of remote learning and the need for adequate support, the varied levels of student motivation, the influence of family and pet support, and the observed lack of networking opportunities and social interaction in online educational settings.

XSL•FO
RenderX

Table 2. An overview of the subcategories, categories, subthemes, and themes.

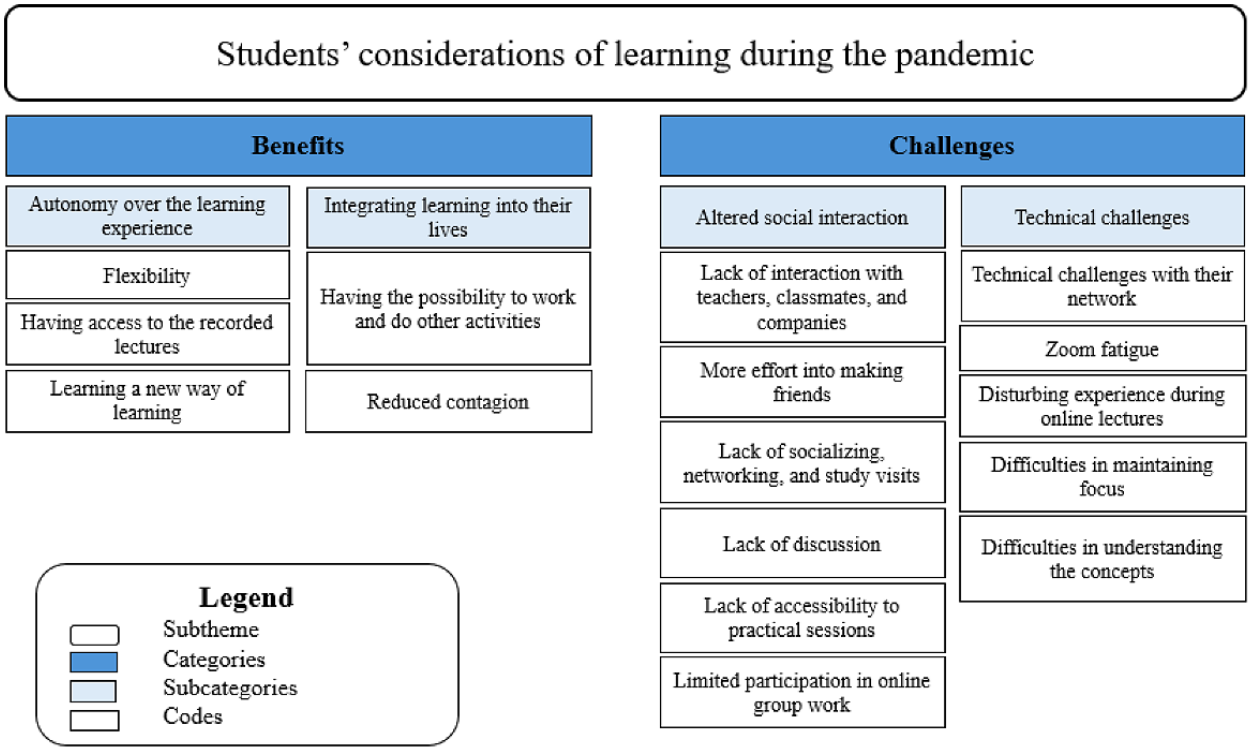
Theme, subtheme, and category	Subcategories
Adapting to hybrid learning	
Students' considerations of learning during the pandemic	
• Benefits	• Autonomy over the learning experience • Integrating learning into their lives
• Challenges	• Altered social interaction • Technical challenges
Moving between learning environments	
• Transition—campus to online education	• Online education as an obligation, not an option • Technological readiness for different learning environments • Psychological impact and support
• Transition—online to campus education	• Technological readiness for different learning environments • Lack of engagement in online learning • Different experiences in interaction with classmates and teachers between online and campus education
Students' well-being and engagement in learning	
• Motivation level • Strategies to stay motivated	• Varied levels of motivation • Group work as a connecting element • Family and pet support
• The impact of online learning on academic performance	• Autonomy • Lack of networking (social interaction)
• Overall experience and recommendations	• Improving online sessions • Improving educational support and delivery

Students' Considerations of Learning During the Pandemic

The students recognized that the changes and requirements necessitated by the COVID-19 pandemic brought both benefits and challenges. Figure 2 illustrates the relationship among the codes, subcategories, and categories for the subtheme *students' consideration of learning during the pandemic*.

XSL•FO
RenderX

Figure 2. A hierarchical structure displaying the subtheme and codes for theme 1—students' considerations related to learning during the pandemic.



Benefits

Benefits were characterized by increased *autonomy over their learning experience* and an ability to *integrate learning more readily into their lives*. Most of the participants mentioned that *flexibility* was one of the biggest benefits during the COVID-19 pandemic. Some mentioned that they could work from home or, in the case of many international students attending the Master's Programme in Health Informatics at KI, even work in their home country:

> *[The] main benefit is that it was less time-consuming. I was living at that time in [area_1] and I would have to travel to [area_2], that time was saved for me.* [2019_Participant 3]

> *I can start with one of the big benefits I had, especially in the beginning was that I was able to start the master's program from Germany where I am originally coming from. So, I moved I think 2 weeks in the program to Sweden. So, that was a big benefit to be more flexible in terms of the location for sure.* [2020_Participant 1]

Having access to the recorded lectures was appreciated as it made it possible for the students to go through the material whenever it was suitable for them. In addition, some of the students believed that they had *learned a new way of learning* and could now learn anything online. Online learning was believed to be more *efficient* as they could save time without the requirement to commute to campus:

> *...For me, it helped me to have the independence to learn as [participant] said. And yeah, and like now I can I have the feeling that I can learn anything online, I can just sign up for a course and in some videos and do it. And yeah, and in my current job, I did this a lot during the last six months. New technologies, new programming languages, everything...* [2020_Participant 7]

> *...I do watch tutorials on YouTube right now, but you get used to it. To manage to solve problems alone actually, which is really interesting because it's something that I do in my current job. If I don't know how to do something, I watch YouTube. I'm not going to ask anybody, and I think that's something with the pandemic as well...I managed to solve the problems myself.* [2020_Participant 4]

Having the possibility to work and do other activities in parallel with their studies was also appreciated by some of the participants:

> *In terms of benefits, it was very beneficial to save the commute time to the university by at least two hours a day. I was able to work in parallel with my studies. That wasn't a big deal. The third benefit, I would say [was] the flexibility to schedule meetings with colleagues in Group work. It's much easier than scheduling a physical meeting...* [2020_Participant 7]

As the risk of being infected and getting sick was high during the COVID-19 pandemic, the *reduced contagion* was also perceived as a benefit by many students as they lived with their families and wished to protect or shield them or, at least, lower the transmission risk:

> *Yes, of course. But before that, I would like to add one thing that I think it hasn't been mentioned by my fellow friend is that the advantages of online class or*

XSL•FO
RenderX

> *online learning during the pandemic is that we were able to refrain from the infection for COVID cases, of course, and I do believe that in Sweden it's well managed about the cases or the infection rate is remain low, but you know the preventive measures that we stay in our home and limiting interaction that's is also one.* [2021_Participant 5]

Challenges

The challenges that students experienced were divided into 2 distinct categories: *altered social interaction* and *technical challenges*. Most of the participants perceived the *lack of interaction with teachers, classmates, and companies* as a significant challenge of online learning. They experienced that the *group feeling* and the feeling of belonging to a bigger group was missing. Some students mentioned that they needed to put *more effort into making friends* and developing collegial relationships with their classmates. Several participants mentioned the challenges with the lack of *socializing, networking, and study visits* through online learning. This was noted through the *lack of discussion* due to people being shy and not turning on their cameras and by students leaving the online lectures directly after they finished:

> *...In this case, it's an international program. And I was not living there, so I moved to this country to learn. But also like to meet new people, make friends, to expand my network. The interaction with the teachers. And as yeah, [participant] said to others you just attend the meeting, and then you close it, and you don't have this interaction discussion afterward or during the class. So that was also a point, and I also like making friends. Of course, I made a lot of friends in this program. But you like the effort was bigger, you know? So, you have to be proactive to let's meet. But in an in-person or personal program, you have more facilities...* [2020_Participant 5]

> *...another aspect that at least made me feel a bit I shouldn't say depressed, but not as happy as I was not meeting all the people in the class. Partly built on that, you should form a team over two years, and that was for me very, very obvious that I almost missed that team spirit or the team.* [2021_Participant 1]

Lack of accessibility to practical sessions on campus and *limited participation in online group work* that might have resulted in the low quality of group work were other challenges mentioned by the students:

> *...I don't have technical problems with attending online lectures...the only thing, as I mentioned before, is we had no opportunity to have someone...to have an instructor while we were doing these programming things...having online lectures is fine but having a technical lab doesn't always work online.* [2019_Participant 2]

> *...regarding the challenge of collaboration...we had like several group assignments where we needed to have a lot of discussions...that was fairly difficult because we tend to have passive collaboration, I mean like normally one or two people lead while the others just agree it's totally different from what we have in the class where everyone can jump in and share their talks during the group work.* [2021_Participant 1]

Technical challenges associated with online learning were discussed by students of all cohorts. Some students mentioned that they were worried about the reliability of their network along with some *technical challenges with their network*. For some students, it was difficult to work with Zoom in the beginning. There was also an adjustment period required to use online platforms for longer periods, such as experiencing *Zoom fatigue* from having several hours of online classes. There were additional technical problems with hybrid sessions, and some of the participants mentioned the *disturbing experience* that they had during online learning as some people spoke at the same time, which disrupted the flow of the conversation. During hybrid learning sessions, there was a lack of interaction between those online and those present in the room on campus:

> *So, in the beginning, I mean [during] the lectures sometimes there were some network issues, and then when you are the first time sitting on Zoom, and everybody has their cameras turned off.* [2019_Participant 2]

> *...Zoom fatigue because I think during that period of time, we had two or three classes in one day. Even with a normal session like from 9:00 to 3:00, we felt really exhausted and it was worsened when we had the online classes, and I was thinking that perhaps some sessions could be minimized. I mean like normally we will like 90 minutes for a single theoretical class. But if online sessions can be reduced to about 30 or 40 minutes and the rest of the time, we could do our self-learning. I think that would have been more beneficial.* [2021_Participant 1]

> *...One negative aspect I noticed was that I didn't feel comfortable interrupting to ask a question. I know I would get distracted because it's an online setting, and sometimes when two people speak at the same time, it can be very disruptive...also there are always some technical difficulties.* [2021_Participant 2]

While some students found online learning beneficial as it allowed them to concentrate better during lectures by multitasking and engaging in other activities simultaneously, others reported *difficulties in maintaining focus*. These students struggled with reduced attention spans, often due to the distractions of being at home with their families or lack of camera use during online lectures, which contributed to a sense of disengagement:

> *...Then the other thing is that I cannot focus properly because my kids are small. When I was at home, it was distracting many, many times and difficult...* [2019_Participant 3]

As a result, some students experienced *difficulties in understanding the concepts* and needed some assistance:

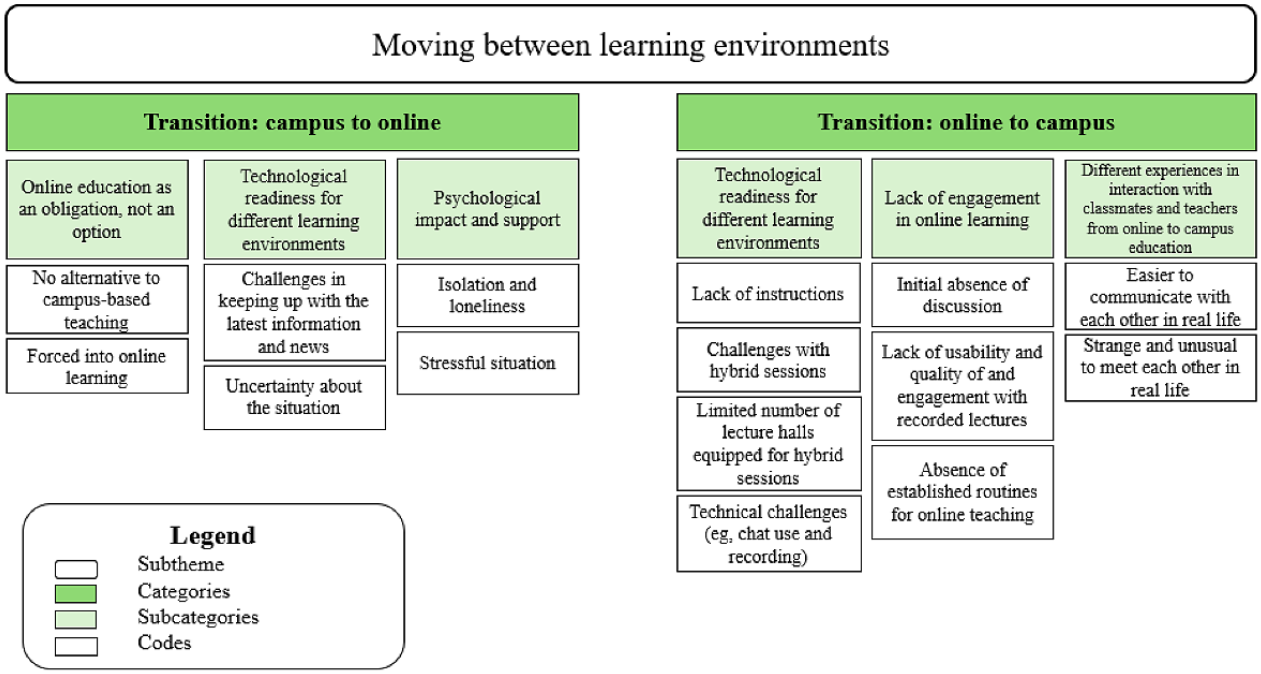
XSL•FO
RenderX

> *...Well, I think the plus point was that my husband works in technology, so I could take his help to understand the concepts and I think that was the main part that kept me going because if I don't understand things, I just...I can't focus and I can't move forward...* [2019_Participant 2]

Moving Between Learning Environments

Figure 3 illustrates the relationship among the codes, subcategories, and categories for the subtheme *moving between learning environments*.

Figure 3. A hierarchical structure displaying the subtheme and codes for theme 2—moving between learning environments.



Transition: Campus to Online Education

The COVID-19 restrictions prompted universities to shift from on-campus to online teaching. The disease spread rapidly, rendering on-campus lectures unfeasible. Numerous international students faced *challenges in keeping up with the latest information and news*. Despite the continuous updates available on the university's website, many international students experienced *uncertainty about the situation* and when everything would go back to normal:

> *...and I think that the transition between these different ways of working, I think that was hard for all of us because, well, especially for me, since I was studying before. So, I got used to online learning and then I went back to traditional learning and was very excited about that. But then it went back to online learning again and I was like, huh, OK...* [2021_Participant 2]

Working *online was an obligation rather than an option* for students, which may have made it easier to accept and manage, at least in the early stages of the pandemic. Having *no alternative to campus-based teaching* and *being forced into it* made online learning an obligation rather than a choice:

> *Well, I mean, I think there was no other choice. Everything was shut down, and there was no question of going back to campus. Nobody knew when things would reopen again, so in the beginning, we were forced into it, left with no other option. But then, I think the transition was OK.* [2019_Participant 1]

However, the psychological impact and requirement of support were important when considering the transition from on-campus to online environments. *Isolation and loneliness* and *a stressful situation* were common themes among students who did not have significant social support, such as family members living in Sweden. Most of the international students leaving their home countries to study the program in Sweden experienced that they were isolated and felt lonely during the COVID-19 pandemic:

> *...Then we changed our routines, but always we had some fear because of the unknown and what would happen next. It was always stressful, and we didn't meet our friends and teachers because from my side I don't have any relatives here, just only people from the university then I felt isolation and loneliness, and something, something alwayswas in the back of my mind and a little difficult to focus on my studies...* [2019_Participant 3]

Transition: Online to Campus Education

Technological readiness for different learning environments was a subcategory of transitioning from on-campus to an online environment and on the return to on-campus learning. At the onset of the pandemic, some students encountered challenges with Zoom due to a *lack of instructions*, particularly affecting those without a technical background. Numerous students encountered issues with audio quality during lectures and faced

XSL•FO
RenderX

challenges with hybrid sessions. The university was not fully prepared for online or hybrid formats due to the *limited number of lecture halls equipped for hybrid sessions*:

> *...I was a nurse, so I wasn't used to using it [Zoom]...No, it was difficult in the beginning to get used to using Zoom and the online campus. Nobody was explaining to you how to use it in person, so you have to do tutorials online for everything you do...* [2020_Participant 4]

> *We would have that many problems in the classroom technology-wise, like not being able to record the meetings, for instance. That was a big loss. Not being able to join functionally from home because it was not great for you. You didn't actually know what was happening in the classroom, so it was. I just wasn't expecting that. The classrooms weren't equipped to handle hybrid learning. If that was presented as an option, so.* [2020_Participant 8]

Some participants noted the *initial absence of discussion* on enhancing online learning at the onset of the COVID-19 pandemic. Nevertheless, they observed an increase in discussions on how online learning worked and how it could be improved as the pandemic progressed. In addition, at the beginning of the pandemic, technical challenges (eg, Zoom difficulties, uncertainties about chat use, and initial problems with recordings) were encountered. As the pandemic continued, these issues were addressed, leading to a gradual improvement in the online learning experience over time:

> *So, I think by the time we started [the online lectures] everything got improved especially when they started to record the lectures. Because I remember at the beginning the lecture was not recorded. But later lectures were recorded, and this was good.* [2019_Participant 2]

> *I mean for the teachers also, in the beginning, it was difficult...now if you see Zoom meetings are usually facilitated by one person who keeps an eye on the chat while the teacher is teaching. And then in between they ask questions. I mean it took a few months...and during the third semester, it became better towards the end of it. So, I think yes, I mean everybody adapted to it because that was the only way left.* [2019_Participant 1]

Other *technical challenges* included the requirement for extra tools such as headsets and screens, issues with hybrid teaching, and inadequate online content delivery, marked by a *lack of usability and quality of and engagement with recorded lectures* and an *absence of established routines for online teaching*:

> *...but that isn't very equitable for those who might not have had headphones or something like that, or maybe lived in a very small apartment, even though I live in a small apartment. Well, then maybe have a family and other people at home, maybe they didn't have any headphones, so maybe that's an aspect of the required equipment for the most optimal way of learning remotely...* [2021_Participant 2]

> *...if you maybe go and look at a lecture or recorded video on YouTube. They are very different, and they are a lot more engaged. Even though they're not live, so I do think that there should be some sort of, yeah, look over, like how you're supposed to have an online lecture to maximize learning. Because, yeah, I've, I found it quite strange that the prerecorded ones were better than the live ones.* [2021_Participant 4]

Some students encountered challenges with active participation in online and hybrid sessions, which they experienced as a lack of engagement in online learning from learners and, at times, educators. They emphasized the need for implementing strategies to enhance engagement during online sessions:

> *It lies on the lecturers. I mean, they're supposed to engage the students, and I understand from their perspective that it's so hard to engage people on a computer, because the students usually don't have their cameras on, usually, maybe even sit in their beds, listening to the lecture. And there are some lecturers who really, really try to be like: please turn on your cameras...I think that it requires maybe some standard routine for how we're supposed to do remote learning, and I mean, just like any technical product that goes into implementation, we need to have change management afterward. We need to maybe have some implementation consultants. So, both students and teachers or professors need to learn how they are supposed to teach, or for students how we are supposed to learn.* [2021_Participant 2]

Students had *different experiences in interaction with classmates and teachers between online and campus education.* Several of them noted that *it was easier to communicate with each other* during the transition from online to on-campus learning given their previous digital interaction through lectures and the WhatsApp group, although some students found it *strange and unusual to meet their classmates* in person for the first time. However, the students experienced limited interaction with teachers within the program. They found it more challenging to engage with teachers compared to their classmates:

> *I think for me I can kind of remember that. Since we first started digitally and then we changed to the campus, we knew each other's faces and how we interacted. So, in a way, it was kind of easier to change...After you've trained a little bit in the digital parts, then it gives you more confidence to talk in person. On the other side maybe, I was more shy to talk to the teachers...it was harder. I felt more the distance in a way, so it's kind of the different sides, but maybe I felt nearer to my colleagues...* [2020_Participant 2]

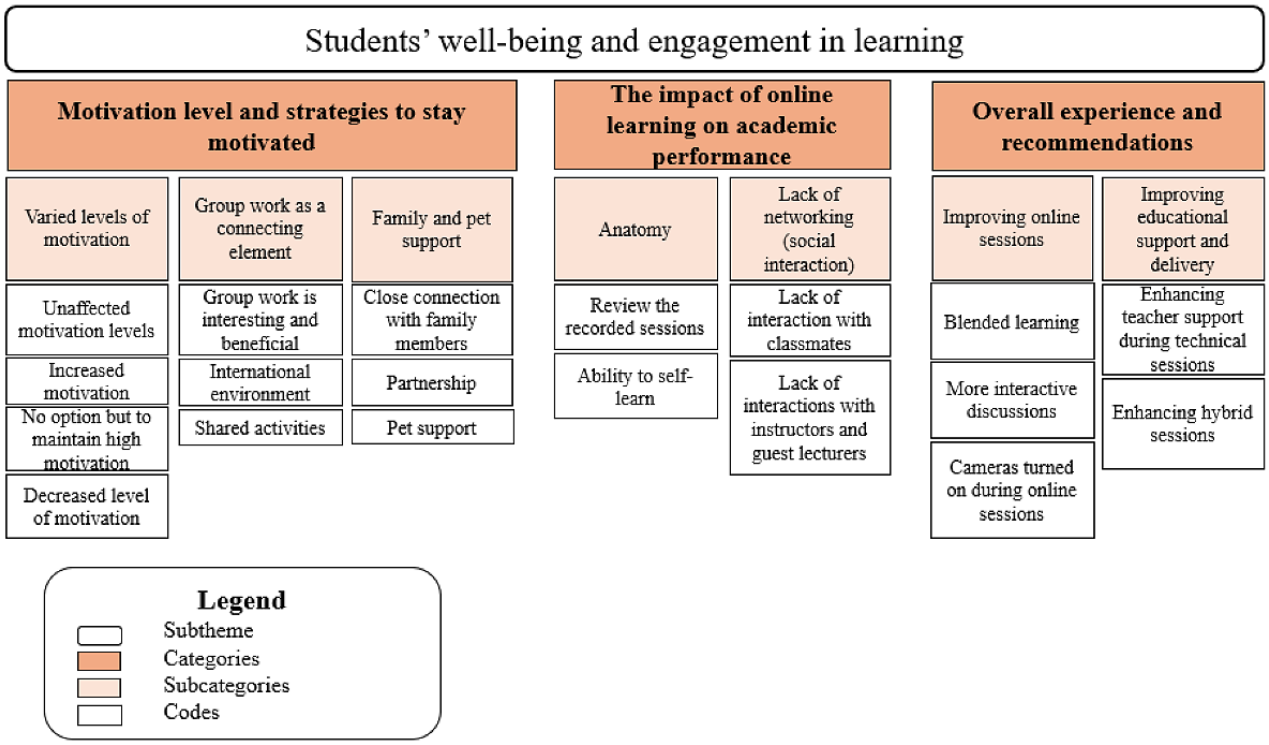
> *I would say I mean. The first day when we met. It was, it was very weird to meet our colleagues. I mean, I'm not used to seeing them in real life. Some were shorter, some were taller, and you know it was a weird experience for the first day, I would say...* [2020_Participant 7]

XSL•FO
RenderX

Students' Well-Being and Engagement in Learning

Figure 4 illustrates the relationship among the codes, subcategories, and categories for the subtheme *students' well-being and engagement in learning.*

Figure 4. A hierarchical structure displaying the subtheme and codes for theme 3—students' well-being and engagement in learning.



Motivation Level and Strategies to Stay Motivated

The *motivationlevels* of students varied depending on their individual circumstances and also throughout the pandemic. Some students noted that their *motivation levels remained unaffected*, maintaining the same commitment to learning despite the pandemic. However, they expressed an *increased motivation* to meet people in person. Several students found that their motivation improved during the pandemic as they had more free time to plan upcoming courses, learn new skills, and engage in activities beyond their studies. Other students expressed that they had *no option but to maintain high motivation* due to visa requirements, family responsibilities, and future career plans. They highlighted that their relocation to Sweden was driven by the pursuit of a better future:

> *I would definitely say that when we moved to the campus, I felt more motivated not because of the learning, maybe because the motivation might be exactly the same, but also more, more motivated to meet people, to actually have an opportunity to meet different people because some of you have lived in Sweden and it's hard most of the time because of the difference in the cultural aspects. So, I think that's something that I remember being nice that I was happy to go to the campus and see these people and talk and learn with them.* [2020_Participant 2]

> *I think we as international students are here because you have your student visa, you have to study, and you have to pass to have your visa renewed. So, we don't have this luxury...you know, I'm bored...I don't want to study; I will skip it. No. We have to do this...We came here to study this program, and you have your plans, and you need to work to achieve your plans.* [2019_Participant 2]

> *...It's the same motivation because I brought my kids [to Sweden] and I have to make the future. That is the motivation.* [2019_Participant 3]

Nevertheless, some students reported a *decreased level of motivation*. Some related this loss of motivation to their backgrounds, particularly feeling uncomfortable with certain courses such as programming and machine learning. Others mentioned that the darkness negatively affected their mood, although it did not impact their motivation to study:

> *If I talk about the motivation, yes it was reduced because I come from a medical background, and the technical things, especially the courses like programming and machine learning...There were so many new things that sometimes I felt I was lost...Within different things, all alone, and I don't know if I'm the only one feeling like this or [if] there's somebody else, or I mean, what else do I need to make it work better? But then I didn't want to stop doing this...Because my main motivation was to learn technology and I wanted to complete it so that I know...* [2019_Participant 1]

> *I think the majority of the time that we had online lectures was during the darkest period in Sweden. And I think that affected the mood also because you then sat in front of your computer and it was dark, and then you sat inside the whole day pretty much.*

XSL•FO
RenderX

> *And then it was dark again. And so obviously that has an impact too, that it just happened to be during the most like the darkest period.* [2021_Participant 4]

Some students preferred online lectures due to the time and energy saved from morning routines such as waking up early, having breakfast, and preparing lunch. However, other students believed that lectures and interactions in person gave them more energy:

> *I really like remote learning and additionally, I'd like to add as I was saying in the beginning, I was able to be in another country because at that time my parents were living in another country and then me and [participant] actually could collaborate because we were at the same time zone.* [2021_Participant 2]

> *I will actually agree here with [participant] that also the lectures in person and interactions in person actually gave me more energy than they took, due to the fact that I was way more active. For example, I cycled to university. Then uh each morning and after class. Which always gives me an energy boost when I move more than also like just walking together on the campus, going for lunch, or getting a coffee or whatsoever. That also gave me I would stay way more energy than it took from me compared to when I would just be in my apartment. Pick up a coffee and then sit back again at the table...* [2020_Participant 1]

Group work was identified as a key factor for gaining and maintaining motivation during the pandemic. Many students believed that *engaging in group work was interesting and beneficial* throughout the program. The *international environment* played a crucial role in keeping students motivated. In addition, maintaining positive relationships with classmates and enjoying *shared activities* helped students stay motivated during the COVID-19 pandemic:

> *I would say that the group work made me more motivated. If I felt down. Honestly, we had very nice groups that we worked with. I enjoyed working with them and I learned a lot from them. But yeah, when you feel down and then we get a challenge to work on some project or some research or yeah, it was studying having some fun, and some jokes here and there. So, it wasn't 100% serious mode. So yeah, group work affected...* [2020_Participant 7]

> *If it was just entirely remote, and I never had the opportunity to meet my classmates one-on-one, I don't think I would be this motivated. Meeting them for the first time, seeing their various background, a lot of experience, and being able to tap into their own personal experiences and professional experiences gave me at least a lot of motivation, and a lot of interest in how better I can be...Or if it was just purely digital, or that involved me not traveling, I don't think I would have been this motivated.* [2021_Participant 3]

Several students noted that having close connections with their *family, a partner, or a pet* played a significant role in keeping them motivated during the pandemic:

> *Yeah, I agree. Actually, that happened to me too with my husband. He was more of a body shadowing. Like, if someone else is working on your side like you, I don't know. You get more folks. Somehow with cats, it's different though. It may not be a routine or like force it to go out. To walk the cat...So, it would actually force me to get up, play a little bit with the cat, and then come back and it was really good.* [2020_Participant 8]

The Impact of Online Learning on Academic Performance

While some students reported no change in their academic performance, others observed an improvement in their overall achievement, among other reasons due to having the opportunity to *review the recorded sessions*. The students expressed a sense of adaptability to online learning, emphasizing their *ability to self-learn*. However, the students found it challenging to grasp technical aspects solely through online learning:

> *For me, I think that the pandemic situation increased my performance. Because lectures were recorded and then I used them again and again until I understood them. And then especially for the technical part...because we are not familiar with technology, then I used again and again for all those recorded things to understand...that was very effective for me.* [2019_Participant 3]

Students experienced a *lack of interaction with classmates, instructors, and guest lecturers*, hindering the exchange of ideas and discussion of questions. This lack of networking could potentially lead to delayed job opportunities:

> *...But yes, I agree with my colleagues that some from the social side there was a lack. Having a bigger network of friends or socializing with your friends and visiting companies in person, those who were giving us lectures like different healthcare companies were giving us guest lectures, but we couldn't go in person to the companies to visit them...Because it was a pandemic, there were no, not many opportunities for summer jobs and internships, but not much open during that time. So, it was kind of that gap, which was present during [the pandemic]...* [2020_Participant 6]

Overall Experience and Recommendations

The students noted that the program altered their approach to problem-solving. However, they expressed a preference for *blended learning* over online learning alone as the latter had a negative impact on some, leading to feelings of isolation. In addition, the students provided suggestions for improvement in both online and campus-based sessions. They desired *more interactive discussions*, especially for technical sessions, to strengthen their knowledge after online lectures. Students also

XSL•FO
RenderX

believed that having *cameras turned on during online sessions* would enhance the learning experience:

> *...So overall, it was a positive experience, in my opinion. I would say I learned a new way of learning I wasn't used to, but sort of adaptability I would say. In terms of challenges, I would say I would agree with my colleagues about the social part. We were not able to socialize as compared to the normal or the offline study. It was very important. As well as building relationships with the instructors, teachers, and companies. And we were not able to do that. One other disadvantage was the interactive discussion with colleagues...* [2020_Participant 7]

> *...I can mention that not having the cameras on for example is another aspect that didn't help in concentration, also being able to socialize with people because this is a very international program, and would be amazing just to be able to have been in the classroom from the beginning to the end...* [2020_Participant 2]

> *We got to see our peers in person and kind of create a connection and sit and eat lunch with each other and talk about things to get to know each other more, which actually can not only motivate us but also affect us in our ways of learning because we can get another perspective on a certain topic or something that we wouldn't have gotten without talking to them during an informal event such as lunch, for instance...Again, I think a mix, a mix of online and on-site learning has made me at least reach my performance goals.* [2021_Participant 2]

The importance of having *teacher support* during the practical sessions in technical courses in the master's program during the pandemic highlights a key factor in improving educational support in online learning:

> *Also, we have some practical sessions for our program, so we must have someone with us in the room to ask because there are technical things in programming. I don't know where the error is in what I am doing, so I need a next eye to see what I'm doing.* [2019_Participant 2]

Enhancing hybrid sessions was mentioned as an area for improvement as students encountered numerous challenges during these sessions throughout the pandemic:

> *Yeah, I missed something. I just remembered that when we switched to in-person, there were some meetings hybrids. So people joined online while we were in the class and this was not really well organized or I don't know if it was our problem or if it could be better, but I think it's something that can be improved or just not having hybrid meetings, but I mean, I think we're not prepared and we just had like 1 computer with the meeting, so only the teacher could hear the student who was at home.* [2020_Participant 5]

Discussion

Principal Findings

This paper reports the experience of master's program students at a higher education institution due to the COVID-19 pandemic requiring adaptation to hybrid learning. Students experienced both benefits and challenges in relation to this. The transition between on-campus and online learning resulted in a hybrid learning era, which is likely here to stay. The students in this study appreciated the flexibility and autonomy provided by online learning, enabling them to integrate their studies into their daily lives. However, the rapid shift to online learning caused significant challenges, such as changes in social interactions, technical difficulties, and feelings of obligation rather than choice in adopting e-learning. The transition from on-campus to online learning caused a psychological impact on students and highlighted the need for better support and technological readiness to adapt to different learning environments. Returning to campus presented mixed experiences, with some students struggling to re-engage in face-to-face learning, whereas others faced challenges in adjusting to renewed forms of social and academic interactions. Students' motivation varied during the pandemic. Group work, family support, and having pets played crucial roles in maintaining students' moods. Despite the challenges, students were positive about hybrid learning to enhance future experiences, emphasizing the need for better networking opportunities and innovative strategies.

The findings of this study highlight implications for educators, students, and higher education institutions to embrace adaptation and foster innovation. These implications include the need for educators to stay current with the latest educational technologies and design teaching strategies and pedagogical approaches suited to both online and in-person settings to effectively foster student engagement. Students must be informed about the technological requirements for online learning and adequately prepared to meet them. Institutions play a critical role in ensuring equitable access to technology, guiding and supporting educators in adopting innovative tools and methods, and offering mental health resources to assist students in overcoming the challenges of evolving educational environments.

Comparison to the Literature

Studies have previously reported both positive and negative consequences of the shift in learning environments. Naciri et al [10] have suggested that students generally responded positively to the rapid shift to online health science education during this crisis, expressing views on aspects such as acceptance, motivation, and engagement. Despite varying socioeconomic conditions across countries, certain key factors such as access to technology, basic computer literacy, well-designed online course pedagogy, and flexibility in learning consistently supported online education. However, students encountered challenges such as inconsistent internet access, difficulties with educational platforms, and hurdles in acquiring clinical skills online. These insights are crucial for enhancing the integration of these technologies into educational frameworks [10]. This is congruent with our results, where

XSL•FO
RenderX

overall the students responded well to the rapid shift to online learning but reported both positive and negative outcomes, with challenges especially in computer laboratory sessions. Our findings also confirm the results of other studies [9,11] on the difficulties regarding limited internet access, technical problems, challenges related to adjustments from traditional to online formats, and impact on students' well-being.

Students in our study were not entirely satisfied with the practical sessions in technical courses, highlighting the need for access to teacher assistance during these sessions. This contrasts with a study that reported satisfaction with clinical teaching and practical sessions remaining adequately high during the shift to e-learning. They also reported notable improvement in student satisfaction concerning course structure, instructor expertise, learning materials, and overall contentment with the courses, as well as a tendency for student grades to improve in the online format [8].

A review study [20] revealed that motivation and self-regulated learning were significant challenges, impacting students' ability to engage critically with the material. It also showed varied attitudes toward online learning, with decreased satisfaction and emotional well-being in many students due to feelings of isolation and increased stress. These findings align with our results as students in this study also experienced loneliness and isolation, which led to a lack of focus on the studies.

Findings regarding the use of technology show that competence, perceived usefulness, ease of use, and facilitated implementation are predictors of learners' attitudes toward and intentions to use technology [18,21,22]. Technologically capable students are likely to associate poor digital implementation by educators with lower satisfaction and self-efficacy. Integration of technology with a student-centered focus is likely to influence the development of autonomy and self-regulated learning [23,24]. For students enrolled in a health informatics program, where technology is a significant aspect of the learning curriculum, it could be extrapolated that they would likely be comfortable and be able to adapt to the use of technologies for learning. Therefore, the reported themes associated with technology use may be more significant for other groups of learners in different educational fields.

In addition to these findings, there has been a significant association reported between instructors' use of effective teaching practices and student motivation. Alongside the quality of teaching and challenges associated with technology, prepandemic studies have indicated that motivation in online learning can be affected by demographic characteristics such as age, gender, employment status, income, and family obligations [22]. However, most studies were conducted at higher education institutions where students had a choice to enroll in online learning; learning in a hybrid environment may be different.

Online laboratory sessions in technical courses have shown some drawbacks in managing requests for help from student groups. During in-person sessions, teaching assistants can manage the requests of a group but, at the same time, give feedback to others (spontaneous interactions or questions that require a short answer). Using online learning tools with students divided into rooms, requests are managed only sequentially without allowing for spontaneous interactions. In addition, managing the request queue was more challenging for teaching assistants due to the unavailability of a specific function in the online learning tools. This seems to be in accordance with previous research on the topic [25-27].

Implications for the Educators, Students, and Higher Education Institutions

The transition to online learning during the COVID-19 pandemic had significant implications for educators, students, and universities. Both educators and students had to make substantial adjustments in their teaching methods and learning styles, respectively. This study highlighted a greater emphasis on active learning strategies, self-discipline and time management, active participation in discussions and group work, and the development of flexibility and resilience.

Although our study focused on the experiences of students, the importance of educators in both the design and delivery of hybrid learning must be recognized. An online replication of a physical classroom can result in bored students and exhausted teachers, with both experiencing *Zoom fatigue*. A change in pedagogy is required to teach in the hybrid learning environment. Hybrid pedagogy is a development from blended learning using elements of both online and face-to-face learning, resulting in no separation between learners in the physical or digital space [13].

A significant issue related to hybrid learning is the required use of technologies. The health informatics students recognized both individual and organizational gaps in infrastructure and a lack of technological preparedness, which was experienced by many institutions. After the pandemic, it is critical that higher education institutions ensure that technology is leveraged to provide students with a good online, physical, or hybrid experience. The pandemic has highlighted the importance of preparedness and resilience in higher education for future crises. To support online and hybrid education, institutions need to invest in robust technological infrastructure and prioritize continuous professional development and training for educators. As highlighted in our study and similar research [9,10,28,29], there is a significant need to equip educators with the skills to adapt to the evolving educational landscape, particularly in the use of digital technologies. Educators need to develop skills to effectively use online platforms and tools to create engaging and interactive learning experiences. Common themes when designing quality online instruction that engages and motivates students are those of interaction, collaboration, communication, and discussion [21,30]. Online and hybrid learning require a greater focus on interactivity; this helps break the monotony and allows for student socialization. Therefore, engaging in ongoing professional development to stay updated with the latest educational technologies is a necessity for educators. Those students who had previously had the opportunity to meet in person on campus before transitioning to online learning may have had a closer sense of community than those whose educational experience transitioned in the opposite direction. Therefore, it is vital for instructors to build and foster a sense of community to keep students motivated. Overall, the onus is

XSL•FO
RenderX

on educators to design and deliver quality teaching appropriate to the environment (web based or in person) to promote student engagement and encourage motivation to learn. To support these efforts, it is equally important that students are informed in advance of the requirements to attend classes online and be advised on the necessary technology (such as headphones, microphone, and camera). Furthermore, institutions should ensure access to technology by providing it on a short-term loan basis or guiding students in applying for grant funding, thereby ensuring equitable access to learning resources. Institutions should also guide and support educators to learn about and incorporate new technologies into their teaching practices [31-33]. In addition, providing extra resources and support to help students navigate the challenges of online learning is crucial. Given the impact that the shift had on students' mental health, higher education institutions should provide mental health resources and support systems to assist students in managing these challenges. Students, educators, and institutions play a crucial role in informing policy makers regarding challenges and implications of future crises by sharing best practices for managing education emergencies. Providing feedback, sharing research findings, and offering concrete examples are essential for ensuring that policy makers are well informed and able to respond to the evolving needs of the educational community.

Strengths and Limitations of This Study

This study is limited by its context-specific nature, focusing on a particular group of students within a specific educational environment (health informatics master's students at KI), which may not fully capture the diverse experiences of learners in other regions, disciplines, or institutional settings. This may limit the generalizability of the results. However, the recommendations in the results can apply to other programs and institutions with similar educational systems and resources. The sample size in this study could be considered a limitation. However, saturation was achieved across the entire sample as no additional new themes emerged during the data analysis process. Although the number of participants from the 2019 cohort was low, the insights gathered from this group were consistent with those gathered from students from the other cohorts.

Conclusions

The shift to hybrid learning in response to the COVID-19 pandemic presented both benefits and challenges for postgraduate students in higher education institutions. Technological preparedness, equitable access to technology, and educator training are crucial factors that institutions must address to support students' learning experiences. In addition, fostering interaction, collaboration, communication, and discussion in online and hybrid learning environments is essential for engaging and motivating students. Ultimately, educators play a key role in designing and delivering quality teaching that promotes student engagement and encourages motivation to learn regardless of the learning environment.

Future studies should aim to explore the impact of the COVID-19 pandemic on students' learning experiences across varied contexts, incorporating cross-institutional comparisons to develop a more comprehensive understanding of how different factors influence learning experiences and outcomes in hybrid and online education. In addition, it is of great importance that future studies examine new teaching methods and pedagogical approaches that can enhance students' engagement and increase communication and interaction between students in hybrid education. Finally, it is crucial to develop strategies and policies that prepare higher education institutions for crises, ensuring that they can effectively transition between educational environments—whether shifting from in-person to online or hybrid modes—while maintaining educational continuity and quality. Future research could use case studies to investigate institutional responses and apply principles of action research to collaboratively develop and refine teaching strategies and crisis management policies.

Acknowledgments

The authors would like to thank all master's students and alumni who participated in this study for their time and valuable insights. The authors received no specific funding for this work. Artificial intelligence technology such as Grammarly (Grammarly Inc) and ChatGPT-3.5 (OpenAI) was used in the preparation of this manuscript to harmonize the text. After using these tools, the authors reviewed and edited the content as needed, and they take full responsibility for the content of the publication.

Authors' Contributions

ND, SB, and NS were involved in the design of the study. These authors prepared the interview guide and piloted it. The focus groups were held online with CS in the role of interviewer. ND, NS, and SB conducted the initial coding. CS joined the analytical process once the initial coding and subcategories had been generated. After identifying subcategories, a comparative analysis was conducted to highlight similarities and differences among the student groups. We then collaboratively categorized and identified subthemes for all 3 groups. All authors contributed to the subsequent writing and review of the manuscript.

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. Li D. The shift to online classes during the COVID-19 pandemic: benefits, challenges, and required improvements from the students' perspective. Electron J E Learn 2022 Jan 21;20(1):pp1-p18. [doi: 10.34190/ejel.20.1.2106]
2. Ruiz JG, Mintzer MJ, Leipzig RM. The impact of E-learning in medical education. Acad Med 2006 Mar;81(3):207-212. [doi: 10.1097/00001888-200603000-00002] [Medline: 16501260]
3. Barrot JS, Llenares II, Del Rosario LS. Students' online learning challenges during the pandemic and how they cope with them: the case of the Philippines. Educ Inf Technol (Dordr) 2021 May 28;26(6):7321-7338 [FREE Full text] [doi: 10.1007/s10639-021-10589-x] [Medline: 34075300]
4. Lederman D. Most teaching is going remote. Will that help or hurt online learning? Inside Higher Ed. URL: https://www.insidehighered.com/digital-learning/article/2020/03/18/most-teaching-going-remote-will-help-or-hurt-online-learning [accessed 2024-04-29]
5. Lin Y, Nguyen H. International students' perspectives on e-learning during COVID-19 in higher education in Australia: a study of an Asian student. Electron J E Learn 2021 Aug 11;19(4):241-251. [doi: 10.34190/ejel.19.4.2349]
6. Wu Z. How a top Chinese university is responding to coronavirus. World Economic Forum. URL: https://www.weforum.org/agenda/2020/03/coronavirus-china-the-challenges-of-online-learning-for-universities/ [accessed 2024-04-29]
7. Worth D. Coronavirus: how to maximise distance learning. Tes Global Ltd. URL: https://www.tes.com/magazine/article/coronavirus-how-maximise-distance-learning [accessed 2024-04-29]
8. Alenezi S, Bahathig A, Soliman M, Alhassoun H, Alkadi N, Albarrak M, et al. Performance and satisfaction during the E-learning transition in the COVID-19 pandemic among psychiatry course medical students. Heliyon 2023 Jun;9(6):e16844 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e16844] [Medline: 37303529]
9. Alsoufi A, Alsuyihili A, Msherghi A, Elhadi A, Atiyah H, Ashini A, et al. Impact of the COVID-19 pandemic on medical education: medical students' knowledge, attitudes, and practices regarding electronic learning. PLoS One 2020;15(11):e0242905 [FREE Full text] [doi: 10.1371/journal.pone.0242905] [Medline: 33237962]
10. Naciri A, Radid M, Kharbach A, Chemsi G. E-learning in health professions education during the COVID-19 pandemic: a systematic review. J Educ Eval Health Prof 2021 Oct 29;18:27 [FREE Full text] [doi: 10.3352/jeehp.2021.18.27] [Medline: 34710319]
11. Kumar A, Sarkar M, Davis E, Morphet J, Maloney S, Ilic D, et al. Impact of the COVID-19 pandemic on teaching and learning in health professional education: a mixed methods study protocol. BMC Med Educ 2021 Aug 19;21(1):439 [FREE Full text] [doi: 10.1186/s12909-021-02871-w] [Medline: 34412603]
12. Nuryana Z, Xu W, Kurniawan L, Sutanti N, Makruf SA, Nurcahyati I. Student stress and mental health during online learning: potential for post-COVID-19 school curriculum development. Compr Psychoneuroendocrinol 2023 May;14:100184 [FREE Full text] [doi: 10.1016/j.cpnec.2023.100184] [Medline: 37038597]
13. Haikalis M, Doucette H, Meisel MK, Birch K, Barnett NP. Changes in college student anxiety and depression from pre- to during-COVID-19: perceived stress, academic challenges, loneliness, and positive perceptions. Emerg Adulthood 2022 Apr 27;10(2):534-545 [FREE Full text] [doi: 10.1177/21676968211058516] [Medline: 35382515]
14. Lederman D. The shift to remote learning: the human element. Inside Higher Ed. URL: https://www.insidehighered.com/digital-learning/article/2020/03/25/how-shift-remote-learning-might-affect-students-instructors-and [accessed 2024-04-29]
15. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
16. Collins III JW, O'Brien NP. The Greenwood Dictionary of Education. Santa Barbara, CA: Greenwood Press; 2003.
17. Gundumogula M. Importance of focus groups in qualitative research. Int J Humanit Soc Sci 2020 Nov 30;8(11):299-302. [doi: 10.24940/theijhss/2020/v8/i11/hs2011-082]
18. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
19. Saladana J. The Coding Manual for Qualitative Researchers. 2nd edition. Thousand Oaks, CA: Sage Publications; 2013.
20. Salas-Pilco SZ, Yang Y, Zhang Z. Student engagement in online learning in Latin American higher education during the COVID-19 pandemic: a systematic review. Br J Educ Technol 2022 May 15;53(3):593-619 [FREE Full text] [doi: 10.1111/bjet.13190] [Medline: 35600418]
21. Munday D. Hybrid pedagogy and learning design influences in a higher education context. Stud Technol Enhanc Learn 2022;2(2):25 [FREE Full text] [doi: 10.21428/8c225f6e.b5af8bae]
22. Hanham J, Lee CB, Teo T. The influence of technology acceptance, academic self-efficacy, and gender on academic achievement through online tutoring. Comput Educ 2021 Oct;172:104252. [doi: 10.1016/j.compedu.2021.104252]
23. Franchi T. The impact of the COVID-19 pandemic on current anatomy education and future careers: a student's perspective. Anat Sci Educ 2020 May 05;13(3):312-315 [FREE Full text] [doi: 10.1002/ase.1966] [Medline: 32301588]
24. Mukhtar K, Javed K, Arooj M, Sethi A. Advantages, limitations and recommendations for online learning during COVID-19 pandemic era. Pak J Med Sci 2020 May 18;36(COVID19-S4):S27-S31. [doi: 10.12669/pjms.36.covid19-s4.2785]
25. McKnight K, O'Malley K, Ruzic R, Horsley MK, Franey JJ, Bassett K. Teaching in a digital age: how educators use technology to improve student learning. J Res Technol Educ 2016 May 21;48(3):194-211. [doi: 10.1080/15391523.2016.1175856]

XSL•FO
RenderX

26. Austen L, Parkin HJ, Jones-Devitt S, McDonald K, Irwin B. Digital capability and teaching excellence: an integrative review exploring what infrastructure and strategies are necessary to support effective use of technology enabled learning (TEL). Sheffield Hallam University Research. URL: https://shura.shu.ac.uk/13750/1/Digital-capability-and-teaching-excellence-2016.pdf [accessed 2024-04-29]
27. May D, Morkos B, Jackson A, Hunsu NJ, Ingalls A, Beyette F. Rapid transition of traditionally hands-on labs to online instruction in engineering courses. Eur J Eng Educ 2022 Mar 05;48(5):842-860. [doi: 10.1080/03043797.2022.2046707]
28. Rashid S, Yadav SS. Impact of COVID-19 pandemic on higher education and research. Indian J Hum Dev 2020 Aug 23;14(2):340-343. [doi: 10.1177/0973703020946700]
29. Jakoet-Salie A, Ramalobe K. The digitalization of learning and teaching practices in higher education institutions during the Covid-19 pandemic. Teach Public Adm 2022 Apr 21;41(1):59-71. [doi: 10.1177/01447394221092275]
30. Teodorescu D, Aivaz KA, Amalfi A. Factors affecting motivation in online courses during the COVID-19 pandemic: the experiences of students at a Romanian public university. Eur J High Educ 2021 Aug 30;12(3):332-349. [doi: 10.1080/21568235.2021.1972024]
31. Turnbull D, Chugh R, Luck J. Transitioning to e-learning during the COVID-19 pandemic: how have higher education institutions responded to the challenge? Educ Inf Technol (Dordr) 2021 Jun 23;26(5):6401-6419 [FREE Full text] [doi: 10.1007/s10639-021-10633-w] [Medline: 34177349]
32. El Said GR. How did the COVID-19 pandemic affect higher education learning experience? An empirical investigation of learners' academic performance at a university in a developing country. Adv Hum Comput Interact 2021 Feb 8;2021:1-10. [doi: 10.1155/2021/6649524]
33. Oliveira G, Grenha Teixeira J, Torres A, Morais C. An exploratory study on the emergency remote education experience of higher education students and teachers during the COVID-19 pandemic. Br J Educ Technol 2021 Jul 18;52(4):1357-1376 [FREE Full text] [doi: 10.1111/bjet.13112] [Medline: 34219758]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
KI: Karolinska Institutet

Edited by B Lesselroth; submitted 27.06.24; peer-reviewed by H Abuhassna, H Gray; comments to author 10.11.24; revised version received 13.12.24; accepted 02.01.25; published 10.02.25.

Please cite as:
Davoody N, Stathakarou N, Swain C, Bonacina S
Exploring the Impact of the COVID-19 Pandemic on Learning Experience, Mental Health, Adaptability, and Resilience Among Health Informatics Master's Students: Focus Group Study
JMIR Med Educ 2025;11:e63708
URL: https://mededu.jmir.org/2025/1/e63708
doi:10.2196/63708
PMID:

©Nadia Davoody, Natalia Stathakarou, Cara Swain, Stefano Bonacina. Originally published in JMIR Medical Education (https://mededu.jmir.org), 10.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Dentists: A Curriculum for the 21st Century

Michelle Mun[1,2], DDS; Samantha Byrne[1], PhD; Louise Shaw[2], PhD; Kayley Lyons[2], PhD

[1]Melbourne Dental School, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, 720 Swanston Street, Melbourne, Australia

[2]Centre for Digital Transformation of Health, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Melbourne, Australia

Corresponding Author:
Michelle Mun, DDS
Melbourne Dental School, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, 720 Swanston Street, Melbourne, Australia

Abstract

Future health professionals, including dentists, must critically engage with digital health technologies to enhance patient care. While digital health is increasingly being integrated into the curricula of health professions, its interpretation varies widely depending on the discipline, health care setting, and local factors. This viewpoint proposes a structured set of domains to guide the designing of a digital health curriculum tailored to the unique needs of dentistry in Australia. The paper aims to share a premise for curriculum development that aligns with the current evidence and the national digital health strategy, serving as a foundation for further discussion and implementation in dental programs.

(JMIR Med Educ 2025;11:e54153) doi:10.2196/54153

KEYWORDS

digital health; digital transformation; informatics; ehealth; dentistry; dental informatics; curriculum; competence; capability; dental education

Introduction

As the world continues to be digitally transformed, there are increasing expectations for health care providers to use technology and handle health information safely and ethically [1]. It is likely that future health professionals will also need to think critically about how digital health technologies can be used to transform models of care [2].

Digital health and informatics remain relatively new curriculum topics for many health professions, including dentistry. Defining the relevant curricular objectives in entry-to-practice degrees can be particularly challenging for several reasons. First, there are several definitions and conceptualizations of the term “digital health” [1,3-6]. Second, the implementation of digital health education in health profession degrees has largely been ad hoc, with different schools adopting varied approaches [7]. This has resulted in inconsistent learning outcomes and a fragmented understanding of digital health competencies for health profession graduates. Third, although a multitude of digital health competency frameworks exist [8-11], there is a notable absence of shared curriculum models specific to dentistry. Therefore, dentistry educators are not aware of or struggle to adopt best practices in the teaching of digital health.

In this viewpoint, we argue that there is a need to integrate digital health education into dentistry curricula to prepare future practitioners for the increasingly digitized health care environment. Specifically, we propose a distinct point of view for defining "digital health" in dental education, a structured set of domains to guide the design of digital health curriculum, and a framework for curriculum development that aligns with current evidence.

Beginning With the End in Mind

In Australia, a country that ranks consistently high for digital health maturity [12], clear digital health objectives are set out through national strategies such as the Australian Digital Health Agency (ADHA) Capability Action Plan (2023) and the National Healthcare Interoperability Plan (2023‑2028) [1,13]. The government body for digital health (ie, the ADHA), peak bodies such as the Australasian Institute of Digital Health (AIDH), and digital health innovation centers all identify building workforce capability as a critical part of achieving digital transformation of health care [1,13-15]. These organizations envision a future where health professionals will work in integrated and multidisciplinary environments. Digital health education in entry-to-practice degrees is thus a core element of advancing workforce capability; however, the specific content, including priority areas of knowledge and skills, must be tailored to the unique demands of each discipline and local context. For dentistry in Australia, this means aligning the curriculum with the national digital health strategy while addressing the current maturity level and future needs of the dental profession.

XSL•FO
RenderX

Reframing "Digital Health" for the Next Generation of Australian Dental Practitioners

A lack of standardized digital health education in entry-to-practice degrees in Australia has been recognized for decades [16-18] and has been demonstrated by gaps in workforce competency [19]. Global interest in the digital transformation of health care is accelerating, catalyzed by the COVID-19 pandemic and advances in artificial intelligence (AI). However, not all health professions have advanced equally considering their digital transformation. In Australia, as in many countries, the dental sector remains traditionally siloed from the rest of the health care system and faces fragmentation in its information systems [20]. In the dental sector, the most progress in digital health has been observed in restorative and surgical procedures, where technology is directly integrated into clinical workflows [21]. For example, conventional manual techniques and laboratory workflows for the design and fabrication of dental restorations have evolved into in-house, fully digitized workflows with the application of intraoral scanners and chairside milling machines [22]. In contrast, dentistry has a relatively nascent data culture [23], with less focus on the broader scope of digital health, which we define in this viewpoint to encompass virtual care, remote monitoring, mobile health (mHealth), wearables, big data analytics, platforms, and "the exchange of data and sharing of relevant information across the health ecosystem creating a continuum of care" [4].

Developing digital health–capable dentists thus involves more than simply teaching the technical aspects of digital tools used in service delivery; it requires a shift towards understanding how digital data can inform clinical decisions, enhance patient care, and contribute to system-wide improvements. This conceptual change is crucial for moving from a service-focused practice to one that leverages digital health as an integral part of modern dental care. The Learning Health Systems (LHS) framework [24] is one example of how to help dental professionals characterize digital health. LHS are health care environments where science, informatics, incentives, and culture align to promote continuous improvement and innovation. In these systems, best practices are embedded in care, patients actively participate, and new knowledge is generated from every care experience [25]. Building on this vision, dental education should emphasize models where digital health is central to both practice and continuous improvement. This approach will foster digital health capability by cultivating a deeper understanding of how and why digital health technologies enable the delivery of high-quality, safe, and sustainable care.

Textbox 1 outlines the questions that can guide the development of a digital health curriculum for entry-to-practice dental education. These questions are intended to help educators and curriculum developers define clear goals aligned with the specific needs of the discipline and the local health care context.

Textbox 1. Defining the goals of digital health curriculum for entry-to-practice degrees.

- What is outlined in national and local digital health strategies for the next 5-10 years? What does the political and funding environment look like?
- What are the digital health-related accreditation standards of the profession?
- What does the current and future digital health maturity of the primary work environments of your graduates look like?
  - Consider the difference in goals for:
    - A rural school where graduates may work in areas with limited digital maturity
    - A health discipline or specialty where graduates will typically work in tertiary care rather than primary care

Considerations for Curriculum Development

The Australian Dental Council (ADC) recently revised its competencies for newly qualified dental practitioners; they updated the requirement to include "using digital technologies and informatics to manage health information and inform person-centred care" [26]. This prompted the authors to develop a digital health curriculum to be implemented in a higher education institution that has graduating dental professionals in Australia. As per the best practice in curriculum development [27], we considered the existing digital health competency and capability frameworks as part of our curricular needs assessment. An environmental literature scan found that only a few frameworks had been created specifically for dentistry or involved dental experts in their consultations, reflecting a lag in dentistry's digital health participation (Multimedia Appendix 1). As a result, not all topics in these existing frameworks were relevant or current to the reality of training dental professionals in Australia, who will predominantly work in small clinics in primary care, in practices with varying digital health maturity [28,29]. An exception was the digital dentistry curriculum proposed by the American College of Prosthodontists [30], which is well-researched but focused solely on digital skills for prosthodontics. This highlighted a gap in resources to support the broader skill set of graduating dentists in Australia, as outlined by the ADC.

The process of designing higher education courses aims to align industry standards with a scaffolded approach for developing effective learning outcomes that produce work-ready graduates. While the ADC's revised competency served as a catalyst for curriculum development, our efforts extended beyond the ADC's scope to meet standards such as those overseen by the Tertiary Education Quality and Standards Agency (TEQSA), which

XSL•FO
RenderX

performs the quality assurance checks for all participants, delivered as part of higher education in Australia. TEQSA's emphasis on authenticity in curricula design, as well as contemporary leading practice [31,32], influenced our approach towards designing a curriculum that not only meets regulatory competencies but also prepares students for practical, professional challenges in the evolving digital health landscape.

Finally, a critical component of our approach was to tailor the curriculum to the local context. While internationally recognized informatics competencies [33] often underpin digital health capability frameworks, they do not alone fully capture the breadth and nuances of digital health proficiency. Digital health encompasses a range of skills, including digitally enabled clinical processes, care pathways, and behavior change management, all of which are shaped by local variations in digital health maturity and sociocultural contexts. Furthermore, curriculum development often occurs under significant time and resource constraints, requiring an approach that is rigorous but targeted. For example, rural schools may not yet prioritize AI competencies if electronic health records are not yet in use locally.

Key Domains

Two frameworks were selected to inform the development of the dental digital health curriculum, both of which are government-sponsored, peer-reviewed, and directly relevant to the Australian setting [Textbox 2]. The domains in Table 1 are an abridged synthesis created by the authors, drawing on elements from the two selected frameworks. This reimagined structure is intended to facilitate the development of a digital health curriculum for dentistry, aligning learning objectives, instruction, and assessment with the national strategy in Australia.

Textbox 2. Frameworks selected to inform development of the dental digital health curriculum.

1. Framework 1 (2018): eHealth Capabilities Framework for Graduates and Health Professionals [34]. This framework was developed by the University of Sydney and eHealth New South Wales, consisting of a tri-phase literature review, focus groups with faculty and government representatives (n=23), and a Delphi method refinement with 4 iterations. The framework is structured in 4 domains and describes recommended knowledge and skills for health professions graduates in digital health.
2. Framework 2 (2021): Digital Health Capability Framework for Allied Health Professionals [35]. This framework was developed by the Department of Health, Victoria, and consisted of a 3-part development program including a competency framework review, expert discussion panel interviews (n=28), and an online survey of Victorian allied health professionals (n=164). This document draws from Framework 1 and is similarly structured into 4 domains of 3-6 subdomains, with the addition of levels of digital health proficiency ranging from Foundation, Consolidation, Expert, and Leadership.

Table . Domains and goals for digital health curriculum for an entry-to-practice dental degree.

Domain	Learning goal	Suggested learning topics
1. Digital transformation of health	Newly graduated dental practitioners will actively lead the digital transformation of dentistry by using technology to deliver patient-centred care and by recognizing the role of data and analytics in improving it.	Electronic health records, digital dentistry (radiography, intraoral scanning, CAD/CAM[a], and other digital workflows) data, interoperability and learning health systems, artificial intelligence
1. Legislation, policy, and governance	Newly graduated dental practitioners will drive improvements in the privacy and security of patient data, and model the safe, ethical, and responsible use of digital health technologies in the dental practice.	Data privacy and cybersecurity
1. Digital health for patients	Newly graduated dental practitioners will promote patient engagement in health care, prescribe appropriate digital resources, and support digital health literacy.	Digital health literacy, patient engagement in health care, and digital health equity
1. Digital professionalism	Newly graduated dental practitioners will model a professional and appropriate digital identity.	Social media and digital professionalism

[a]CAD/CAM: computer-aided design/computer-aided manufacturing.

The first domain recognizes that along with technical proficiency in digital clinical workflows, dental practitioners must be able to think in multidisciplinary terms of the flow of data and information across health care [13]. Dental practitioners must understand the importance of informatics, interoperability, and a quality improvement mindset to be the building blocks for creating LHS [24,25].

The second domain recognizes the role of the dental practitioner in safe and ethical governance of patient data across digital workflows, noting that health care is the consistently top-reporting sector for data breaches in Australia [36].

The third domain recognizes the shift from the paternalistic model of health care towards a person-centered one where the person receiving care plays an active role in shared health care decision-making. The OpenNotes mandate in the US is a good

XSL•FO
RenderX

example of this [37]. This domain is also particularly relevant to the rapid pace of AI development and the accessibility of generative AI models that patients may use to access health (mis)information. Dental practitioners must understand digital health literacy; how patients may engage with digital health technologies and services; and the uses, ethics, benefits, and risks of AI in health care.

The fourth domain recognizes that dental practitioners must develop a professional identity, which is multidimensional across social media and the internet. The obligation for a dental practitioner to uphold their professional code of conduct is binding for both their in-person and digital profiles [38].

This holistic overview of digital health in dentistry is a step towards addressing the observation that digital health education tends to be focused on medical degrees—mostly in electives or single-unit areas such as telehealth—and in utilizing diverse approaches for delivery, development, and assessment [39]. A similar observation was found during our curricular needs assessment, revealing a strong focus in single content areas such as telehealth and digital dentistry, but confirming opportunities to facilitate a more coordinated and comprehensive learning pathway to support full digital health competency.

Final Thoughts

Viewing dentistry through a "digital health" lens may seem like a small matter. However, the change in perspective for dental educators is important. Dentistry has traditionally focused on individual patient care and procedural intervention, but contemporary health care is increasingly shaped by system-level forces. AI, interoperability, value-based care, and increasing consumer participation are now current realities [40-43]. The potential for digital health to drive meaningful systemic improvements in oral health and health care cannot be truly realized without first building the necessary capability at the graduate level. Consequently, these topics can and should be taught in a structured manner in entry-to-practice dental education.

Critically, although newer generations are often seen as digitally adept, they do not automatically master the necessary digital skills simply from being exposed to technology [44]. This gap in digital competency underscores the importance of intentional curriculum design. Universities are increasingly using the approach of constructive alignment to enhance outcome-based education [45], and this approach should be used to design a longitudinal digital health curriculum that can align with the intended graduate attributes.

This viewpoint has outlined the premise for designing a digital health curriculum in dentistry, using a structured set of domains based on current evidence and adapted to the Australian context. The proposed domains provide a foundation for educators to build a curriculum that aligns with the unique needs of dental professionals and the national strategy for digital health. This approach is intended for integration into the University of Melbourne's dentistry program and aims to encourage the further development and discussion of digital health education within dental programs, both nationally and globally.

Conclusion

It can be difficult for educators to define digital health curriculum that is both evidence-based and relevant to their discipline and local context; to design it is to predict the future. However, keeping pace involves changing our view of digital health in dentistry. A common understanding about the language of digital health is important for developing health professionals who will be able to navigate the environment of the modern health care system. We found that existing digital health capability frameworks were useful to define a view of digital health across an entry-to-practice dental degree, and high level roadmaps and frameworks are valuable to envision a future-ready dental graduate who can embrace the next wave of digital transformation. This perspective will be useful for developing the curriculum aligned with the national vision of building workforce capability and realizing the aim of safe, connected care.

Acknowledgments

Our sincere thanks to Alastair Sloan, Melbourne Dental School's Head of School, and the Centre for Digital Transformation of Health for supporting this educational initiative.

Authors' Contributions

Conceptualization: MM
Supervision: SB, KL
Writing – original draft: MM
Writing – review & editing: MM, SB, LS, KL

Conflicts of Interest

None declared.

Multimedia Appendix 1
Digital health capability and competency frameworks considered for curriculum development in dentistry.
[DOCX File, 22 KB - mededu_v11i1e54153_app1.docx]

XSL•FO
RenderX

References

1. Australian Digital Health Agency. National digital health capability action plan. 2022. URL: https://www.digitalhealth.gov.au/sites/default/files/documents/national-digital-health-capability-action-plan.pdf [accessed 2023-08-31]
2. Kraus S, Schiavone F, Pluzhnikova A, Invernizzi AC. Digital transformation in healthcare: Analyzing the current state-of-research. J Bus Res 2021 Feb;123:557-567. [doi: 10.1016/j.jbusres.2020.10.030]
3. Fatehi F, Samadbeik M, Kazemi A. What is Digital health? Review of definitions. Stud Health Technol Inform 2020 Nov 23;275:67-71. [doi: 10.3233/SHTI200696] [Medline: 33227742]
4. World Health Organization. Global strategy on digital health 2020–2025. 2021. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2023-08-31]
5. European Commission. EHealth: digital health and care: overview. 2017. URL: https://health.ec.europa.eu/ehealth-digital-health-and-care/overview_en [accessed 2024-09-19]
6. U.S. Food & Drug Administration. What is digital health?. 2020. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health [accessed 2024-09-19]
7. Gray K, Dattakumar A, Maeder A, Butler-Henderson K, Chenery H. Advancing eHealth education for the clinical health professions. 2014. URL: http://clinicalinformaticseducation.pbworks.com/w/file/fetch/74500403/PP10_1806_Gray_report_2014.pdf [accessed 2024-07-26]
8. Brice S, Almond H. Health professional digital capabilities frameworks: a scoping review. J Multidiscip Healthc 2020;13:1375-1390. [doi: 10.2147/JMDH.S269412] [Medline: 33173300]
9. Longhini J, Rossettini G, Palese A. Digital health competencies among health care professionals: systematic review. J Med Internet Res 2022 Aug 18;24(8):e36414. [doi: 10.2196/36414] [Medline: 35980735]
10. Jimenez G, Spinazze P, Matchar D, et al. Digital health competencies for primary healthcare professionals: a scoping review. Int J Med Inform 2020 Nov;143:104260. [doi: 10.1016/j.ijmedinf.2020.104260] [Medline: 32919345]
11. Nazeha N, Pavagadhi D, Kyaw BM, Car J, Jimenez G, Tudor Car L. A digitally competent health workforce: scoping review of educational frameworks. J Med Internet Res 2020 Nov 5;22(11):e22706. [doi: 10.2196/22706] [Medline: 33151152]
12. Global Digital Health Monitor. State of global health report 2023. URL: https://digitalhealthmonitor.org/stateofdigitalhealth23 [accessed 2024-07-26]
13. Australian Digital Health Agency. Connecting Australian healthcare – national healthcare interoperability plan 2023-2028. 2023. URL: https://www.digitalhealth.gov.au/sites/default/files/documents/national-healthcare-interoperability-plan-2023-2028.pdf [accessed 2023-10-16]
14. Australasian institute of digital health. About workforce advancement. URL: https://digitalhealth.org.au/ [accessed 2024-07-26]
15. Centre for Digital Transformation of Health. Connecting digital innovation to health. URL: https://mdhs.unimelb.edu.au/digitalhealth [accessed 2024-07-26]
16. Garde S, Harrison D, Huque M, Hovenga EJS. Building health informatics skills for health professionals: results from the Australian Health Informatics Skill Needs Survey. Aust Health Rev 2006 Feb;30(1):34-45. [Medline: 16448376]
17. Gray K, Sim J. Factors in the development of clinical informatics competence in early career health sciences professionals in Australia: a qualitative study. Adv Health Sci Educ Theory Pract 2011 Mar;16(1):31-46. [doi: 10.1007/s10459-010-9238-3] [Medline: 20544387]
18. Dattakumar A, Gray K, Henderson KB, Maeder A, Chenery H. We are not educating the future clinical health professional workforce adequately for e-health competence: findings of an Australian study. Stud Health Technol Inform 2012;178:33-38. [Medline: 22797016]
19. Butler-Henderson K, Dalton L, Probst Y, Maunder K, Merolli M. A meta-synthesis of competency standards suggest allied health are not preparing for a digital health future. Int J Med Inform 2020 Dec;144:104296. [doi: 10.1016/j.ijmedinf.2020.104296] [Medline: 33091830]
20. Kalenderian E, Zouaidi K, Yeager J, et al. Learning from data in dentistry: Summary of the third annual openwide conference. Learn Health Syst 2024 Apr;8(2):e10398. [doi: 10.1002/lrh2.10398] [Medline: 38633022]
21. de Ahumada Servant P, Martín-Martín D, Romero I. Digital transformation of oral health care: measuring the digitalization of dental clinics. Soc Indic Res 2024. [doi: 10.1007/s11205-024-03366-z]
22. Mourouzis P, Tolidis K. CAD/CAM systems. In: Delantoni A, Orhan K, editors. Digital Dentistry: An Overview and Future Prospects, 1st edition: Springer; 2024:47-66. [doi: 10.1007/978-3-031-52826-2_5]
23. Schwendicke F, Krois J. Data dentistry: how data are changing clinical care and research. J Dent Res 2022 Jan;101(1):21-29. [doi: 10.1177/00220345211020265] [Medline: 34238040]
24. Institute of Medicine (US). Institute of medicine roundtable on value and science-driven health care. In: Grossmann C, Powers B, McGinnis JM, editors. Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary: National Academies Press; 2011. URL: https://www.ncbi.nlm.nih.gov/books/NBK83568/ [accessed 2024-10-30]
25. Smith M, Saunders R, Stuckhardt L, McGinnis JM. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America: National Academies Press (US); 2013. [doi: 10.17226/13444] [Medline: 24901184]

XSL•FO
RenderX

26. Australian Dental Council. Professional competencies of the newly qualified dental practitioner. 2022. URL: https://www.adc.org.au/files/accreditation/competencies/ADC_Professional_Competencies_of_the_Newly_Qualified_Practitioner.pdf [accessed 2023-08-31]
27. Schneiderhan J, Guetterman TC, Dobson ML. Curriculum development: a how to primer. Fam Med Community Health 2019;7(2):e000046. [doi: 10.1136/fmch-2018-000046] [Medline: 32148703]
28. Richardson A. Dental services in Australia: industry report Q8531. : IBISWorld; 2022. URL: https://www.ibisworld.com/ [accessed 2023-08-31]
29. Nanayakkara S, Zhou X, Spallek H. Impact of big data on oral health outcomes. Oral Dis 2019 Jul;25(5):1245-1252. [doi: 10.1111/odi.13007] [Medline: 30474902]
30. American College of Prosthodontists. Digital dentistry curriculum for predoctoral and advanced education in prosthodontics. 2018. URL: https://www.prosthodontics.org/assets/1/7/ACP_Digital_Dentistry_Curriculum.pdf [accessed 2023-08-31]
31. Orrel J. Expert advice on designing authentic assessments for online delivery. : Tertiary Education Quality and Standards Agency; 2022. URL: https://www.teqsa.gov.au/sites/default/files/2022-10/assessment-for-online-delivery-orrell-4may2020.pdf [accessed 2024-10-30]
32. Contact North. How assessment is changing in the digital age – five guiding principles. 2020. URL: https://teachonline.ca/tools-trends/how-assessment-changing-digital-age-five-guiding-principles [accessed 2024-10-30]
33. Valenta AL, Berner ES, Boren SA, et al. AMIA Board White Paper: AMIA 2017 core competencies for applied health informatics education at the master's degree level. J Am Med Inform Assoc 2018 Dec 1;25(12):1657-1668. [doi: 10.1093/jamia/ocy132] [Medline: 30371862]
34. Brunner M, McGregor D, Keep M, et al. An eHealth capabilities framework for graduates and health professionals: mixed-methods study. J Med Internet Res 2018 May 15;20(5):e10229. [doi: 10.2196/10229] [Medline: 29764794]
35. Littlewood N, Downie S, Sawyer A, Feely K, Govil D, Gordon B. Development of a digital health capability framework for allied health practitioners: an Australian first. IJAHSP 2021;20(3):22. [doi: 10.46743/1540-580X/2022.2234]
36. Office of the Australian Information Commissioner. Notifiable data breaches report july to december 2023. 2023. URL: https://www.oaic.gov.au/privacy/notifiable-data-breaches/notifiable-data-breaches-publications/notifiable-data-breaches-report-july-to-december-2023 [accessed 2024-07-26]
37. OpenNotes. Putting patient safety first with OpenNotes. URL: https://www.opennotes.org/ [accessed 2024-07-26]
38. Australian Health Practitioner Regulation Agency. Social media: how to meet your obligations under the national law. URL: https://www.ahpra.gov.au/Resources/Social-media-guidance.aspx [accessed 2024-07-26]
39. Tudor Car L, Kyaw BM, Nannan Panday RS, et al. Digital health training programs for medical students: scoping review. JMIR Med Educ 2021 Jul 21;7(3):e28275. [doi: 10.2196/28275] [Medline: 34287206]
40. Shan T, Tay FR, Gu L. Application of artificial intelligence in dentistry. J Dent Res 2021 Mar;100(3):232-244. [doi: 10.1177/0022034520969115] [Medline: 33118431]
41. Vujicic M, David G. Value-based care in dentistry: Is the future here? J Am Dent Assoc 2023 Jun;154(6):449-452. [doi: 10.1016/j.adaj.2023.04.001] [Medline: 37097278]
42. Childers C, Marron J, Meyer EC, Abel GA. Clinical ethics consultation documentation in the era of open notes. BMC Med Ethics 2023 May 3;24(1):27. [doi: 10.1186/s12910-023-00904-1] [Medline: 37138339]
43. Schwendicke F, Chaurasia A, Wiegand T, et al. IADR e-oral health network and the ITU/WHO focus group AI for health. Artif Intell Oral Dent Healthc: Core Educ Curr J Dent 2023;128:104363. [doi: 10.1016/j.jdent.2022.104363] [Medline: 36410581]
44. Cham K, Edwards ML, Kruesi L, Celeste T, Hennessey T. Digital preferences and perceptions of students in health professional courses at A leading Australian university: a baseline for improving digital skills and competencies in health graduates. AJET 2021;38(1):69-86. [doi: 10.14742/ajet.6622]
45. Biggs JB, Tang CS. Teaching for Quality Learning at University: What the Student Does, 4th edition: Society for Research into Higher Education& Open University Press; 2011.

Abbreviations

ADC: Australian Dental Council
ADHA: Australian Digital Health Agency
AI: artificial intelligence
AIDH: Australasian Institute of Digital Health
CAD/CAM: computer-aided design/computer-aided manufacturing
LHS: learning health system
mHealth: mobile health
TEQSA: Tertiary Education Quality and Standards Agency

XSL•FO
RenderX

Edited by P Kanzow; submitted 31.10.23; peer-reviewed by M Pang, S Brice; revised version received 30.10.24; accepted 31.10.24; published 08.01.25.

Please cite as:
Mun M, Byrne S, Shaw L, Lyons K
Digital Dentists: A Curriculum for the 21st Century
JMIR Med Educ 2025;11:e54153
URL: https://mededu.jmir.org/2025/1/e54153
doi:10.2196/54153

© Michelle Mun, Samantha Byrne, Louise Shaw, Kayley Lyons. Originally published in JMIR Medical Education (https://mededu.jmir.org), 8.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feedback From Dental Students Using Two Alternate Coaching Methods: Qualitative Focus Group Study

Lulwah Alreshaid[1,2,3], BDS, PhD; Rana Alkattan[1,2,3], BDS, MSD, PhD

[1]Department of Restorative and Prosthetic Dental Sciences, College of Dentistry, King Saud bin Abdulaziz University for Health Sciences, P.O. Box 22490, Riyadh, Saudi Arabia

[2]Ministry of National Guard Health Affairs, King Abdullah International Medical Research Centre, Riyadh, Saudi Arabia

[3]Dental Services, King Abdulaziz Medical City, Ministry of the National Guard Health Affairs, Riyadh, Saudi Arabia

Corresponding Author:
Rana Alkattan, BDS, MSD, PhD
Department of Restorative and Prosthetic Dental Sciences, College of Dentistry, King Saud bin Abdulaziz University for Health Sciences, P.O. Box 22490, Riyadh, Saudi Arabia

Abstract

Background: Student feedback is crucial for evaluating the effectiveness of institutions. However, implementing feedback can be challenging due to practical difficulties. While student feedback on courses can improve teaching, there is a debate about its effectiveness if not well-written to provide helpful information to the receiver.

Objective: This study aimed to evaluate the impact of coaching on proper feedback given by dental students in Saudi Arabia.

Methods: A total of 47 first-year dental students from a public dental school in Riyadh, Saudi Arabia, completed 3 surveys throughout the academic year. The surveys assessed their feedback on a Dental Anatomy and Operative Dentistry course, including their feedback on the lectures, practical sessions, examinations, and overall experience. The surveys focused on assessing student feedback on the knowledge, understanding, and practical skills achieved during the course, as aligned with the defined course learning outcomes. The surveys were distributed without coaching, after handout coaching and after workshop coaching on how to provide feedback, designated as survey #1, survey #2, and survey #3, respectively. The same group of students received all 3 surveys consecutively (repeated measures design). The responses were then rated as neutral, positive, negative, or constructive by 2 raters. The feedback was analyzed using McNemar test to compare the effectiveness of the different coaching approaches.

Results: While no significant changes were found between the first 2 surveys, a significant increase in constructive feedback was observed in survey #3 after workshop coaching compared with both other surveys ($P<.001$). The results also showed a higher proportion of desired changes in feedback, defined as any change from positive, negative, or neutral to constructive, after survey #3 ($P<.001$). Overall, 20.2% reported desired changes at survey #2% and 41.5% at survey #3 compared with survey #1.

Conclusions: This study suggests that workshops on feedback coaching can effectively improve the quality of feedback provided by dental students. Incorporating feedback coaching into dental school curricula could help students communicate their concerns more effectively, ultimately enhancing the learning experience.

(JMIR Med Educ 2025;11:e68309) doi:10.2196/68309

KEYWORDS

student feedback; coaching; dental education; student evaluation; teaching methods; educational intervention

Introduction

Feedback, a cornerstone of effective performance improvement, plays a crucial role in various domains, including education. Understanding how feedback is delivered and received is essential to maximize its impact. Several models provide frameworks for analyzing feedback processes, such as Hattie and Timperley's [1] model, which categorizes feedback based on its focus (task, process, and self-regulation), and Kluger and DeNisi's [2] Feedback Intervention Theory, which explores the instructional, motivational, emotional, and learning effects of feedback. These models highlight the complexities of feedback delivery and the importance of considering the recipient's needs and the specific context. These models go beyond simple evaluation and rather focus on providing actionable information that supports student learning and development.

Feedback is a critical method of measuring the effectiveness of performance and outcome of any institution. More importantly, if these institutions play an important role in education, health, or essential services, it is crucial to use student feedback to ensure the successful performance of these institutions. Feedback is often challenging to execute due to interaction issues or practical applicability [3,4]. Challenges arise from a complex interaction between the providers and recipients'

XSL•FO
RenderX

performance [4]. An example of these challenges could be the fear of recognizing unsatisfactory performance, discouragements, and liability. However, feedback's primary purpose is to improve the outcome. Delivering productive feedback to assess teaching procedures and students' experience is critical for effective learning and developing a solid connection between feedback providers and recipients [5-7]. In addition, it serves to evaluate teaching strategies. By aligning with the principles of key feedback models, the overall learning experience can be enhanced for both students and faculty.

Giving feedback to recipients can be complex; however, various techniques have been reported in the literature; 1 of the popular techniques is the "compliment sandwich," in which the recipient receives 1 criticism between 2 positive comments [8]. In contrast, another effective technique is to eliminate the negative connotation of feedback, in which the feedback provider mentions the mistakes and provides some solutions [9]. In any case, it is important to note that effective feedback comprises structure, content, and time [10]. When this feedback is expected from students to instructors, another level of challenge can be anticipated [11]; certain boundaries between the students and their instructors may restrict students' ability to express themselves freely. Students may also perceive end-of-course feedback as a mere administrative requirement fulfilling curricular mandates, potentially diminishing their perceived value and engagement in the process. Hence, to give constructive feedback, it is essential to guide students to the fact that the goal is not to deliver the feedback by criticizing but to enhance the feedback process to be more effective and constructive [12,13].

Many educational institutions imply student and professor feedback concerning courses in which they are both involved [14]. The feedback from the students usually involves a set of surveys to rate a course and the instructor giving that course. This process could assist the instructors in better recognizing areas of strengths and weaknesses, ultimately improving the educational experience [15-17]. Debate emerges that questions the effectiveness of such feedback [15,18-20]. A recent study found that implementing feedback could be beneficial if incorporated into the curriculum while also providing instructors with how to receive such feedback and how to adapt to these comments [17,21,22]. Furthermore, another author highlighted the importance of student evaluation and excelling in education, which could provide the instructor with minor adjustments to reform the course [21-23]. In contrast, some instructors note that this feedback will not encourage them to modify their courses [23]. Furthermore, some instructors might find it difficult to solely base altering decisions on input provided by students, arguing that some aspects will affect the student's ability to provide trustworthy information based on factors such as the ability to construct critical feedback or complex circumstances, including age, gender, or educational background [18,24].

Although previous studies assessed the effect of feedback given by students on teaching quality and the improvement of feedback over a certain period [14,18], the relations between coaching to give and receive feedback and the feedback received from students after coaching have not been investigated among dental students in Saudi Arabia. Teaching students how to provide reflective, constructive feedback to elicit better outcomes for course, curriculum, and general educational development would be significant. Thus, the primary objective of this study was to evaluate the feedback given by students in the College of Dentistry, King Saud bin Abdulaziz University for Health Sciences (KSAU-HS), after using 2 different coaching approaches on how to provide feedback. The secondary objective was to improve the effectiveness of the feedback given by dental students after exposing them to 2 different coaching approaches on how to provide feedback. The null hypothesis of the current study is twofold as there is no difference in the nature of the feedback given by the students using 1 coaching methods, and both the coaching methods increase the proportion of constructive student feedback equally.

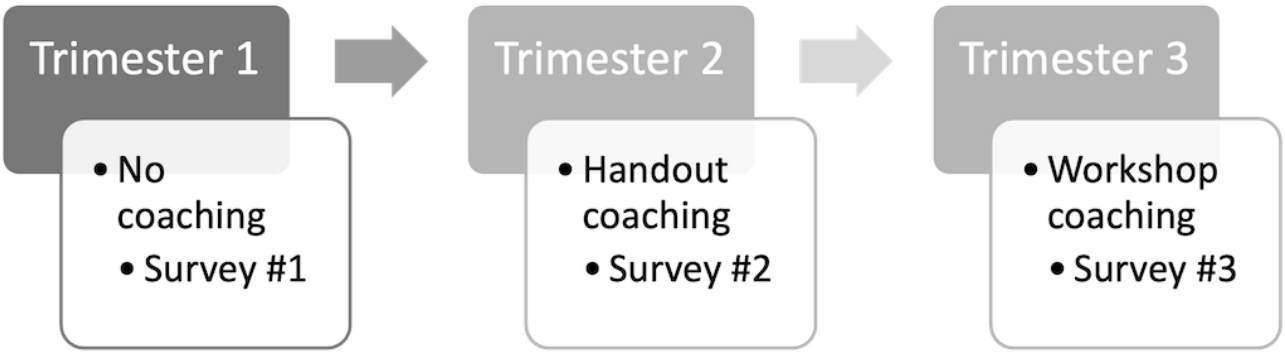
Methods

Overview

In total, 50 students were invited to participate in the study. Students were asked to complete 3 surveys in open-ended question format at the end of each trimester (repeated measures design, Figure 1). These surveys asked the same questions but were specific to each trimester. Invited students were asked to provide feedback on the RSTO 311 course. This study focused on first-year dental students enrolled in their introductory dental course, which serves as their initial exposure to fundamental dental concepts, including tooth anatomical landmarks, cavity preparation techniques, and restorative procedures in both theoretical and simulated clinical environments. Each survey consisted of 5 questions, adopted by Hajhamid and Somogyi‐Ganss [16]. The first question was to indicate the 2-digit number assigned to each student by a research assistant who never interacted with the students to ensure anonymity. The second question was about the lectures given during the trimester. The third question was about the practical sessions taken during the trimester. The fourth question was about the quizzes, written and practical exams taken during the trimester. The fifth and last question was about the overall course (Multimedia Appendix 1). Based on the course learning outcomes, the students are expected to meet minimum criteria of knowledge and understanding as well as practical skills; thus, the survey focused on these aspects of the course.

XSL•FO
RenderX

Figure 1. Study design showing coaching methods for feedback given in surveys #1, #2, and #3.



At the beginning of the course, an invitation was sent to all students taking the RSTO 311 course. The students were offered a bonus (2 grades) if they participated in the study. If they wished not to participate, they could write an essay about a topic related to their course and get the same bonus grades. The survey was designed using Google Forms and was sent by email to all participating students; the survey link was sent by the same research assistant who assigned the 2-digit numbers to participating students. Consent was obtained from all participating students at the beginning of each survey. Before completing the first survey at the end of the first trimester, no coaching or instructions were given to the students on how to receive and provide feedback. Before completing the second survey at the end of the second trimester, students were coached by reading a 2-page handout on how to receive and provide feedback, which can be covered in approximately 10 minutes (Multimedia Appendix 2). The handout explains the different types of feedback, as well as steps and examples for giving constructive feedback. Before completing the third and last survey at the end of the third trimester, students were coached by attending a 1-hour workshop on how to receive and provide constructive feedback. The workshop was given by a faculty member who was not involved in the course or the research study. The workshop similarly explained different theories on giving and receiving feedback and demonstrated to the students how to improve feedback with examples. Both the handout and the workshop were based on a previously published paper [16].

All 3 surveys' answers were evaluated independently by 2 raters, the course director and the co-course director of the course. The answers were rated as either neutral, positive, negative, or constructive feedback. Any disagreement between the 2 evaluators' ratings of the survey answers was discussed and agreed upon before the analysis. Answers were considered neutral feedback if there were no positive, negative, or constructive comments. Answers were deemed positive if general praise was included. Answers were considered negative if the provided nonconstructive criticism. Finally, answers were considered constructive if there were any suggestions to improve the course in any aspect, even if they contained any positive or negative comments. When rater disagreement was noted in cases where responses were not clearly considered as positive or negative feedback and did not include any constructive comment or intent for improvement, responses were rated as neutral. Data were collected and analyzed based on the ratings given by the 2 evaluators and then compared between surveys.

Student feedback was collected in textual form and subsequently coded into 4 categories: positive, negative, neutral, and constructive. The reliability of this coding process was ensured through independent assessments by 2 raters. Kappa statistics was used to assess inter-rater reliability between the 2 raters. The ratings followed a nominal scale (1=neutral, 2=positive, 3=negative, and 4=constructive); hence, frequency and proportions were reported for the ratings as descriptive statistics. Inferential statistical analysis was used to test rating changes over time (McNemar test). The level of significance of .05 was used for inferential analysis with P values <.05 reported as statistically significant. Analysis was performed combined for 4 questions as well as separately for each question. IBM SPSS Statistics software (version 29) was used for descriptive and inferential analysis.

Ethical Considerations

An institutional review board ethical approval was obtained from King Abdullah International Medical Research Center for this cross-sectional study (IRB/3004/23). This study was conducted during the academic year 2023‑2024 among first-year dental students who took the Dental Anatomy and Operative Dentistry course (RSTO 311) at the College of Dentistry, KSAU-HS, a public dental school in Riyadh, Saudi Arabia. The RSTO 311 is a yearly course divided over 3 trimesters. The course has theoretical and practical components: 30 lectures and 40 practical sessions. The students were assessed based on weekly continuous assessment, 3 quizzes, 3 written exams, and 3 practical exams.

Results

Of the 50 students in the class, 47 participants (25 male and 22 female participants) were included who completed all 3 surveys at 3 different time points, giving a participation rate of 94%. Out of 50, 1 student dropped the course, 1 refused to participate, and 1 failed to complete the third survey.

The 2 raters provided a total of 564 ratings each. Overall, 541 out of 564 ratings matched, suggesting a 95.9% level of agreement. The κ value was 0.941, which, being above 0.9, indicates an almost perfect level of agreement between the raters, demonstrating a high degree of reliability in the classification of responses. Discrepancy in data was discussed, re-evaluated, and a final agreement was reached and recorded. The following are randomly selected examples presented from students' feedback:

XSL•FO
RenderX

Neutral feedback: "No complaints about it"

Positive feedback: "The course provided a solid foundation in the subject matter, it was a valuable learning opportunity"

Negative feedback: "The work was hard and tiring and the time was not enough"

Constructive feedback: "In some anatomy lectures, clearer explanations were needed. Providing a short video would offer better visualization for students"

Within-subject analysis was conducted separately for each of the 4 questions in the 3 surveys. No significant changes were observed between survey #1 and #2 in any of the 4 questions, separately or combined. However, there were statistically significant changes between survey #1 and #3 with regards to increase in proportion of constructive ratings for questions 2 - 4 as well as for the 4 questions combined. Significant change in ratings was also found in survey #3 relative to survey #2 for questions 1 - 3 as well as for the 4 questions combined (Table 1).

Table . Ratings for each of the 4 questions at each of the 3 surveys.

Question and rating		Survey #1, n (%)	Survey #2, n (%)	Survey #3, n (%)
#1[a]				
	Neutral	10 (21.3)	9 (19.1)	13 (27.7)
	Positive	4 (8.5)	2 (4.3)	0 (0)
	Negative	15 (31.9)	25 (53.2)	9 (19.1)
	Constructive	18 (38.3)	11 (23.4)	25 (53.2)
#2[b]				
	Neutral	13 (27.7)	11 (23.4)	11 (23.4)
	Positive	3 (6.4)	8 (17)	1 (2.1)
	Negative	15 (31.9)	14 (29.8)	1 (2.1)
	Constructive	16 (34)	14 (29.8)	34 (72.3)
#3[c]				
	Neutral	27 (57.4)	24 (51.1)	15 (31.9)
	Positive	6 (12.8)	2 (4.3)	1 (2.1)
	Negative	4 (8.5)	10 (21.3)	6 (12.8)
	Constructive	10 (21.3)	11 (23.4)	25 (53.2)
#4[d]				
	Neutral	17 (36.2)	16 (34)	12 (25.5)
	Positive	1 (2.1)	1 (2.1)	2 (4.3)
	Negative	18 (38.3)	12 (25.5)	8 (17)
	Constructive	11 (23.4)	18 (38.3)	25 (53.2)
All 4 combined[e]				
	Neutral	67 (35.6)	60 (31.9)	51 (27.1)
	Positive	14 (7.4)	13 (6.9)	4 (2.1)
	Negative	52 (27.7)	61 (32.4)	24 (12.8)
	Constructive	55 (29.3)	54 (28.7)	109 (58)

[a] McNemar test for survey 2 vs survey 1: $\chi^2 5=5.86$, $P=.32$; McNemar test for survey 3 vs survey 1: $\chi^2 5=6.64$, $P=.25$; McNemar test for survey 3 vs survey 2: $\chi^2 3=13.07$, $P=.004$

[b] McNemar test for survey 2 vs survey 1: $\chi^2 6=4.63$, $P=.59$; McNemar test for survey 3 vs survey 1: $\chi^2 5=18.26$, $P=.003$; McNemar test for survey 3 vs survey 2: $\chi^2 5=23.30$, $P<.001$

[c] McNemar test for survey 2 vs survey 1: $\chi^2 6=6.10$, $P=.41$; McNemar test for survey 3 vs survey 1: $\chi^2 5=15.87$, $P=.01$; McNemar test for survey 3 vs survey 2: $\chi^2 5=14.53$, $P=.006$

[d] McNemar test for survey 2 vs survey 1: $\chi^2 5=5.31$, $P=.38$; McNemar test for survey 3 vs survey 1: $\chi^2 4=10.81$, $P=.03$; McNemar test for survey 3 vs survey 2: $\chi^2 5=5.62$, $P=.35$

[e] McNemar test for survey 2 vs survey 1: $\chi^2(6)=5.28$, $P=.51$; McNemar test for survey 3 vs survey 1: $\chi^2(5)=33.43$, $P<.001$; McNemar test for survey 3 vs survey 2: $\chi^2(5)=45.28$, $PP<.001$

XSL•FO
RenderX

Table 2 shows the proportion of constructive versus nonconstructive (positive, negative, or neutral) ratings for each question and for all 4 questions combined. A significant increase in the proportion of constructive ratings was found between survey #1 and survey #3 for questions 2 - 4 as well as for the 4 questions combined. A significant increase in the proportion of constructive ratings was also found between survey #2 and survey #3 for questions 1 - 3 as well as for the 4 questions combined.

Table . Proportion of constructive ratings.

Question and rating		Survey #1	Survey #2	Survey #3
#1				
	Nonconstructive[a]	29 (61.7)	36 (76.6)	22 (46.8)
	Constructive	18 (38.3)	11 (23.4)	25 (53.2)
			MN_1(b) *P*=.19	MN_1[b](b) *P*=.23
				MN_2[c](b) *P*=.003
#2				
	Nonconstructive	31 (66)	33 (70.2)	13 (27.7)
	Constructive	16 (34)	14 (29.8)	34 (72.3)
			MN_1(b) *P*=.83	MN_1(b) *P*<.001
				MN_2(b) *P*<.001
#3				
	Nonconstructive	37 (78.7)	36 (76.6)	22 (46.8)
	Constructive	10 (21.3)	11 (23.4)	25 (53.2)
			MN_1(b) *P*>.99	MN_1(b) *P*=.003
				MN_2(b) *P*<.001
#4				
	Nonconstructive	36 (76.6)	29 (61.7)	22 (46.8)
	Constructive	11 (23.4)	18 (38.3)	25 (53.2)
			MN_1(b) *P*=.14	MN_1(b) *P*=.01
				MN_2(b) *P*=.14
All 4 combined				
	Nonconstructive	133 (70.7)	134 (71.7)	79 (42)
	Constructive	55 (29.3)	54 (28.7)	109 (58)
			MN_1(b) *P*>.99	MN_1(b) *P*<.001
				MN_2(b) *P*<.001

[a] nonconstructive ratings include positive, negative and neutral.

[b] MN_1(b)=McNemartest using binomial distribution to examine change from survey #1.

[c] MN_2(b)=McNemartest using binomial distribution to examine change from survey #2.

For each question, the change from survey #1 was coded as desired versus not desired. Desired change was defined as any change from positive, negative, or neutral to constructive. All other changes were coded as not desired. The proportion of desired changes is summarized in Table 3. Survey #3 showed a higher proportion of desired changes compared with survey #2. For the 4 questions combined, 20.2% had desired changes at survey #2% and 41.5% at survey #3 compared with survey #1. In survey #3, the most frequent changes reported overall for the 4 questions combined were: neutral to constructive (17.6%), negative to constructive (16.5%) and constructive to constructive (16.5%).

XSL•FO
RenderX

Table . The proportion of desired changes in surveys #2 and #3 compared with survey #1.

Proportion of desired changes	Survey (#2 versus #1), n (%)	Survey (#3 versus #1), n (%)
Question 1	7 (14.9%)	16 (34%)
Question 2	10 (21.3%)	23 (48.9%)
Question 3	9 (19.1%)	19 (40.4%)
Question 4	12 (25.5%)	20 (42.6%)
Four questions combined	38 (20.2%)	78 (41.5%)

Discussion

Principal Findings

This study compared student responses without coaching, coaching using a feedback handout, or coaching using a feedback workshop before completing the surveys. Results demonstrate that handout coaching showed no significant difference compared with no coaching with respect to the number of neutral, positive, negative, or constructive ratings. However, workshop coaching significantly increased the number of constructive ratings compared with both no coaching and handout coaching ($P<.001$, Table 1). Therefore, the null hypothesis was rejected. The reason for these results could be due to the fact that handouts were distributed to the students, and they were asked to read the 2-page document independently. This method does not involve student and instructor interaction and is hence, less engaging. There was also no measure of whether the students in fact read the handout and grasped the information. Thus, no significant changes were noted between survey #1 and survey #2. Workshop coaching, on the other hand, was done in a classroom setting with 1 faculty member present, ensuring a 100% attendance rate of all participating students. Furthermore, the students were able to ask questions regarding the information presented in the workshop and were asked to fill out survey #3 immediately after the workshop, before leaving the classroom.

The proportion of constructive feedback, compared to nonconstructive feedback, significantly increased after workshop coaching (Table 2). The workshop-based format provided multiple examples in a story format from past student feedback, whereas the handout only stated the description of proper feedback writing without detailed examples compared with the examples presented in the workshop. The educational value of workshop coaching has been previously established, wherein the students are “active learners” and can engage in asking questions during the learning process [25,26]. Information presented in video format can also enhance information retention, owing to reduced student cognitive loading and optimized use of visual learner memory [27]. Furthermore, the key learning points are emphasized during the workshop, and audio-visual learning is more likely to keep the students more attentive and engaged in the content being delivered [28]. This is also demonstrated in Table 3, where the most frequently reported changes in feedback from survey #1 (no coaching) to survey #3 (workshop coaching) were from neutral and negative to constructive, reported in this study as “desired changes”.

The effectiveness of workshop coaching can also be understood through several educational and psychological frameworks. For example, the constructivist learning theory emphasizes the importance of social interaction and guided learning in developing cognitive skills [29]. The workshop format, which encourages active participation and immediate feedback from the instructor, aligns with this theory by fostering an environment where students engage with and construct their knowledge of feedback writing through scaffolding, wherein support is provided by a more knowledgeable person. This approach helps students internalize new feedback techniques through direct interaction and reflection on real examples. Furthermore, the importance of emotional intelligence in feedback delivery cannot be overlooked. According to Goleman [30], empathy and self-regulation are key components of emotional intelligence that influence how feedback is communicated. In the workshop setting, students are not only taught the mechanics of constructive feedback but also how to consider the emotional impact of their words, enhancing their ability to offer feedback that is both critical and supportive. This connection to emotional intelligence helps explain why the workshop coaching produced a higher proportion of constructive feedback compared with the handout coaching.

Comparison With Previous Work

In any educational environment, student satisfaction is an essential criterion for quality assessment [31]. Student evaluations of teaching are surveys typically used to collect, analyze, and interpret teaching quality [32]. Hence, every year, students are asked to evaluate the course material and provide feedback. In this study, the survey questions provided to the students concerned the lectures, practical sessions, and examinations at KSAU-HS. They were distributed immediately after the end of each trimester to ensure the feedback was relevant and firsthand. The purpose of these distributed surveys was to gather information on the course teaching, practical sessions, and facilities so that an action plan may be set to ensure improvement. However, most student feedback tends to be general or rely on their personal experience rather than providing helpful information related to the learning experience [33]. As this study is based on open-ended questions, analyzing responses can be quite intricate unless the process is made more structured. Hence, this study evaluated student responses after a handout and workshop coaching.

Written comments add value to both students and educators when compared with scale-type questions [34]. The students are given the possibility to explain their perspective beyond Likert-type scales and raise further topics that may not have

XSL•FO
RenderX

been covered in closed-ended questions [35]. Written comments are more informative for educators, and suggestions are beneficial when compared with receiving a statistical summary of quantitative results [36]. “Student evaluations of teaching” instruments can be a source of valuable thoughts from students and can help educators gain insight into how students perceive their learning experience and how different students learn best in a given setting [37]. However, these benefits can only be reliable after bringing a little order to the chaos of written responses.

The main purpose of the study was to improve the quality of feedback provided by the students. To the best of our knowledge, this is the first study introducing interactive workshop coaching for proper feedback among teaching institutes in Saudi Arabia. The workshop was able to improve the constructive criticism given by the students compared with self-learning using the handout. It is likely that the lower performance with handout coaching reflected less motivation, responsibility, or independence of the students [38]. These results are contrary to a previous similar study, in which both the handout and workshop coaching similarly improved student feedback [16]. The difference in results could be attributed to the nature of the dental school between both studies. This study was performed in a governmental dental school where students are not obliged to pay tuition fees. On the contrary, since their education is financed largely by loans, students from the Canadian private dental school may be more encouraged to commit to assigned tasks [39,40]. It is also worth noting that dental students at our institution are more familiar with lecture- and workshop-based learning as opposed to self-directed learning; as most dental schools in Saudi Arabia have not completely shifted from teacher-centered learning to a more interactive or evidence-based style [41]. Furthermore, culturally, expressing opinions, especially those with negative connotations or suggestive tones, may not necessarily be favored [42]. However, the results of this study clearly show the benefits of workshop coaching in directing students to provide their perception towards the course. This emphasizes the importance of including such a coaching approach for first-year students as part of the academic curriculum at the beginning of their studies.

Limitations

One of the limitations of this study was the inclusion of only first year students, as students in older years may have responded differently to the handout coaching, likely being more familiar with independent self-learning. Students in older years may also be more exposed to course-based surveys compared with first-year students. This also reduced the sample size of the participants. Furthermore, the difference between the topics covered over the 3 trimesters of the course may have influenced the feedback given by the students. In addition, when the students were given the third survey, they had already been exposed to both handout and workshop coaching on proper feedback, and this emphasis on appropriate feedback writing may have led to the higher number of constructive comments in survey #3. Furthermore, self-reported student feedback is subject to various biases, such as recall bias, acquiescence bias, social desirability bias, and cultural influences, which could impact the accuracy of the responses. Finally, the incentive of the bonus grades may have introduced self-selection bias; however, as the incentive was offered to all students equally, whether they participated in the survey or chose to submit an essay assignment, this may have mitigated the bias.

Conclusions

This study compared the effectiveness of 3 approaches, no coaching, handout coaching, and workshop coaching, on improving the quality of feedback provided by dental students. The results show that workshop coaching significantly increased the number of constructive feedback ratings, compared with both no coaching and handout coaching. This study encourages a more expressive feedback culture that facilitates student or instructor interaction in a constructive manner, wherein instructors can receive and implement feedback to improve the educational process. This suggests that interactive, instructor-led workshops foster a more engaged learning environment, encouraging students to provide higher-quality feedback. Given these findings, educators can implement interactive workshops focused on teaching students how to provide constructive feedback. These workshops should encourage active engagement through real-life examples and peer discussions. Given that the study shows significant benefits in first-year students, feedback coaching can be introduced early in the academic program. Building on the concept of scaffolding, educators could start with guided feedback exercises during the workshop, gradually increasing the level of independence as students become more proficient. Educators can also integrate emotional intelligence training into feedback workshops by helping students understand how to express feedback empathetically and how to regulate their emotions while providing feedback. Further studies evaluating different coaching methods to enhance student feedback are needed, with consideration to assign different methods to each study group. Future research should also investigate the impact of standardized coaching protocols on the quality of student feedback and use the data to improve assessment and learning outcomes.

Acknowledgments

We would like to thank the research assistant, Dr Jabir Al Humaid, for his valuable contribution.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

RSTO 311 course evaluation survey
[PDF File, 78 KB - mededu_v11i1e68309_app1.pdf]

Multimedia Appendix 2
How to provide constructive feedback
[PDF File, 91 KB - mededu_v11i1e68309_app2.pdf]

References

1. Hattie J, Timperley H. The power of feedback. Rev Educ Res 2007 Mar;77(1):81-112. [doi: 10.3102/003465430298487]
2. Kluger AN, DeNisi A. The effects of feedback interventions on performance: A historical review, A meta-analysis, and A preliminary feedback intervention theory. Psychol Bull 1996 Mar;119(2):254-284. [doi: 10.1037/0033-2909.119.2.254]
3. Jonsson A. Facilitating productive use of feedback in higher education. Active Learning in Higher Education 2013 Mar;14(1):63-76. [doi: 10.1177/1469787412467125]
4. Kowalski K. Giving and receiving feedback: part II. J Contin Educ Nurs 2017 Oct 1;48(10):445-446. [doi: 10.3928/00220124-20170918-04] [Medline: 28954179]
5. Biggs J. Teaching For Quality Learning at University: What the Student Does, 2nd edition: Open University Press; 2003.
6. Harris LR, Brown GTL, Harnett JA. Understanding classroom feedback practices: a study of New Zealand student experiences, perceptions, and emotional responses. Educ Asse Eval Acc 2014 May;26(2):107-133. [doi: 10.1007/s11092-013-9187-5]
7. McCarthy J. Evaluating written, audio and video feedback in higher education summative assessment tasks. Iss Educa Res 2015;25(2):153-169.
8. Cannon MD, Witherspoon R. Actionable feedback: Unlocking the power of learning and performance improvement. AMP 2005 May;19(2):120-134. [doi: 10.5465/ame.2005.16965107]
9. Emory CL. Pearls: giving and receiving feedback. Clin Orthop Relat Res 2019 Jan;477(1):35-36. [doi: 10.1097/CORR.0000000000000538] [Medline: 30586065]
10. Kruidering-Hall M, O'Sullivan PS, Chou CL. Teaching feedback to first-year medical students: long-term skill retention and accuracy of student self-assessment. J Gen Intern Med 2009 Jun;24(6):721-726. [doi: 10.1007/s11606-009-0983-z] [Medline: 19384559]
11. Ben-Porath S, Webster D. Free Speech and Education, 1st edition: Routledge; 2022.
12. van de Ridder JMM, Stokking KM, McGaghie WC, ten Cate OTJ. What is feedback in clinical education? Med Educ 2008 Feb;42(2):189-197. [doi: 10.1111/j.1365-2923.2007.02973.x] [Medline: 18230092]
13. Bienstock JL, Katz NT, Cox SM, et al. To the point: medical education reviews--providing feedback. Am J Obstet Gynecol 2007 Jun;196(6):508-513. [doi: 10.1016/j.ajog.2006.08.021] [Medline: 17547874]
14. Kember D, Leung DYP, Kwan KP. Does the use of student feedback questionnaires improve the overall quality of teaching? Assessment & Evaluation in Higher Education 2002 Sep;27(5):411-425. [doi: 10.1080/0260293022000009294]
15. Boerboom TBB, Jaarsma D, Dolmans DHJM, Scherpbier AJJA, Mastenbroek NJJM, Van Beukelen P. Peer group reflection helps clinical teachers to critically reflect on their teaching. Med Teach 2011;33(11):e615-e623. [doi: 10.3109/0142159X.2011.610840] [Medline: 22022915]
16. Hajhamid B, Somogyi-Ganss E. Improving effectiveness of dental students' feedback and course evaluation. J Dent Educ 2021 Jun;85(6):794-801. [doi: 10.1002/jdd.12548] [Medline: 33502807]
17. Arreola RA. Developing a Comprehensive Faculty Evaluation System: Magna Publications; 2004.
18. Gormally C, Evans M, Brickman P. Feedback about teaching in higher ed: neglected opportunities to promote change. CBE Life Sci Educ 2014;13(2):187-199. [doi: 10.1187/cbe.13-12-0235] [Medline: 26086652]
19. Smither JW, London M, Reilly RR. Does performance improve following multisource feedback? A Theoretical model, meta - analysis, and review of empirical findings. Pers Psychol 2005 Mar;58(1):33-66. [doi: 10.1111/j.1744-6570.2005.514_1.x]
20. Overeem K, Wollersheim H, Driessen E, et al. Doctors' perceptions of why 360-degree feedback does (not) work: a qualitative study. Med Educ 2009 Sep;43(9):874-882. [doi: 10.1111/j.1365-2923.2009.03439.x] [Medline: 19709012]
21. Ward JR, McCotter SS. Reflection as a visible outcome for preservice teachers. Teac Teacher Educ 2004 Apr;20(3):243-257. [doi: 10.1016/j.tate.2004.02.004]
22. Watts M, Lawson M. Using a meta-analysis activity to make critical reflection explicit in teacher education. Teach Teacher Educ 2009 Jul;25(5):609-616. [doi: 10.1016/j.tate.2008.11.019]
23. Schneider G. Student evaluations, grade inflation and pluralistic teaching: moving from customer satisfaction to student learning and critical thinking. In: Forum for Social Economics: Taylor & Francis; 2013.
24. McColskey W, Leary MR. Differential effects of norm-referenced and self-referenced feedback on performance expectancies, attributions, and motivation. Contemp Educ Psychol 1985 Jul;10(3):275-284. [doi: 10.1016/0361-476X(85)90024-4]

XSL•FO
RenderX

25. Mahler SA, Wolcott CJ, Swoboda TK, Wang H, Arnold TC. Techniques for teaching electrocardiogram interpretation: self-directed learning is less effective than a workshop or lecture. Med Educ 2011 Apr;45(4):347-353. [doi: 10.1111/j.1365-2923.2010.03891.x] [Medline: 21401682]
26. Haidet P, Morgan RO, O'Malley K, Moran BJ, Richards BF. A controlled trial of active versus passive learning strategies in A large group setting. Adv Health Sci Educ Theory Pract 2004;9(1):15-27. [doi: 10.1023/B:AHSE.0000012213.62043.45] [Medline: 14739758]
27. Chen CM, Wu CH. Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. Comput Educ 2015 Jan;80:108-121. [doi: 10.1016/j.compedu.2014.08.015]
28. Shqaidef AJ, Abu-Baker D, Al-Bitar ZB, Badran S, Hamdan AM. Academic performance of dental students: A randomised trial comparing live, audio recorded and video recorded lectures. Eur J Dent Educ 2021 May;25(2):377-384. [doi: 10.1111/eje.12614] [Medline: 33021047]
29. Vygotsky LS. In: Cole M, Jolm-Steiner V, Scribner S, Souberman E, editors. Mind in society: development of higher psychological processes: Harvard University Press; 1978.
30. Goleman D. Emotional intelligence, 10th edition: Bantam Books; 2007.
31. Douglas JA, Douglas A, McClelland RJ, Davies J. Understanding student satisfaction and dissatisfaction: an interpretive study in the UK higher education context. Stu High Educ 2015 Feb 7;40(2):329-349. [doi: 10.1080/03075079.2013.842217]
32. Zabaleta F. The use and misuse of student evaluations of teaching. Teach High Educ 2007 Feb;12(1):55-76. [doi: 10.1080/13562510601102131]
33. Richardson JTE. Instruments for obtaining student feedback: a review of the literature. Assess Evalu Higher Educ 2005 Aug;30(4):387-415. [doi: 10.1080/02602930500099193]
34. Pan D, Tan GSH, Ragupathi K, Booluck K, Roop R, Ip YK. Profiling teacher/teaching using descriptors derived from qualitative feedback: formative and summative applications. Res High Educ 2009 Feb;50(1):73-100. [doi: 10.1007/s11162-008-9109-4]
35. Spooren P, Brockx B, Mortelmans D. On the validity of student evaluation of teaching. Rev Educ Res 2013 Dec;83(4):598-642. [doi: 10.3102/0034654313496870]
36. Svinicki MD. Encouraging your students to give feedback. New Drctns for Teach & Learn 2001 Sep;2001(87):17-24 [FREE Full text] [doi: 10.1002/tl.24]
37. Lewis KG. Making sense of student written comments. New Drctns for Teach & Learn 2001 Sep;2001(87):25-32 [FREE Full text] [doi: 10.1002/tl.25]
38. Beckert L, Wilkinson TJ, Sainsbury R. A needs-based study and examination skills course improves students' performance. Med Educ 2003 May;37(5):424-428. [doi: 10.1046/j.1365-2923.2003.01499.x] [Medline: 12709183]
39. Karibe H, Suzuki A, Sekimoto T, et al. Cross-cultural comparison of the attitudes of dental students in three countries. J Dent Educ 2007 Nov;71(11):1457-1466. [doi: 10.1002/j.0022-0337.2007.71.11.tb04417.x] [Medline: 17971576]
40. Matthew IR, Walton JN, Dumaresq C, Sudmant W. The burden of debt for Canadian dental students: part 3. Student indebtedness, sources of funding and the influence of socioeconomic status on debt. J Can Dent Assoc 2006 Nov;72(9):819. [Medline: 17109801]
41. Ahmad MS, Bhayat A, Fadel HT, Mahrous MS. Comparing dental students' perceptions of their educational environment in Northwestern Saudi Arabia. Saudi Med J 2015 Apr;36(4):477-483. [doi: 10.15537/smj.2015.4.10754] [Medline: 25828286]
42. Ahmad M, Al Shorman H, Mahrous M. Assessment of the educational environment in a newly established dental college. J Educ Ethics Dent 2013;3(1):6. [doi: 10.4103/0974-7761.126935]

Abbreviations

KSAU-HS: King Saud bin Abdulaziz University for Health Sciences
KSAU-HS: King Saud bin Abdul-Aziz University

Edited by S Nedunchezhiyan, T Gladman; submitted 02.11.24; peer-reviewed by D Jocic, MZ Nassani; revised version received 07.02.25; accepted 11.02.25; published 18.03.25.

Please cite as:
Alreshaid L, Alkattan R
Feedback From Dental Students Using Two Alternate Coaching Methods: Qualitative Focus Group Study
JMIR Med Educ 2025;11:e68309
URL: https://mededu.jmir.org/2025/1/e68309
doi:10.2196/68309

XSL•FO
RenderX

© Lulwah Alreshaid, Rana Alkattan. Originally published in JMIR Medical Education (https://mededu.jmir.org), 18.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Citation Accuracy Challenges Posed by Large Language Models

Manlin Zhang[1*], MPhil; Tianyu Zhao[2*], MSc

[1]Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University, Shenyang, China

[2]Department of Science and Technology Studies, University College London, Gower Street, London, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Tianyu Zhao, MSc
Department of Science and Technology Studies, University College London, Gower Street, London, United Kingdom

Related Articles:

https://mededu.jmir.org/2025/1/e63400

Comment in: https://mededu.jmir.org/2025/1/e73698

(JMIR Med Educ 2025;11:e72998) doi:10.2196/72998

KEYWORDS

chatGPT; medical education; Saudi Arabia; perceptions; knowledge; medical students; faculty; chatbot; qualitative study; artificial intelligence; AI; AI-based tools; universities; thematic analysis; learning; satisfaction; LLM; large language model

Large language models (LLMs) such as DeepSeek, ChatGPT, and ChatGLM have significant limitations in generating citations, raising concerns about the quality and reliability of academic research. These models tend to produce citations that are correctly formatted but fictional in content, misleading users and undermining academic rigor. In the recent study titled "Perceptions and earliest experiences of medical students and faculty with ChatGPT in medical education: qualitative study," the section addressing concerns about ChatGPT deserves a deeper discussion [1].

There are several reasons for the citation issues in LLMs, which can be analyzed as follows. First, most LLMs cannot access paid subscription databases and therefore solely rely on open-access resources [2]. This limits the citations generated by LLMs to open-access journals, potentially omitting more significant research published in subscription-based journals. Second, LLMs are trained on vast amounts of text data and generate content by analyzing patterns and structures in text. However, they lack the ability to understand the content of the text or think critically, implying that they cannot judge the accuracy and reliability of information. Third, the algorithms underlying LLMs are often opaque, leaving users unable to understand the specific processes of information handling. This makes it difficult for users to determine the reliability of citations generated by LLMs and to effectively evaluate their results. Recent research also stated that half of generated search results lack citations, and only 75% of those with citations support the claims, posing trust concerns as user reliance grows[3].

Recently, an experiment conducted by the Journal of Clinical Anesthesia involved publishing a fictional article titled "Spinal Cord Ischemia After ESP Block" to test the spread and citation of a fabricated academic content. Surprisingly, the fictional article was widely cited, over 400 times, including in some journals with high impact factors[4], revealing a lack of rigor in academic citation practices, where many authors may not check the original literature and instead copy references directly. This incident sparked widespread discussion about academic citation practices, emphasizing the importance of critical thinking by scholars while citing materials.

The use of fictional citations by LLMs poses a multifaceted problem: it misleads users into drawing incorrect conclusions and making inappropriate decisions, undermines the rigor and credibility of academic research, and hinders the dissemination of knowledge by limiting access to accurate scientific information [5]. The issue of LLMs generating fictional citations is complex and requires the combined efforts of multiple stakeholders for resolution. Developers must continuously improve the LLM technology and algorithms, users must increase their awareness and critical evaluation skills while using LLMs, and academic institutions must strengthen the management and education in academic practices. Only through these efforts can we ensure that LLMs play a positive role in academic research and promote the dissemination and progress of knowledge.

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Abouammoh N, Alhasan K, Aljamaan F, et al. Perceptions and earliest experiences of medical students and faculty with ChatGPT in medical education: qualitative study. JMIR Med Educ 2025 Feb 20;11:e63400. [doi: 10.2196/63400] [Medline: 39977012]
2. Perianes-Rodríguez A, Olmeda-Gómez C. Effects of journal choice on the visibility of scientific publications: a comparison between subscription-based and full open access models. Scientometrics 2019 Dec;121(3):1737-1752. [doi: 10.1007/s11192-019-03265-y]
3. Peskoff D, Stewart B. Credible without credit: domain experts assess generative language models. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics 2023 Jul;2:427-438. [doi: 10.18653/v1/2023.acl-short.37]
4. Marcus A, Oransky I, De Cassai A. Please don't cite this editorial. J Clin Anesth 2025 Jan 8:111741. [doi: 10.1016/j.jclinane.2025.111741] [Medline: 39779384]
5. Rasul T, Nair S, Kalendra D, et al. The role of ChatGPT in higher education: benefits, challenges, and future research directions. JALT 2023 May 10;6(1):41-56. [doi: 10.37074/jalt.2023.6.1.29]

Abbreviations

LLM: large language model

Edited by S Nedunchezhiyan; submitted 23.02.25; this is a non–peer-reviewed article;accepted 12.03.25; published 02.04.25.

Please cite as:
Zhang M, Zhao T
Citation Accuracy Challenges Posed by Large Language Models
JMIR Med Educ 2025;11:e72998
URL: https://mededu.jmir.org/2025/1/e72998
doi:10.2196/72998

©Manlin Zhang, Tianyu Zhao. Originally published in JMIR Medical Education (https://mededu.jmir.org), 2.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Authors' Reply: Citation Accuracy Challenges Posed by Large Language Models

Mohamad-Hani Temsah[1], MD; Ayman Al-Eyadhy[1], MD; Amr Jamal[2], MBBS; Khalid Alhasan[1], MBBS; Khalid H Malki[3], PhD

[1]Pediatric Department, College of Medicine, King Saud University, King Abdullah Road, Riyadh, Saudi Arabia

[2]Department of Family and Community Medicine, King Saud University Medical City, Riyadh, Saudi Arabia

[3]Research Chair of Voice, Swallowing, and Communication Disorders, Department of Otolaryngology-Head and Neck Surgery, College of Medicine, King Saud University, Riyadh, Saudi Arabia

Corresponding Author:
Mohamad-Hani Temsah, MD
Pediatric Department, College of Medicine, King Saud University, King Abdullah Road, Riyadh, Saudi Arabia

Related Articles:

https://mededu.jmir.org/2025/1/e72998

https://mededu.jmir.org/2025/1/e63400

(JMIR Med Educ 2025;11:e73698) doi:10.2196/73698

KEYWORDS

ChatGPT; Gemini; DeepSeek; medical education; AI; artificial intelligence; Saudi Arabia; perceptions; medical students; faculty; LLM; chatbot; qualitative study; thematic analysis; satisfaction; RAG retrieval-augmented generation

We appreciate the thoughtful critique of our manuscript "Perceptions and earliest experiences of medical students and faculty with ChatGPT in medical education: qualitative study" [1] by Zhao and Zhang [2]. Concerns over the generation of hallucinated citations by large language models (LLMs), such as OpenAI's ChatGPT, Google's Gemini, and Hangzhou's DeepSeek, warrant exploring advanced and novel methodologies to ensure citation accuracy and overall output integrity [3].

The LLMs have demonstrated a propensity to generate well‐formatted yet fictitious references—a limitation largely attributed to restricted access to subscription-based databases and their reliance on probabilistic text generation [4]. As LLMs evolve, future iterations may integrate more reliable retrieval-based architectures, enhancing their capacity to cite legitimate sources while reducing fabricated references [4,5]. However, until such improvements are systematically validated, scholars must remain cautious.

One suggested enhancement is using retrieval-augmented generation (RAG) [6]. This approach integrates up-to-date external information, substantially improving real-world applicability. However, even RAG-based systems can misinterpret or distort source content under high-trust conditions. To address this, the authors developed Hallucination-Aware Tuning (HAT) [6]. HAT trains dedicated detection models to generate labels and detailed descriptions of identified hallucinations. These descriptions are then used by GPT-4 to correct discrepancies. The combination of corrected and original outputs forms a preference dataset that, when used for Direct Preference Optimization training, yields LLMs with reduced hallucination rates and improved answer quality [6].

We also propose another solution aimed at fundamentally reducing citation errors: the development of "Reference-Accurate" academic LLM by major global publishers. Leading journals could develop their own specialized LLM, trained exclusively on rigorously verified academic literature from robust databases. This targeted training would ensure that every generated reference is accurate and directly traceable to published work. Ideally, these publisher-backed LLMs would be made freely available to promote open science.

Therefore, we recommend a dual approach that combines advanced RAG methodologies with publisher-developed academic LLMs. Comparative studies should be conducted to evaluate the citation accuracy, factual consistency, and overall performance of RAG-HAT-tuned models against these publisher-specific models. Collaborative efforts among academic institutions, publishers, and AI developers are essential to establish standardized protocols and reliable training datasets. Such partnerships would not only enhance the reliability of LLM-generated outputs but also foster greater trust in AI-assisted scholarly communication.

Moreover, the broader academic community bears responsibility for critically appraising AI-generated content. While LLMs can streamline information retrieval and synthesis, human oversight remains indispensable for safeguarding academic integrity. Rather than dismissing AI-driven tools due to their current flaws, we advocate for further research to ensure greater alignment

XSL•FO
RenderX

with evidence-based scholarship and authentic publications. Future LLM iterations may rapidly overcome these limitations, but until then, transparency, responsible usage, and ongoing improvements in AI training remain imperative.

In conclusion, while RAG augmented by HAT represents a potential advancement in reducing hallucinations, the development of specialized, reference-accurate academic LLMs by publishers may offer a promising pathway. By integrating both strategies and ensuring human oversight, the academic community can ensure that AI-driven tools reliably support the rigor and transparency essential to scholarly research.

Conflicts of Interest

None declared.

References

1. Abouammoh N, Alhasan K, Aljamaan F, et al. Perceptions and earliest experiences of medical students and faculty with ChatGPT in medical education: qualitative study. JMIR Med Educ 2025 Feb 20;11:e63400. [doi: 10.2196/63400] [Medline: 39977012]
2. Zhang M, Zhao T. Citation accuracy challenges posed by large language models. JMIR Med Educ 2025 [FREE Full text] [doi: 10.2196/72998]
3. Temsah A, Alhasan K, Altamimi I, et al. DeepSeek in healthcare: revealing opportunities and steering challenges of a new open-source artificial intelligence frontier. Cureus 2025 Feb;17(2):e79221. [doi: 10.7759/cureus.79221] [Medline: 39974299]
4. Aljamaan F, Temsah MH, Altamimi I, et al. Reference hallucination score for medical artificial intelligence chatbots: development and usability study. JMIR Med Inform 2024 Jul 31;12:e54345. [doi: 10.2196/54345] [Medline: 39083799]
5. Howard A, Hope W, Gerada A. ChatGPT and antimicrobial advice: the end of the consulting infection doctor? Lancet Infect Dis 2023 Apr;23(4):405-406. [doi: 10.1016/S1473-3099(23)00113-5] [Medline: 36822213]
6. Song J, Wang X, Zhu J, et al. RAG-HAT: a hallucination-aware tuning pipeline for LLM in retrieval-augmented generation. 2024 Presented at: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing; Miami, Florida, US. [doi: 10.18653/v1/2024.emnlp-industry.113]

Abbreviations

HAT: Hallucination-Aware Tuning
LLM: large language model
RAG: retrieval-augmented generation

Edited by S Nedunchezhiyan; submitted 10.03.25; this is a non–peer-reviewed article;accepted 12.03.25; published 02.04.25.

Please cite as:
Temsah MH, Al-Eyadhy A, Jamal A, Alhasan K, Malki KH
Authors' Reply: Citation Accuracy Challenges Posed by Large Language Models
JMIR Med Educ 2025;11:e73698
URL: https://mededu.jmir.org/2025/1/e73698
doi:10.2196/73698

© Mohamad-Hani Temsah, Ayman Al-Eyadhy, Amr Jamal, Khalid Alhasan, Khalid H Malki. Originally published in JMIR Medical Education (https://mededu.jmir.org), 2.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Examining Multimodal AI Resources in Medical Education: The Role of Immersion, Motivation, and Fidelity in AI Narrative Learning

Chris Jacobs, MB BChir, BSc, MRes, MD(Res)

Department of Psychology, University of Bath, Claverton Down, Bath, United Kingdom

Corresponding Author:
Chris Jacobs, MB BChir, BSc, MRes, MD(Res)
Department of Psychology, University of Bath, Claverton Down, Bath, United Kingdom

Related Articles:

https://mededu.jmir.org/2025/1/e63865

Comment in: https://mededu.jmir.org/2025/1/e72336

(JMIR Med Educ 2025;11:e72190) doi:10.2196/72190

KEYWORDS

artificial intelligence; cinematic clinical narrative; cinemeducation; medical education; narrative learning; AI; medical students; preclinical education; long-term retention; pharmacology; AI tools; GPT-4; image; applicability; CCN

I read with great interest the recent study by Bland [1], “Enhancing Medical Student Engagement Through Cinematic Clinical Narratives: Multimodal Generative AI–Based Mixed Methods Study,” which explored the use of cinematic clinical narratives (CCNs) in medical education. The findings highlighted the potential of multimodal generative artificial intelligence (AI) to enhance engagement and knowledge retention among medical students. While the study effectively demonstrated novel use of AI to modernize case learning, further exploration of engagement mechanisms and broader learning theories may deepen our understanding of how these approaches can be optimized for educational impact.

Engagement in learning is multifaceted and can be linked to immersion [2] and intrinsic motivation [3]. As Bland observed, students exhibited heightened situational interest in CCNs, reinforcing the idea that immersive learning environments can enhance attention and recall. However, beyond the constructivist learning theory discussed in the study, additional models could be considered to expand the theoretical framework for understanding these results. One such model is the Cognitive Affective Model of Immersive Learning framework [4], which emphasizes the interplay between representational fidelity, cognitive load, and technological mediation in shaping learner experiences. Exploring the Cognitive Affective Model of Immersive Learning and similar frameworks could provide a more nuanced perspective on how technology interacts with learner motivation and engagement.

Another area for further inquiry is the role of pretest and posttest methodologies in evaluating learning outcomes. As the study rightly acknowledges, medical education research is often limited to single posttest designs [5], which may not capture students’ initial levels of engagement or attitudes toward learning. Additionally, fidelity of experience is important for recall, as realistic and contextually accurate learning environments can enhance memory retention and application. Implementing pretest assessments could help quantify shifts in engagement, allowing researchers to distinguish between students who are inherently motivated and those whose interest is primarily triggered by the intervention itself. Such an approach could offer valuable insights into how CCNs influence different learner profiles.

Moreover, the conditions under which learning occurs significantly affect student engagement. The study employed a classroom-based intervention, which effectively bridged the gap between controlled laboratory settings and real-world educational environments. Future research might explore how CCNs perform across varied instructional contexts, including self-paced online learning and clinical simulation settings, to assess their adaptability and impact on different learning situations and scenarios.

Bland provides good evidence that multimodal AI resources hold promise in medical education, underscoring the need to adapt to the evolving preferences of modern medical students who increasingly turn to social media for information and engagement. Expanding this research to integrate additional learning theories and measures, with varied instructional settings will help establish best practices for how to use these innovative tools. I appreciate the author’s novel approach and am eager to see how future developments in this field will shape medical education.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Bland T. Enhancing medical student engagement through cinematic clinical narratives: multimodal generative AI–based mixed methods study. JMIR Med Educ 2025 Jan 6;11:e63865. [doi: 10.2196/63865] [Medline: 39791333]
2. Jacobs C, Maidwell-Smith A. Learning from 360-degree film in healthcare simulation: a mixed methods pilot. J Vis Commun Med 2022 Oct;45(4):223-233. [doi: 10.1080/17453054.2022.2097059] [Medline: 35938350]
3. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
4. Makransky G, Petersen GB. The Cognitive Affective Model of Immersive Learning (CAMIL): a theoretical research-based model of learning in immersive virtual reality. Educ Psychol Rev 2021 Sep;33(3):937-958. [doi: 10.1007/s10648-020-09586-2]
5. Scerbo MW, Calhoun AW, Hui J. Research and hypothesis testing: moving from theory to experiment. In: Nestel D, Hui J, Kunkler K, Scerbo MW, Calhoun AW, editors. Healthcare Simulation Research: A Practical Guide: Springer International Publishing; 2019:161-167. [doi: 10.1007/978-3-030-26837-4_22]

Abbreviations

AI: artificial intelligence
CCN: cinematic clinical narrative

Edited by S Nedunchezhiyan; submitted 05.02.25; this is a non–peer-reviewed article;accepted 14.02.25; published 18.03.25.

Please cite as:
Jacobs C
Examining Multimodal AI Resources in Medical Education: The Role of Immersion, Motivation, and Fidelity in AI Narrative Learning
JMIR Med Educ 2025;11:e72190
URL: https://mededu.jmir.org/2025/1/e72190
doi:10.2196/72190

© Chris Jacobs. Originally published in JMIR Medical Education (https://mededu.jmir.org), 18.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Enhancing AI-Driven Medical Translations: Considerations for Language Concordance

Stephanie Quon[1], BASc; Sarah Zhou[2], BSc

[1]Faculty of Medicine, University of British Columbia, 2194 Health Sciences Mall, Vancouver, BC, Canada
[2]Faculty of Science, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Stephanie Quon, BASc
Faculty of Medicine, University of British Columbia, 2194 Health Sciences Mall, Vancouver, BC, Canada

Related Articles:

Companion article: https://mededu.jmir.org/2024/1/e51435

https://mededu.jmir.org/2025/1/e71721

(JMIR Med Educ 2025;11:e70420) doi:10.2196/70420

KEYWORDS

letter to the editor; ChatGPT; AI; artificial intelligence; language; translation; health care disparity; natural language model; survey; patient education; accessibility; preference; human language; communication; language-concordant care

We commend the recent publication by Dzuali et al [1], which explored the application of ChatGPT for translating patient education materials into multiple languages. This important study highlights a critical area where artificial intelligence (AI) can potentially bridge gaps in language-concordant care. To further this research, we would like to raise several points to enrich the discussion and understanding of the findings.

The study demonstrates that while ChatGPT provides clinically usable translations for Spanish and Russian, its performance with Mandarin is suboptimal. This inconsistency raises important questions regarding the linguistic complexities and structural differences between English and Mandarin, which may hinder the accuracy and appropriateness of translations. Previous research has shown that the nuanced sentence structures and specialized terminology in Mandarin pose challenges for AI models such as ChatGPT, suggesting the need for more refined approaches when using AI for translation in linguistically distinct languages [2].

Being familiar with the Mandarin language, we have firsthand experience with the challenges that come with translating between languages with distinct linguistic structures. Mandarin, with its nuanced sentence structures and specialized terminology, presents difficulties for large language models such as ChatGPT. These challenges are compounded by differences in grammar, idiomatic expressions, and cultural contexts, which may lead to inaccuracies and misunderstandings in translations. Therefore, this study could provide additional insight into how cultural context influences translation quality. Mandarin, for example, involves not only linguistic precision but also an understanding of cultural nuances that could affect comprehension [3]. Future studies could explore how AI models such as ChatGPT are trained to account for these contextual factors to ensure culturally appropriate translations.

Another area for potential exploration in this study is the testing of alternative prompts and the impact they may have on translation quality. While the study focuses on a single translation prompt—“Translate this into <target language>”—the variability of AI-generated translations could be better evaluated through a variety of prompts. Utilizing multiple prompts could reveal a broader range of performance outcomes, especially for linguistically complex languages such as Mandarin and Russian. Other studies have shown that different AI prompts can produce vastly different results [4].

Lastly, the study heavily relies on the involvement of board-certified dermatologists for posttranslation review, which is applicable to the context of dermatology-related information, but may not fully address the extent of errors and misinformation. While human oversight is essential, the study could benefit from a more robust evaluation of how different levels of human intervention—such as linguistic experts or specialists in medical translation—might improve translation accuracy [5]. Future research should explore how different combinations of AI-generated translations and human review from varied sources could optimize clinical usability.

Overall, while ChatGPT shows promise for improving access to language-concordant patient education, further refinement and validation are required. This study is an important milestone in starting this discussion surrounding AI-translation in medical contexts, and we commend the authors for their valuable contribution to advancing the field. They clearly demonstrate a meticulous approach, thoughtful analysis, and commitment to improving patient care through innovative solutions.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Dzuali F, Seiger K, Novoa R, et al. ChatGPT may improve access to language-concordant care for patients with non-English language preferences. JMIR Med Educ 2024 Dec 10;10:e51435. [doi: 10.2196/51435] [Medline: 39657144]
2. Jiao W, et al. Is chatgpt A good translator? A preliminary study. arXiv. Preprint posted online on Oct 1, 2023. [doi: 10.48550/arXiv.2301.08745]
3. Duff P, et al. Learning Chinese: Linguistic, Sociocultural, and Narrative Perspectives: Walter de Gruyter; 2013, Vol. 5.
4. Oppenlaender J, Linder R, Silvennoinen J. Prompting AI art: an investigation into the creative skill of prompt engineering. Int J Hum Comput 2024:1-23. [doi: 10.1080/10447318.2024.2431761]
5. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]

Abbreviations

AI: artificial intelligence

Edited by T Leung; submitted 20.12.24; this is a non–peer-reviewed article;accepted 27.01.25; published 11.04.25.

Please cite as:
Quon S, Zhou S
Enhancing AI-Driven Medical Translations: Considerations for Language Concordance
JMIR Med Educ 2025;11:e70420
URL: https://mededu.jmir.org/2025/1/e70420
doi:10.2196/70420

© Stephanie Quon, Sarah Zhou. Originally published in JMIR Medical Education (https://mededu.jmir.org), 11.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Author's Reply: Examining Multimodal AI Resources in Medical Education: The Role of Immersion, Motivation, and Fidelity in AI Narrative Learning

Tyler Bland, PhD

Department of Medical Education, University of Idaho, 875 Perimeter Drive MS 4061, Moscow, ID, United States

Corresponding Author:
Tyler Bland, PhD
Department of Medical Education, University of Idaho, 875 Perimeter Drive MS 4061, Moscow, ID, United States

Related Articles:

https://mededu.jmir.org/2025/1/e63865

https://mededu.jmir.org/2025/1/e72190

Abstract

(JMIR Med Educ 2025;11:e72336) doi:10.2196/72336

KEYWORDS

artificial intelligence; cinematic clinical narrative; cinemeducation; medical education; narrative learning; pharmacology; AI; medical students; preclinical education; long-term retention; AI tools; GPT-4; image; applicability; CCN

I extend my sincere appreciation for the thoughtful critique [1] of my study, "Enhancing Medical Student Engagement Through Cinematic Clinical Narratives: Multimodal Generative AI–Based Mixed Methods Study" [2]. The author's insights regarding engagement mechanisms, theoretical expansion, and methodological refinements offer valuable perspectives that contribute to the broader discourse on the pedagogical applications of generative artificial intelligence in medical education.

While the Cognitive Affective Model of Immersive Learning framework originated to explain learning with immersive virtual reality technologies [3], I concur that its underlying principles are applicable to my study. The debate over the role of media versus instructional methods in learning has been longstanding. While some argue that the medium itself shapes cognition, social structures, and cultural norms [4], others reject this notion, asserting that media are merely delivery mechanisms and that instructional methods alone drive learning outcomes [5]. The Cognitive Affective Model of Immersive Learning reframes this debate by emphasizing that it is not the medium (eg, immersive virtual reality) that inherently enhances learning, but rather how instructional methods leverage the unique affordances of that medium. In the context of cinematic clinical narratives (CCNs), the structured narrative and multimodal design capitalize on engagement mechanisms similar to those observed in immersive learning. Future research could further examine how instructional design within CCNs optimally harnesses these principles to promote knowledge retention and clinical application.

The author's recommendation of the integration of pretest and posttest methodologies is well-founded. While the published study employed posttest assessments to measure comprehension, incorporating pretest measures would facilitate a more granular evaluation of baseline knowledge and attitudinal shifts attributable to CCNs. Furthermore, longitudinal assessments could provide critical insights into the durability of knowledge retention and the sustained impact of CCNs over extended timeframes. I aim to incorporate these into future studies.

The author's call for broader contextual applications of CCNs beyond traditional classroom settings is well-taken. While the study examined CCN implementation within a structured learning environment, I am currently working on converting CCNs into self-contained short films that can be viewed online for self-directed learning. This adaptation aims to provide learners with greater flexibility while maintaining the engagement and narrative-driven structure of CCNs. Investigating how these self-contained films perform across varied instructional modalities could yield valuable insights into their scalability and applicability within diverse educational contexts.

Finally, I concur with the author's observation that medical students are increasingly turning to digital platforms such as social media for information and engagement. Medical educators should take note and examine the factors that make these

XSL•FO
RenderX

platforms so compelling. By understanding the draw of these digital environments, educators can incorporate similar characteristics into medical school learning materials to meet students where they are. Expanding CCN research to explore how elements such as interactivity, brevity, and personalization influence learner engagement could provide valuable insights into modernizing medical education. I am grateful for the astute observations and constructive recommendations by the author. These perspectives will undoubtedly inform my future research directions and further the integration of artificial intelligence–driven methodologies in my studies on medical education.

Conflicts of Interest

None declared.

References

1. Jacobs C. Examining Multimodal AI Resources in Medical Education: The Role of Immersion, Motivation, and Fidelity in AI Narrative Learning. JMIR Med Educ 2025;11:e72190. [doi: 10.2196/72190]
2. Bland T. Enhancing medical student engagement through cinematic clinical narratives: multimodal generative AI–based mixed methods study. JMIR Med Educ 2025 Jan 6;11(1):e63865. [doi: 10.2196/63865] [Medline: 39791333]
3. Makransky G, Petersen GB. The Cognitive Affective Model of Immersive Learning (CAMIL): a theoretical research-based model of learning in immersive virtual reality. Educ Psychol Rev 2021 Sep;33(3):937-958. [doi: 10.1007/s10648-020-09586-2]
4. McLuhan M. Understanding Media: The Extensions of Man: McGraw-Hill; 1964.
5. Clark RE. Media will never influence learning. ETR&D 1994 Jun;42(2):21-29. [doi: 10.1007/BF02299088]

Abbreviations

CCN: cinematic clinical narrative

Edited by S Nedunchezhiyan; submitted 07.02.25; this is a non–peer-reviewed article;accepted 14.02.25; published 18.03.25.

Please cite as:
Bland T
Author's Reply: Examining Multimodal AI Resources in Medical Education: The Role of Immersion, Motivation, and Fidelity in AI Narrative Learning
JMIR Med Educ 2025;11:e72336
URL: https://mededu.jmir.org/2025/1/e72336
doi:10.2196/72336

© Tyler Bland. Originally published in JMIR Medical Education (https://mededu.jmir.org), 18.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Authors' Reply: Enhancing AI-Driven Medical Translations: Considerations for Language Concordance

Joyce Teng[1], MD, PhD; Roberto Andres Novoa[1,2*], MD; Maria Alexandrovna Aleshin[1*], MD; Jenna Lester[3*], MD; Kira Seiger[4*], MD, MBA; Fiatsogbe Dzuali[3*], MD; Roxana Daneshjou[1,5], MD, PhD

[1]Department of Dermatology, Stanford University, 700 Welch Rd, Stanford, United States
[2]Department of Pathology, Stanford University, Redwood City, CA, United States
[3]Department of Dermatology, University of California, San Francisco, CA, United States
[4]Department of Dermatology, University of Washington, Seattle, WA, United States
[5]Department of Biomedical Data Science, Stanford University, Stanford, CA, United States
[*]these authors contributed equally

Corresponding Author:
Joyce Teng, MD, PhD
Department of Dermatology, Stanford University, 700 Welch Rd, Stanford, United States

Related Articles:

https://mededu.jmir.org/2024/1/e51435

Comment in: https://mededu.jmir.org/2025/1/e70420

(JMIR Med Educ 2025;11:e71721) doi:10.2196/71721

KEYWORDS

ChatGPT; artificial intelligence; language; translation; health care disparity; natural language model; survey; patient education; accessibility; preference; human language; communication; language-concordant care

We appreciate the thoughtful insights shared by Quon and Zhou [1] regarding our study on the application of ChatGPT in translating patient education materials [2]. We wholly agree that the linguistically distinct languages, such as Mandarin, can present challenges in capturing all the nuances and achieving precise translations.

In response to the comment regarding the use of multiple prompts, we acknowledge the complexity and variability in artificial intelligence (AI)-generated translations. However, it is important to consider the practical limitations within a clinical setting. Asking providers to use various prompts in real time may not be feasible due to time constraints and the need for efficiency in patient care. We believe that focusing on a single, effective prompt can streamline the translation process while we explore avenues for improvement in the AI's capabilities. This could be a productive avenue for future research.

Addressing the concern regarding the reliance on board-certified dermatologists for post-translation review, we want to clarify that, in addition to being board-certified dermatologists, all reviewers were native speakers in the language they reviewed, including fluency in Mandarin at a college level. This proficiency allows for a confluence of both clinical and linguistic insights when evaluating translations, reinforcing the validity of our findings. We appreciate the importance of rigor in translation review and remain committed to enhancing the integrity of our translated materials.

Overall, while we recognize the areas where ChatGPT can improve, we also see its current utility as a valuable tool for expanding access to language-concordant care in clinical settings. Our study serves as a helpful step toward identifying and addressing the limitations of AI translations, and we welcome continued dialogue to refine these practices.

Conflicts of Interest

None declared.

References

1. Quon S, Zhou S. Enhancing AI-driven medical translations: considerations for language concordance. JMIR Med Educ 2025;11. [doi: 10.2196/70420]
2. Dzuali F, Seiger K, Novoa R, et al. ChatGPT may improve access to language-concordant care for patients with non-English language preferences. JMIR Med Educ 2024 Dec 10;10:e51435. [doi: 10.2196/51435] [Medline: 39657144]

XSL•FO
RenderX

Abbreviations

AI: artificial intelligence

Edited by T Leung; submitted 24.01.25; this is a non–peer-reviewed article;accepted 27.01.25; published 11.04.25.

Please cite as:
Teng J, Novoa RA, Aleshin MA, Lester J, Seiger K, Dzuali F, Daneshjou R
Authors' Reply: Enhancing AI-Driven Medical Translations: Considerations for Language Concordance
JMIR Med Educ 2025;11:e71721
URL: https://mededu.jmir.org/2025/1/e71721
doi:10.2196/71721

© Joyce Teng, Roberto Andres Novoa, Maria Alexandrovna Aleshin, Jenna Lester, Kira Seiger, Fiatsogbe Dzuali, Roxana Daneshjou. Originally published in JMIR Medical Education (https://mededu.jmir.org), 11.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Refining Established Practices for Research Question Definition to Foster Interdisciplinary Research Skills in a Digital Age: Consensus Study With Nominal Group Technique

Jana Sedlakova[1,2], PhD; Mina Stanikić[1,2,3], MD, PhD; Felix Gille[1,2], PhD; Jürgen Bernard[1,4], Prof Dr; Andrea B Horn[1,5,6], Dr rer nat; Markus Wolf[1,6], Dr phil; Christina Haag[1,2,3], PhD; Joel Floris[1,7], Dr; Gabriela Morgenshtern[1,4], MSc; Gerold Schneider[1,8], Prof Dr; Aleksandra Zumbrunn Wojczyńska[1,9], Dr; Corine Mouton Dorey[1,10], Dr med, Dr phil; Dominik Alois Ettlin[1,9], PD, Dr med; Daniel Gero[1,11], PD, MD, PhD; Thomas Friemel[1,12], Prof Dr; Ziyuan Lu[1,7], Dr med; Kimon Papadopoulos[1,13], MPH; Sonja Schläpfer[1,14], MSc; Ning Wang[1], PhD; Viktor von Wyl[1,2,3], Prof Dr

[1]Digital Society Initiative, University of Zurich, Zurich, Switzerland
[2]Institute of Implementation Science in Healthcare, Faculty of Medicine, University of Zurich, Zurich, Switzerland
[3]Epidemiology, Biostatistics and Prevention Institute, Faculty of Medicine, University of Zurich, Zurich, Switzerland
[4]Department of Informatics, Faculty of Business, Economics and Informatics, University of Zurich, Zurich, Switzerland
[5]Center for Gerontology, University of Zurich, Zurich, Switzerland
[6]Department of Psychology, Faculty of Arts and Social Sciences, University of Zurich, Zurich, Switzerland
[7]Institute of Evolutionary Medicine, Faculty of Medicine, University of Zurich, Zurich, Switzerland
[8]Department of Computational Linguistics, Faculty of Business, Economics and Informatics, University of Zurich, Zurich, Switzerland
[9]Center of Dental Medicine, Faculty of Medicine, University of Zurich, Zurich, Switzerland
[10]Institute of Biomedical Ethics and History of Medicine, Faculty of Medicine, University of Zurich, Zurich, Switzerland
[11]Department of Surgery and Transplantation, University Hospital of Zurich, Zurich, Switzerland
[12]Department of Communication and Media Research, Faculty of Arts and Social Sciences, University of Zurich, Zurich, Switzerland
[13]Institute of Implementation Science in Healthcare, Faculy of Medicine, University of Zurich, Zurich, Switzerland
[14]Institute for Complementary and Integrative Medicine, University Hospital of Zurich, Zurich, Switzerland

Corresponding Author:
Jana Sedlakova, PhD
Digital Society Initiative
University of Zurich
Raemistrasse 69
Zurich, 8001
Switzerland
Phone: 41 0786753991
Email: sedlakova@ifi.uzh.ch

Abstract

Background: The increased use of digital data in health research demands interdisciplinary collaborations to address its methodological complexities and challenges. This often entails merging the linear deductive approach of health research with the explorative iterative approach of data science. However, there is a lack of structured teaching courses and guidance on how to effectively and constructively bridge different disciplines and research approaches.

Objective: This study aimed to provide a set of tools and recommendations designed to facilitate interdisciplinary education and collaboration. Target groups are lecturers who can use these tools to design interdisciplinary courses, supervisors who guide PhD and master's students in their interdisciplinary projects, and principal investigators who design and organize workshops to initiate and guide interdisciplinary projects.

Methods: Our study was conducted in 3 steps: (1) developing a common terminology, (2) identifying established workflows for research question formulation, and (3) examining adaptations of existing study workflows combining methods from health research and data science. We also formulated recommendations for a pragmatic implementation of our findings. We conducted

XSL•FO
RenderX

a literature search and organized 3 interdisciplinary expert workshops with researchers at the University of Zurich. For the workshops and the subsequent manuscript writing process, we adopted a consensus study methodology.

Results: We developed a set of tools to facilitate interdisciplinary education and collaboration. These tools focused on 2 key dimensions— content and curriculum and methods and teaching style—and can be applied in various educational and research settings. We developed a glossary to establish a shared understanding of common terminologies and concepts. We delineated the established study workflow for research question formulation, emphasizing the "what" and the "how," while summarizing the necessary tools to facilitate the process. We propose 3 clusters of contextual and methodological adaptations to this workflow to better integrate data science practices: (1) acknowledging real-life constraints and limitations in research scope; (2) allowing more iterative, data-driven approaches to research question formulation; and (3) strengthening research quality through reproducibility principles and adherence to the findable, accessible, interoperable, and reusable (FAIR) data principles.

Conclusions: Research question formulation remains a relevant and useful research step in projects using digital data. We recommend initiating new interdisciplinary collaborations by establishing terminologies as well as using the concepts of research tasks to foster a shared understanding. Our tools and recommendations can support academic educators in training health professionals and researchers for interdisciplinary digital health projects.

(JMIR Med Educ 2025;11:e56369) doi:10.2196/56369

KEYWORDS

research question; digitalization; digital data; data science; health research; interdisciplinary

Introduction

Background

Health research increasingly leverages the abundance of data from our "digital lives," including mobility data, social media data, or data from wearables [1,2]. Such digital data are commonly "unstructured" because it may not conform to a tabular format (eg, images, videos, sound, and free text) and often require specific expertise for harvesting; transforming; preprocessing; and creating meaningful insights into health, disease, and treatment [1,3-5]. Moreover, such digital data are often originally generated for nonresearch purposes and without addressing a specific research question [6]. In turn, they may lack standard quality attributes found in digital data collected for specific research purposes, such as depth, completeness, or consistency, which present methodological complexities to meaningfully use these data [1]. Therefore, reusing these digital unstructured data for health research requires diverse expertise, skills, and interdisciplinary collaboration between health domain experts (eg, clinicians and health scientists) and data scientists as method experts (eg, from data science, computer science, or statistics) [5,7,8].

Such interdisciplinary collaborations are often faced with challenges due to the seemingly conflicting research approaches between the disciplines. In addition to differences in terminologies and concept definitions, the prevailing emphasis of linear deductive approaches in health research contrasts with the often more explorative and iterative approaches used in data science [7]. In health research, it is customary to predefine key elements of the scientific process, including a research question and related hypothesis, in a protocol and scientific report (eg, STROBE [Strengthening the Reporting of Observational Studies in Epidemiology] or PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses] guidelines) [9-12]. These standard practices are deeply influenced by the tradition of clinical trials and treatment development, which place a strong emphasis on measurement validity, robustness, scientific rigor, and safety [13], as errors in study conduct or treatment could place study participants at risk. By contrast, data science generally tends to emphasize exploration, pattern discovery, or hypothesis generation as well as more iterative and inductive analysis approaches [1,14]. Some health researchers may perceive this greater emphasis on iterative approaches as lacking scientific rigor or focus on specific research questions.

For young researchers, interdisciplinary digital health collaborations might be particularly challenging because they need to balance traditional scientific methods with more iterative data-driven techniques. This dual demand highlights the importance of fostering interdisciplinary skills in education, enabling students to balance the rigorous demands of hypothesis-driven research with the iterative and inductive approaches of data science. Addressing these complexities represents an educational challenge for both established and young researchers.

Despite broad recognition of their importance, both practical and teaching or educational guidance on how to manage and overcome the challenges of interdisciplinary digital health collaborations are scarce. Such guidance is also important for educational purposes to foster skills for interdisciplinary collaboration among both young and established researchers as well as health professionals. Continuous education for experienced researchers is equally important to keep them updated with evolving methods and foster effective collaboration across disciplines.

This Study

To address this need, our study focuses on skill development to successfully navigate interdisciplinary collaborations and education in health-related research fields. We reviewed established workflows for research question formulation and investigated whether and how established workflows in health research may require adaptations to accommodate inductive and exploratory data science practices and novel analysis techniques. The study findings were translated into a set of tools and recommendations designed to facilitate interdisciplinary education and collaboration. These tools focus on 2 key

XSL•FO
RenderX

dimensions—*content and curriculum* and *methods and teaching style*—and can be applied in various educational and research settings. Lecturers can use them to design interdisciplinary courses, supervisors can guide PhD and master's students in their interdisciplinary projects, and principal investigators can design and organize workshops to initiate and guide interdisciplinary projects. By implementing these tools, educators and researchers can create more cohesive and productive educational resources for interdisciplinary collaborations. In the following sections, we offer our insights and a more detailed outline of how our study findings can inform both the *content* and *methods* dimensions, using an existing interdisciplinary course as an example.

The aims and findings of our study are intended to be globally relevant and applicable to all researchers using digital data in the context of health research and health care. Importantly, they also provide academic educators with a clear workflow and practical recommendations for discussing and addressing the challenges of interdisciplinary collaboration. As the focus is on research question formulation, a fundamental aspect of the research process, these recommendations are especially valuable for educational purposes, helping educators guide researchers and students through this essential phase of research projects.

To achieve our aims, we chose a consensus study approach that is appropriate to harmonize and bridge insights from experts from diverse research disciplines. Moreover, we focused our effort on the different approaches of research question formulation as the guiding example for this study because it represents a central step in guiding the research process and subsequent study design decisions. This process also served as an illustrative example to highlight the differences in research approaches between health research and data science.

Methods

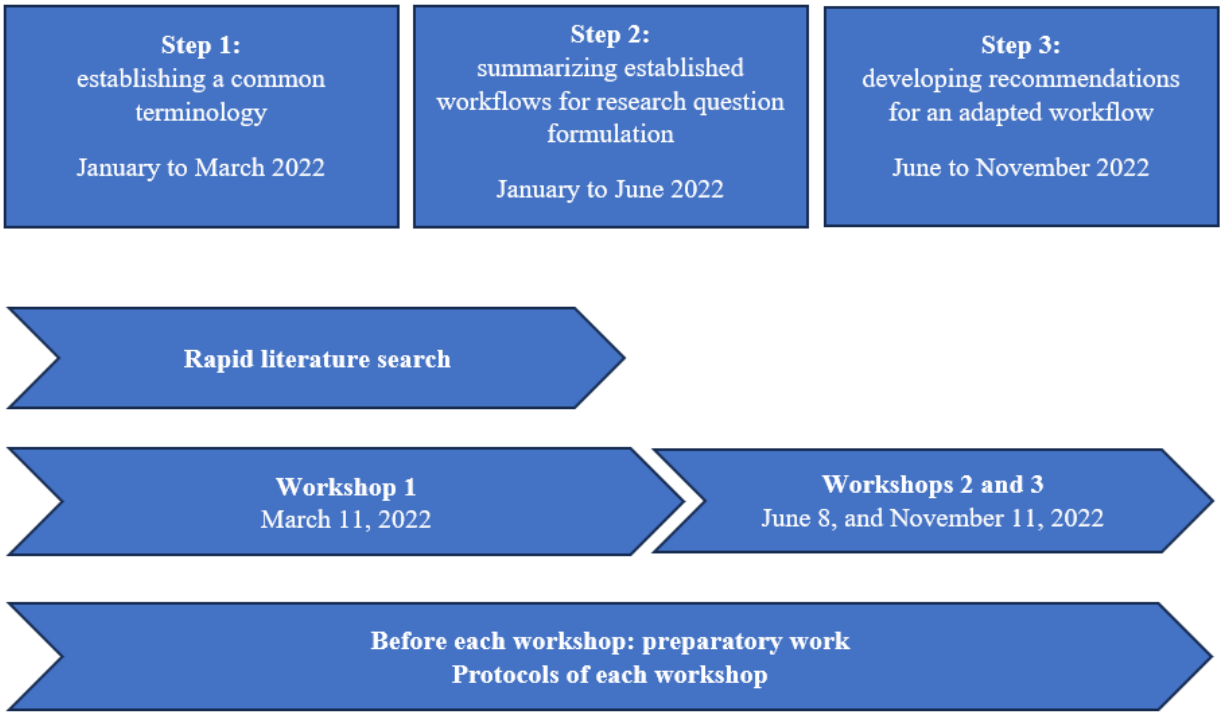
Consensus Methods

We used the nominal group technique with expert groups to gather insights from a diverse range of experts. This approach aimed to foster interdisciplinary skills and knowledge and achieve consensus on adapting research question development.

This study was structured by the following three high-level steps (Figure 1):

1. To create a common terminology to facilitate interdisciplinary and transdisciplinary collaborations that are required for research projects reusing digital data (ie, repurposing data originally generated for nonresearch purposes)
2. To describe the "established workflow" for research question formulation in health research on the basis of existing literature
3. To formulate suggestions and recommendations for adapting the "established workflow"

Figure 1. Study flow.



To inform steps 1 and 2, a rapid literature review was performed to identify established concepts for defining a research question in health research and data science as well as in other fields (refer to the Preparatory Research: Literature Search for the "Established Workflow" and an Example Scenario section). Expert inputs were gathered in a series of three 1.5-hour expert workshops. To foster a focused discussion in the workshops, participants were asked to complete preworkshop tasks. These inputs were summarized by JS and VvW and presented at the beginning of each workshop to discuss potential disagreements and allow participants to explain or comment on their and others' inputs. The consensus and agreement of each objective

XSL•FO
RenderX

were reached by an iterative, deliberative process. This included expert inputs before workshops, discussions during the workshop, and finally, expert feedback on and approval of the consolidated findings. These findings were synthesized, formulated, and shared by JS and VvW after each workshop. Furthermore, each participant was actively involved in the manuscript writing. These methods facilitated the systematic collection of input from participants in group and individual settings, enabling a comprehensive understanding of experts' knowledge and consolidating diverse perspectives. Workshops were recorded after receiving consent from the team. Workshop minutes, including the results of the preworkshop tasks, were sent for approval to the expert group. When necessary, individual researchers were contacted after workshops for clarification on specific issues raised during the workshops. The documentation and reporting of the workshop and the Accurate Consensus Reporting Document (ACCORD) checklist [15] for the consensus methodology are available in Multimedia Appendices 1 and 2.

The first 2 steps were accomplished in workshops 1 and 2. Building on these results, a third workshop was dedicated to identifying the need for the adaptation of established research practices in the health field to streamline collaboration with data scientists and to better integrate and communicate the need for research principles and standards, including open science and reproducibility.

Participants

The consensus meetings in the form of expert groups were led by JS and VvW, who led a previous project focusing on challenges and best practices of digital data, which inspired this study. Furthermore, JS is a scientific manager of the scientific community whose members were recruited for the consensus exercise. VvW's expertise lies in epidemiology and digital health research, and JS's expertise is mainly in digital ethics considering health research and health care. The workshop participants were recruited by JS and VvW among the diverse members of the Digital Society Initiative (DSI) Health Community at the University of Zurich. The members from the DSI were selected because it is a competence center for digital transformation that fosters interdisciplinary collaborations and projects studying the interplay and implications of digital transformations in society. Participants were included if they had experience with projects using digital data or planned to be involved in such projects. The workshop was promoted on the DSI website, through newsletters, and via word-of-mouth within the community. A total of 21 researchers from different disciplines and from all career stages participated in the workshops. This number of participants enabled to have an expert group with sufficient diversity to foster discussions and include insights from diverse disciplines. Of the 21 researchers, 13 (62%) represented health research, 3 (14%) represented data science, and 7 (33%) represented the social sciences and humanities.

Preparatory Research: Literature Search for the "Established Workflow" and an Example Scenario

A rapid literature search was conducted to inform the planning of the workshops and to develop a project roadmap (by JS and VvW). To gather information on the established workflow for research question formulation (steps 1 and 2), we searched the literature for publications, reviews, and course guidelines written either in English or German in PubMed and Google Scholar databases (search terms are provided in Multimedia Appendix 1). The search was further complemented by retrieval of guidelines from universities in Switzerland, Germany, the United States, and the United Kingdom, for which we searched on selected university websites. In addition, coauthors contributed materials they were familiar with or had previously used for teaching or research purposes. On the basis of this literature, we proposed the initial model for the "established workflow" that combines existing well-established frameworks and practices. To guide the discussions of our workshops, we developed an example scenario of digital data reuse for health research, which was communicated to participants before the workshops (Multimedia Appendix 3).

Ethical Considerations

The study followed the recommended procedures of the ethics committee of the Medical Faculty of the University of Zurich by completing the Data Protection/Ethics Self-Assessment Tool and received an exempt status. Participants were informed about the study's scope and goals as well as the nature of their involvement. They provided consent before the workshop and were informed that they could withdraw from the study at any time without providing a reason. The participants did not receive any compensation for their participation. The only personal information collected for the study was sociodemographic data, which were anonymized.

Results

Establishing a Common Terminology

Anticipating that a lack of harmonization concerning terminologies and concepts may hinder an effective interdisciplinary workshop collaboration, we aimed to establish a shared understanding of common terminologies. To this end, the workshop leaders (JS and VvW) developed a glossary before the first workshop, which was discussed and further refined by collecting written feedback from the participants after workshop 1 (Table 1).

The workshop discussions concerning the glossary centered around discipline-specific interpretations of concepts such as "research task," "research objectives," "research aims," and "research goals," whose interpretations were dependent on the embedding in different research methodologies, such as qualitative or quantitative research approaches. A central discussion centered around the recognition of different "research tasks," that is, high-level research aims from a methodological viewpoint, including, for example, exploration, confirmation, prediction, methods development, or theory development. For prediction and classification tasks, participants mentioned 2 subcategories of analyses, which are supervised learning methods that rely on labeled data and outcomes and include the broad class of (multivariable) regression models. By contrast, unsupervised methods (eg, neural networks) aim to find new data structures and features without the need for prior labeling and are often developed in a less linear, inductive manner.

XSL•FO
RenderX

Table 1. Common terminology for interdisciplinary research projects using digital data.

Terms	Definitions
Confirmatory research	• Hypothesis-driven research, experimental research, or research aiming at testing and confirming a hypothesis in a broader context of a theory. This research is also referred to as hypothetico-deductive research in some disciplines.
Exploratory research	• Data-driven research that aims at exploring new patterns and associations to formulate hypotheses.
Hypothesis	• A tentative, hypothetical prediction of the nature and direction of relationships between sets of data, phrased as a declarative statement. It is an assumption about scientific laws, causation, or empirical regularities. A hypothesis should be testable or falsifiable. This refers to quantitative evidence-based health research.
Unstructured data	• Raw data that are not in a predefined structure (eg, tables) or data that may be structured but still require substantial preprocessing or feature extraction (eg, continuous sensor data).
Principles and criteria of good research and research practice	• A set of values and norms for good conduct of research, including validity, scientific integrity, objectivity, and ethical study conduct.
Research aim	• The research aim is the overall, general, and long-term intention of a research project. The research aim describes the "what" of the research—where we aspire to be at the end.
Research design	• Research design describes the general outline of data collection (eg, cross-sectional and longitudinal studies) and analytical methods (eg, randomization, observational, and with or without control group) to answer the RQ[a]. It describes the "how" of research.
Research objective	• The specific goal linked to a RQ [16].
Research problem	• The research problem describes the rationale for a study, for example, by highlighting the societal or medical needs. It describes the "why"—the specific needs a study wants to address.
RQ	• A clear and concise question determining the research aim, objective, design, methodology, data collection, and analysis. The RQ narrows the aim and objective of the research. The process of defining a good RQ is dynamic and iterative. The RQ is refined through the different steps of the research cycle. We define the RQ in the context of quantitative evidence-based health research.
Research task	• A research task describes a high-level classification of aims or tasks in research, including descriptive research, exploratory research, confirmatory research, prediction and classification, theory development, or methods development.
Reuse of digital data	• The process of harvesting, transforming, and using structured or unstructured digital information that was initially generated for purposes other than research.
Theory or model	• A systematic, structured explanation or representation of facts, phenomena, or processes that sets the ground for research design, formulation of hypotheses, and predictions.
Tools to specify RQ	• Frameworks and tools that facilitate the development of specific aspects of defining the RQ or study design.
Types of RQ	• The type of RQ determines the main approach for achieving the research aim. Usually, there is a difference between quantitative and qualitative RQs that reflect quantitative and qualitative approaches. • Quantitative approaches use statistical and mathematical methods to address precise questions, typically using a deductive approach with a strong emphasis on the framework and structure. • By contrast, qualitative approaches use, for example, open-ended responses, focus groups, and interview-based techniques and focus on individual experiences and singularities. It seeks to determine or discover a process or define experiences. RQs tend to be inductive, flexible, adaptable, and nondirectional [17].

[a]RQ: research question.

The group further discussed the central role of hypotheses and linear, highly structured research approaches in health research, for example, in confirmatory research tasks (confirming a hypothesis, eg, by use of randomized controlled experiments or trials) and, to some extent, research focusing on predictions tasks (developing prediction models or classifiers to predict future events or out-of-sample attributes). In health research, it is generally recommended that the development of prediction methods follows a protocol that includes careful selection of predictors and (external) validation of the final model [18]. At the same time, it was also pointed out that some quantitative research tasks, such as methods development (ie, the development and validation of data analysis methods) or exploratory research (ie, detection of patterns and associations

XSL•FO
RenderX

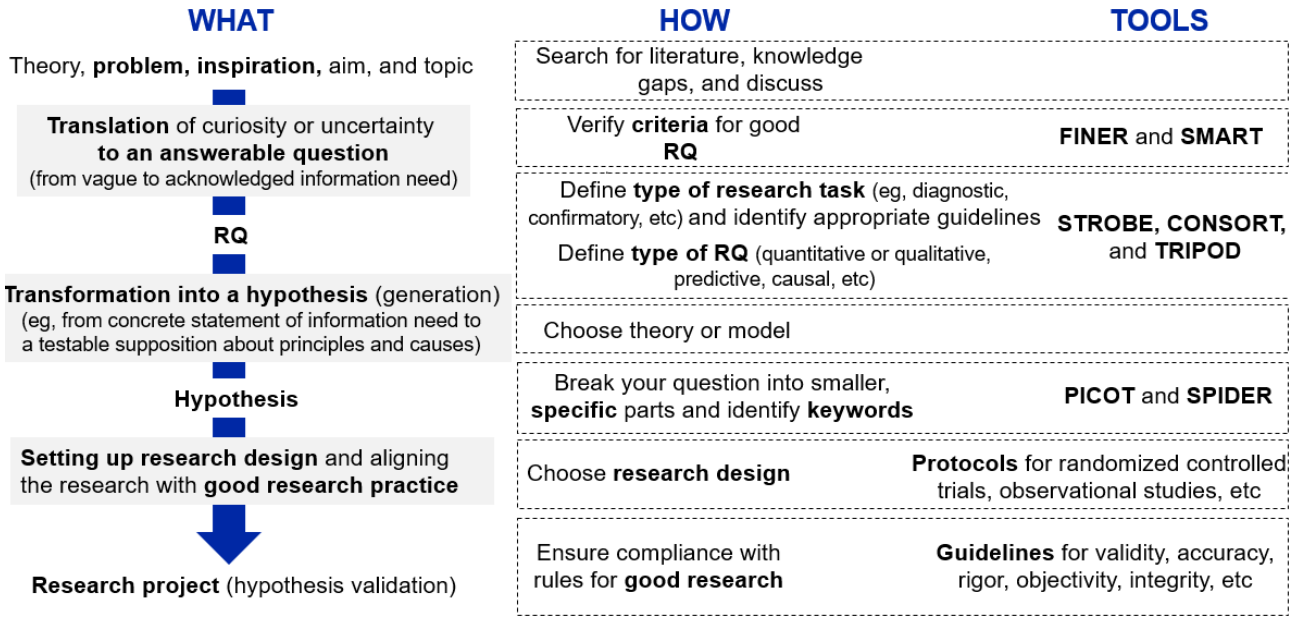
to generate new hypotheses), as well as qualitative research, generally depend much less on the specification of hypotheses. Workshop participants with a qualitative background argued that hypotheses can be implicitly involved in the research project. In qualitative research, it is common for the research question to evolve due to the necessity to critically reflect and adjust the study focus in each research step. As a result, the overall research process in qualitative research and some quantitative tasks, such as methods development, can be more iterative and dialectical when compared to deductive or confirmatory health research.

These discussions led to a key insight that interdisciplinary collaborations may be streamlined through the identification and discussion of the most appropriate "research task" early on, which can help guide subsequent discussions about the research question and the role of hypotheses in a common direction.

Summarizing Established Workflows for Research Question Formulation

The first 2 workshops were dedicated to better understanding how different disciplines approach the initial steps of a research project, including research question definition and study design choices. Informed by our rapid literature review, Figure 2 illustrates a summary workflow for established research design practices in health research. The vertical axis of Figure 2 illustrates the recommended steps for defining a research question (the "what"), starting from finding inspiration to developing a hypothesis, designing an appropriate study, and validating the hypothesis. Aligned with these definition steps, Figure 2 displays established practices (the "how") to execute the recommended steps. The third column references various frameworks and checklists aiding the implementation of each recommended step (the "tools").

Figure 2. Workflow of the recommended practices for defining a good research question. CONSORT: Consolidated Standards of Reporting Trials; FINER: Feasible, Interesting, Novel, Ethical, and Relevant; PICOT: Population, Intervention, Comparison, Outcome, and Time; RQ: research question; SMART: Specific, Measurable, Achievable, Realistic, and Timely; SPIDER: Sample, Phenomenon of Interest, Design, Evaluation, Research Type; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis.



The workflow of established practices generally starts with identifying a meaningful problem or question to be addressed in a study. The inspiration often emerges from a real-world challenge or knowledge gaps, but it can also be derived from existing theories or be triggered by discussions among colleagues. Workshop members also mentioned the influential role of funding criteria (ie, to increase chances for funding success) or topic-specific funding calls. This inspiration, curiosity, or uncertainty then needs to be translated into an answerable question [19-21]. Although we found little guidance in the literature on how to operationalize this step, it is often recommended to check the research question against the FINER (feasible, interesting, novel, ethical, and relevant) and SMART (specific, measurable, achievable, realistic, and timely) quality attributes to ensure its suitability for testing in a research study [16,20,22,23].

The wording of the research question itself may already imply a specific research task (eg, exploratory, confirmatory, or qualitative research) [21]. We differentiate between research aim and research objective. Research aim is the overall goal of the research, whereas research objective is the specific goal linked to a research question [16]. Having clarity on the research task will also facilitate the identification of appropriate reporting guidelines, such as STROBE (for observational studies), CONSORT (Consolidated Standards of Reporting Trials; for randomized controlled studies), or TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis; for the development of prediction models). These reporting guidelines primarily intend to guide the communication of study results but can also be useful in converting the research question into a study design.

Ultimately, decisions regarding the study design should be guided by the research question, while also considering practical

XSL•FO
RenderX

limitations and available means and resources [23]. Frameworks such as PICOT (population, intervention, comparison, outcome, and time) and similar tools (eg, SPIDER [sample, phenomenon of interest, design, evaluation, research type] for qualitative research) [9,20,24] provide useful starting points for defining the study design. We include both the FINER and PICOT tools and their equivalents to ensure the best possible quality of the research question. Some research studies have shown that using only PICOT might be suboptimal [23]. The PICOT framework is frequently applied in health research, as PICOT is already defined above [21]. Further high-level study design decisions concern the study duration and measurement frequency (longitudinal vs cross-sectional studies), the allocation of study participants into comparator groups (randomization vs "as is" in observational research), as well as numerous practical aspects concerning study execution (eg, sample size and methods of data collection and analysis) [21,24]. Study design and study execution also need to adhere to the principles and criteria of good research and research practice to achieve valid, reliable, and accurate results [25]. Moreover, the research must comply with the standards of objectivity, reproducibility, and research integrity [26,27].

Overall, the workshop discussions confirmed that the workflow summary (Figure 2) represents a useful starting point for interdisciplinary collaborations to illustrate the established practices and to explore conceptual differences between health research, data science, and other scientific disciplines.

Developing Recommendations for an Adapted Workflow

Building on the proposed example scenario for using digital data in health research (Multimedia Appendix 3) and the established workflow description from step 1, the participants then discussed 2 types of workflow adaptations to better reflect practices and approaches from data science (Figure 2). These included the following: (1) structural modifications by changing the sequence of workflow steps (ie, introducing additional steps that should become standard in a novel workflow—the "what") and (2) the need for introducing additional contextual constraints or novel quality criteria (ie, modifications that do not change the workflow but may impact their execution—the "how").

Modifications to the "What": The Steps in Research Questions Workflow

Overall, the workshop participants perceived that the general sequence of the established workflow (the "what") still applies to studies using (structured and unstructured) digital data. Complementary steps with their potential pitfalls were proposed to better reflect the additional challenges of working with digital unstructured data (Multimedia Appendix 4.

First, for unstructured data, preprocessing and feature extraction should be allocated a distinct workflow step to emphasize the need for thorough consideration during study planning and execution, to ensure that the data are usable, credible, and useful for the research question at hand [28-31]. On one hand, the assessment of data quality and validity is more challenging. On the other hand, preprocessing and feature extraction through machine learning require additional assumptions and may lead to predictions and derived parameters with uncertain distributional characteristics (eg, normal ranges) or propagation of algorithmic errors and biases.

Second, the selection of appropriate analysis methods to address the research question as a new workflow step would underscore the importance of scientific rigor [1,31-38]. For example, deciding between pretrained deep learning models requires preliminary investigations about the model features and the training database, which goes beyond the choice of more standard statistical techniques (eg, regression models) [39,40].

Finally, the general importance of efforts to render science reproducible and transparent was identified as a new step in the workflow.

Modifications of the "How" of the Research Question Workflow

The workshop group identified potential contextual and methodological changes to research practices (the "how"; Table 2). These proposed changes can be grouped into 3 clusters.

XSL•FO
RenderX

Table 2. Proposed changes to the "how" parts of the research question formulation workflow.

Change number	Contextual constraints and quality criteria	Description	Steps this is applicable to
I	Consider real-life incentives and constraints in defining research problems	• The decision about RQ[a] can be strongly influenced by other nonacademic factors such as the availability of funding or data.	RQ
II	Acknowledge feasibility and resource constraints	• The choice of research design and data analysis tools involves costs that must be considered, particularly to ensure compliance with scientific integrity.	Research design
III	Declare limitations in RQ scope	• Each RQ has limitations; it is important to define what RQ can and cannot answer.	RQ
IV	Allow for and document iterations in RQ development and analysis	• Proper documentation is important for ensuring transparency and helps with evaluating and tracking the decisions regarding the iterations in RQ.	RQ
V	Acknowledge and respond to the increasing need for interdisciplinary expertise	• For the feasibility of the RQ, it is important to consider the needed expertise and skills. This becomes particularly important in research involving digital unstructured data as it requires an interdisciplinary set of skills.	All steps
VI	Enhance reproducibility	• Reproducibility in data science means obtaining consistent results using the same input data and methods. On a higher level, reproducibility in science also refers to the ability to duplicate findings if the same methods are used [41]. • Reproducibility in science also refers to the concept of making data; computational steps, methods, and codes; and conditions of analysis transparent and available, so that others can verify the findings.	All steps
VII	Enhance replicability	• Replicability refers to applying the same methods from a different study on different data. Observed differences in findings should be explicable by data-specific differences between studies.	All steps
VIII	Enhance robustness	• Robustness refers to analyses that apply the same database but use different methods. Observed differences in findings should be explicable by method-specific differences between studies. Within the same study, robustness is often evaluated by sensitivity analyses that use the same data but vary methods (eg, by applying different model parameters).	All steps
IX	Critically assess generalizability	• Generalizability means that the study results or outcomes are also applicable in other study settings and samples.	All steps

[a]RQ: research question.

The first cluster includes the acknowledgment of what we have labeled as real-life constraints (change numbers I and II) and limitations in the scope of research questions (change number III). Appropriately addressing such real-life constraints can be fostered by greater transparency and experience exchange.

The second cluster of proposed contextual changes pertains to enabling more interdisciplinary and iterative workflows (change numbers IV and V). Reasons for iterative approaches include more complex choices of analysis methods, the need for verifying the validity and robustness of model results, or the need to manually search for the best model parametrization. These challenges also require a greater emphasis on interdisciplinary collaborations that combine subject-domain knowledge and data science expertise.

The final cluster reflects the need for strengthening research quality criteria to foster open science, better reproducibility, and greater transparency (change numbers VI-IX). As analytical methods and databases become more complex, there is also an increasing need for transparency; adequate documentation; as well as publicly available analysis protocols, software codes, data, and analysis files. Studies should critically examine their findings under changing data or method combinations, thus exploring reproducibility, robustness, replicability, or generalizability (the 3RG criteria) and enhancing the overall quality of research. An important means to achieve these goals are open science and Findable, Accessible, Interoperable, and Reusable (FAIR) data principles [42].

Recommendations Toward a Pragmatic Approach of Teaching and Conducting Research Question Formulation

The workshop discussions produced a set of specific recommendations to promote approaches for defining good

XSL•FO
RenderX

research questions for reusing digital data. These recommendations are also well suited for educational use, helping to navigate the challenges of interdisciplinary collaboration and to foster interdisciplinary skills.

Iterative Research Question Formulation

As a principle, data collection, preprocessing, and analysis methods should follow the research question; researchers should not lose sight of the research aim, objective, and question. Defining a good research question is a fundamental and universal first step of science, which ideally should not be preceded by the choice of data or methods. However, the linear process of defining a research question common to health research may need several iterations to ensure that the complexity and feasibility of reusing and integrating digital data are accounted for.

The lecture instructors, supervisors, and principal investigators of interdisciplinary projects can apply this recommendation by emphasizing the importance of research question formulation in interdisciplinary projects. Furthermore, they can facilitate a discussion or create exercises for students to practice how the linear process of research question definition changes into a more iterative process when collaborating with other disciplines.

Reconciling Linear And Iterative Approaches: Continuum of Research Tasks

To reconcile the apparent conceptual differences between health research and data science approach research projects, we propose to reframe the scientific process as a continuum of knowledge accumulation over the course of multiple studies. Such a continuum can consist of several different research tasks (projects) combining deductive and inductive research approaches. Not all research tasks will involve explicit research questions or hypotheses. However, systematic reflections on how study results can inform new hypotheses and research questions and how they could be tested in future studies could become an integral part of a study, for example, as a last step in exploratory analyses.

Lecture instructors, supervisors, and principal investigators of interdisciplinary projects can use this recommendation to emphasize the continuum of research tasks. They can create exercises consisting of different research tasks where students practice combining deductive and inductive approaches in research design. These exercises can guide students to recognize that research is not always a straightforward process of hypothesis testing but may involve exploratory tasks that inform future studies. Instructors can also encourage students to reflect systematically on their research results, guiding them to think about how current findings can shape future hypotheses and research directions. This reflection can be incorporated into project work, where students work on iterative research tasks, examining how knowledge accumulates across studies and how inductive and deductive methods interact throughout this process. This practice prepares students to handle the nonlinear nature of interdisciplinary research, especially when bridging health research and data science.

Research Quality Criteria

The complexities involved in digital data preprocessing and analysis require careful design decisions and thorough reporting to ensure adherence to research quality standards. The reuse of existing, digital unstructured data and the need for extensive preprocessing may obfuscate or compound issues of external and internal validity [14]. Moreover, the use of machine learning techniques such as deep neural networks may generate "unexplainable" predictions or classifications that challenge the transparency and open science paradigms. The verification of "whether the data measure what they are supposed to measure (in the context of the research question)" [14] remains crucial and deserves appropriate attention, but it may become more difficult to achieve. We recommend that researchers systematically scrutinize interim results to ensure that they are "on the good track." Such checks can, for example, include the replication of results from different studies. Furthermore, transparency in reporting and reproducibility are key to scientific rigor.

Lecture instructors and supervisors can emphasize the importance of maintaining research quality in interdisciplinary projects. They can design exercises where students practice making careful design decisions in their research projects, ensuring that issues of validity, transparency, and reproducibility are addressed throughout the process. One approach could be to guide students in developing protocols for systematic checks of their interim results. Instructors can also promote transparency by teaching students how to document their research processes thoroughly, facilitating reproducibility and open science principles. By applying these exercises, students learn to critically evaluate the quality of their research.

Take Active Measures to Foster Interdisciplinarity

We recommend reflecting these aspects appropriately in teaching and training of next-generation researchers as well as in establishing new interdisciplinary research groups or collaborations. Therefore, in teaching, it is important to also convey a realistic view of how research works in practice. Students should be sensitized to real-world challenges and the need for pragmatic decision-making, while still striving for the basic principles of "good research practices." The literature review and our own experiences suggest that students are mostly taught the "ideal model," and thus, they are often not well prepared for the realities of research. It seems preferable to discuss challenges openly and to expose students to ethical and practical dilemmas early on.

Lecture instructors and supervisors can sensitize students toward real-world challenges. They can prepare specific exercises where students can reflect on the problems that might arise from real-life constraints.

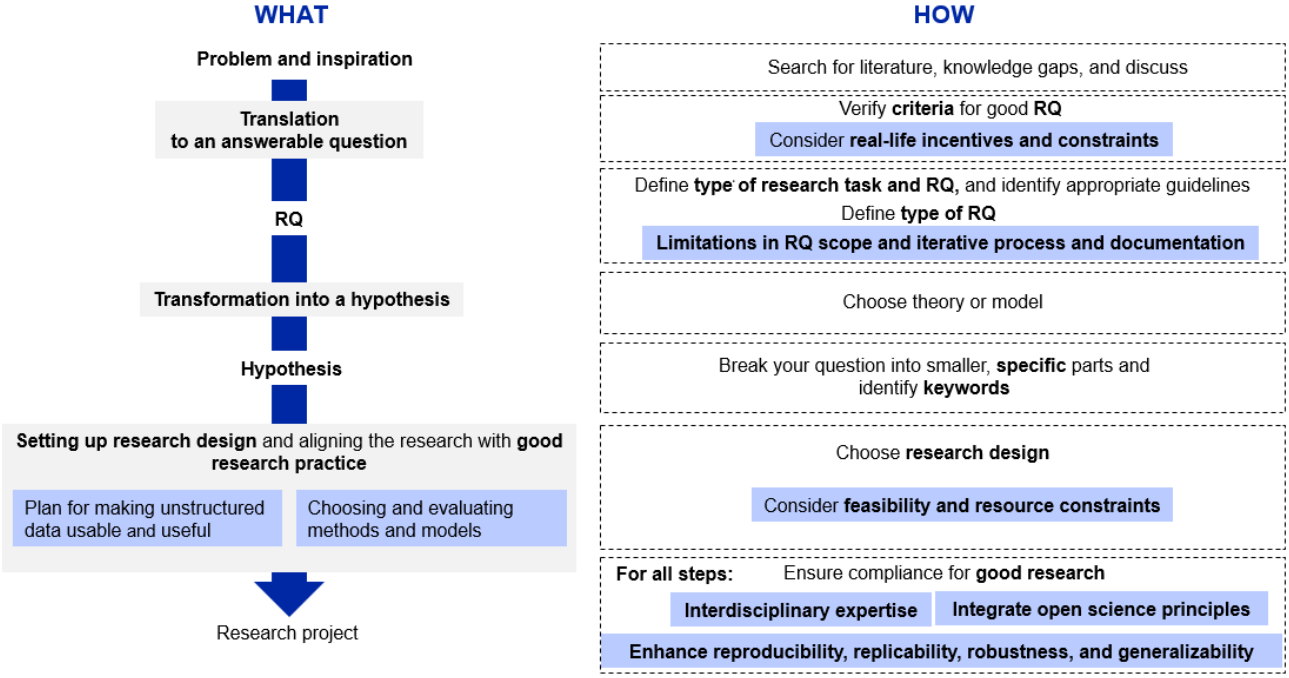
The added complexity and additional financial needs for education for interdisciplinary collaboration and open science should be acknowledged by funding agencies.

XSL•FO
RenderX

Specific Tools to Inform the Teaching of Interdisciplinary Courses on Real-World Data Analyses

Our study provides practical tools to guide the content and curricula of courses focused on interdisciplinary projects and collaborations. A more detailed description of the application of the study results to teaching is provided in Multimedia Appendix 5. The structure of our workshops (Figure 1) and the results of each workshop can be directly translated into the tools focusing on both dimensions of *content and curriculum* as well as *methods and teaching styles*. In terms of *content and curriculum*, the glossary with key concepts and terminology can be used to introduce students to interdisciplinary work. The workflow (Figure 2) combined with the glossary can serve as an interdiction to research practices for students with a different background, for example, humanities. Finally, our adapted workflow (Figure 3) sensitizes students for additional topics of transparency, FAIR data, reproducibility, and open science. Regarding *methods and teaching styles*, the sequence of workshops (Figure 1) and their results as outlined in the *content and curriculum section* can be directly translated into teaching phases, which build on top of each other. As illustrated by the example described in Multimedia Appendix 5, the structure of 3 teaching phases is useful and effective for teaching interdisciplinary research collaborations. A key insight from our workshop (modifications to the "how"—cluster 1) consisted of the need to acknowledge and address real-world challenges in study planning and execution. In our experience, case studies and illustrations of the scientific process of real-world examples are greatly appreciated by students.

Figure 3. Adapted workflow. The blue-colored boxes contain the added aspects in the "what" and "how" categories.



Discussion

Principal Findings

This study examined how interdisciplinary research collaborations between health research and data science can be streamlined by creating a shared conceptual understanding of terminologies and best practice workflows and by acknowledging or merging approaches from other disciplines. In a series of interactive workshops, our interdisciplinary group of coauthors concluded that the workflow of established practices for formulating a research question, generating hypotheses, and defining research designs remain valid. We argue that the reuse of digital data does not substantially change scientific activity, particularly the fundamental step of defining a good research question [1,43]. Achieving clarity on the research question benefits data analysis and interpretation by providing structure and informing the study design workflow. Moreover, a shared understanding of the research question and study workflow facilitates the inclusion of diverse domain knowledge to ensure research quality and result quality [6,14,44]. In line with this, the group noted general tendencies in research toward more open, transparent, and reproducible research, which are influenced by recommended data science practices. Along those lines, health research should increasingly foster good scientific practices that help to align the reuse of digital data with principles of reproducibility, robustness, generalizability, validity [1,32], transparency, and open science.

Our set of tools and recommendations can also be integrated into medical education by providing academic educators with a structured approach to teaching research question formulation in the context of using digital data in health research. By emphasizing the importance of both hypothesis-driven and data-driven research methods, educators can guide researchers in navigating the interdisciplinary challenges of health research and data science. The importance of creating a common terminology and discussion about scientific principles can further increase awareness about the challenges of interdisciplinary collaboration between health researchers and

XSL•FO
RenderX

data scientists. The proposed workflow and recommendations equip researchers with the tools to address the challenges of research question definition for interdisciplinary projects. The clear and practical steps provided by the workflow ensure that students not only grasp theoretical concepts but also apply them effectively in real-world scenarios, preparing them for collaborative, data-driven environments in health care and research.

For implementation, the set of tools and recommendations could be integrated into medical curricula and PhD programs through dedicated courses, workshops, or modules focusing on research methods and interdisciplinary collaboration for young researchers. Medical educators can adopt these recommendations to structure class discussions, assignments, and group projects, ensuring that students are exposed to both research approaches. In Multimedia Appendix 5, we provide an example of an interdisciplinary course implementing this set of tools. To evaluate the effectiveness of this implementation, a combination of qualitative and quantitative assessments can be used. Surveys and feedback from both students and educators can measure how well the workflow improves understanding and application of interdisciplinary research question formulation.

Our interdisciplinary effort recognized and discussed several potential obstacles toward bridging the approaches of established health research and data science. In the following sections, we repeat 4 key insights from our workshop interactions on how such obstacles can be overcome. First, we noted substantial differences in the use of terminologies across disciplines. For interdisciplinary collaborations, it is important to clarify key terms and concepts early on and to develop a shared understanding of the research aim and research question.

Second, in the early stages of the project, workshop participants expressed confusion about different types of analysis methods and their relationship with specific research tasks and high-level aims, such as prediction and classification, confirmatory research, or exploratory research. Agreeing on the high-level conceptual framework of “research tasks” helped structure the workshop discussions effectively. The discussion around the concepts of “research task” also fostered insights about commonalities and overlaps between concepts of data and health research. For example, many data science tasks can be classified as exploratory or prediction or classification tasks, which have conceptual counterparts in health research methodologies, each with corresponding reporting quality guidelines. Referring to specific research tasks rather than making global statements about data science or health research resonated well with the workshop participants and facilitated the discussions considerably.

Third, by introducing the concept of a “research task,” the group was also better able to examine the relationships among research aims, objectives, and tasks and how they are reflected in the workflow of established practices. Participants believed that exploratory or prediction or classification tasks, in particular, did not fit well into the workflow because such work is often not strictly hypothesis driven. However, 2 insights helped to align the workflow framework with the task concept: answering a research question may involve multiple research tasks in the same analysis, such as using prediction and classification tasks for data preprocessing, and later using these predictions in a confirmatory analysis, for example, as an exposure variable. Moreover, the scientific process can be viewed as a continuum of studies. From this perspective, the workflow of established practices can also be seen as a higher-level discovery cycle that spans across multiple studies. For example, an initial study may explore initial exploratory hypotheses or generate a first iteration of a prediction model, thus leading to new hypotheses. Indeed, exploratory and inductive methods can be useful to keep an open mind and become inspired by empirical data. In this way, the research tasks can be seen as a continuum—where data-driven research ends, hypothesis-driven research can start. Follow-up studies could then explore the hypotheses or validate the prediction model (whose structure can also be considered a hypothesis) in new data or in a confirmatory analysis. In combination, these multiple research tasks or study sequences are still likely to conform to the proposed workflow of recommended practices.

Finally, reusing unstructured and structured digital data brings new ethical challenges, such as privacy and consent issues, and problems with (public) trust and data diversity [45-49]. Traditional ethical assessments for data use in research and ethics review committees might not be well suited to address the challenges of digital data and might need adaptations [45,50]. Weighing the potential benefits and risks of using digital data becomes more complex. This problem is accentuated because the availability and production of digital data are often not based on a scientific decision, and rather, other factors such as political or social phenomena play a role [1]. While the need for novel ethical mechanisms to guide researchers is to be found in recently developed self-assessment tools for ethical data use [51,52], these new ethical mechanisms need further refinement to be widely adopted.

Strengths and Limitations

The strength of the expert groups was that participants represented a diverse group in terms of disciplines and career stages. However, it is possible that not all potentially relevant viewpoints were represented. A further strength was that the inputs from experts were collected systematically via different channels (eg, discussions, preworkshop tasks, and commenting on documents) throughout the consensus process. This allowed to harmonize and synthesize knowledge and insights from diverse disciplines. Experts also had several opportunities to review discussion outcomes and final summaries through workshop protocol and involvement in manuscript writing.

There are limitations regarding our proposed workflow. First, it represents an idealized process for defining a good research question, which is often challenged by funding and resource constraints or established norms. Some parts of the workflow might not be explicitly applicable to all types of research. The example scenario used to develop the workflow was based on hypothesis-driven deductive research, which often uses relational and causal research questions. We did not explicitly include inductive, qualitative approaches in health research, but we see the deductive and inductive research on a spectrum [53]. This limitation does not prevent the overall concept of the workflow

XSL•FO
RenderX

from being applied to other types of research, such as inductive, data-driven, or exploratory research.

Finally, although the literature review was conducted with great care and the expert group included several experienced researchers and faculty from different scientific disciplines, it was not possible to conduct a fully systematic search across all research disciplines due to resource constraints. Therefore, it is possible that some potentially relevant concepts and guidelines were not included.

Conclusions

In an age of digital transformation, established scientific practices with a strong focus on formulating research question design remain relevant and useful for gaining clarity about research aims. We recommend initiating new collaborations in the health domain with a review of terminologies and concepts to avoid misconceptions and problems further downstream in the research process. Our terminology and workflow may serve as tools to be used in medical education to support young and established researchers in interdisciplinary health research projects. To this end, we found the concept of "research tasks" particularly useful to foster a shared understanding among our collaborators. In addition, we recommend adapting the way the established workflow is taught to prospective researchers in health research and other disciplines, incorporating concepts from open science, the 3RG criteria, and the "science as a continuum" paradigm. We also call for funding agencies and publishers to incentivize and acknowledge investments in defining good research questions for complex novel data and analysis methods.

Data Availability

All data are available in the manuscript and multimedia appendices.

Authors' Contributions

JS contributed to conceptualization, data curation, investigation, methodology, project administration, resources, visualization, and writing the original draft. VvW contributed to the conceptualization, investigation, methodology, and writing the original draft. All the other authors contributed to writing, reviewing, and editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study and workshop protocols and preparatory tasks.
[PDF File (Adobe PDF File), 1792 KB - mededu_v11i1e56369_app1.pdf]

Multimedia Appendix 2
Accurate Consensus Reporting Document (ACCORD) checklist.
[PDF File (Adobe PDF File), 127 KB - mededu_v11i1e56369_app2.pdf]

Multimedia Appendix 3
Example scenario of digital data reuse to structure workshop discussions.
[PDF File (Adobe PDF File), 20 KB - mededu_v11i1e56369_app3.pdf]

Multimedia Appendix 4
Proposed modifications to the "what": additional steps to the workflow.
[PDF File (Adobe PDF File), 118 KB - mededu_v11i1e56369_app4.pdf]

Multimedia Appendix 5
Application of teaching tools: example from the course Interactive Data Science in Digital Health.
[DOCX File , 36 KB - mededu_v11i1e56369_app5.docx]

References

1. Caliebe A, Leverkus F, Antes G, Krawczak M. Does big data require a methodological change in medical research? BMC Med Res Methodol 2019 Jun 17;19(1):125 [FREE Full text] [doi: 10.1186/s12874-019-0774-0] [Medline: 31208367]
2. Chiavilli M, Campagnini S, Baretta T, Castagnoli C, Paperini A, Politi AM, et al. Design and implementation of a Stroke Rehabilitation Registry for the systematic assessment of processes and outcomes and the development of data-driven prediction models: the STRATEGY study protocol. Front Neurol 2022 Oct 10;13:919353 [FREE Full text] [doi: 10.3389/fneur.2022.919353] [Medline: 36299268]

XSL•FO
RenderX

3. Sim I. Mobile devices and health. N Engl J Med 2019 Sep 05;381(10):956-968. [doi: 10.1056/NEJMra1806949] [Medline: 31483966]
4. Batko K, Ślęzak A. The use of big data analytics in healthcare. J Big Data 2022;9(1):3 [FREE Full text] [doi: 10.1186/s40537-021-00553-4] [Medline: 35013701]
5. Sedlakova J, Daniore P, Horn Wintsch A, Wolf M, Stanikic M, Haag C, et al. Challenges and best practices for digital unstructured data enrichment in health research: a systematic narrative review. PLOS Digit Health 2023 Oct 11;2(10):e0000347 [FREE Full text] [doi: 10.1371/journal.pdig.0000347] [Medline: 37819910]
6. Kitchin R. Big data, new epistemologies and paradigm shifts. Big Data Soc 2014 Apr 01;1(1). [doi: 10.1177/2053951714528481]
7. Goldsmith J, Sun Y, Fried LP, Wing J, Miller GW, Berhane K. The emergence and future of public health data science. Public Health Rev 2021 Apr 26;42:1604023 [FREE Full text] [doi: 10.3389/phrs.2021.1604023] [Medline: 34692178]
8. Mirin N, Mattie H, Jackson L, Samad Z, Chunara R. Data science in public health: building next generation capacity. Harv Data Sci Rev 2022 Oct 27;4(4). [doi: 10.1162/99608f92.18da72db]
9. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. Qual Health Res 2012 Oct;22(10):1435-1443. [doi: 10.1177/1049732312452938] [Medline: 22829486]
10. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008] [Medline: 18313558]
11. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
12. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. PLoS Med 2007 Oct 16;4(10):e297 [FREE Full text] [doi: 10.1371/journal.pmed.0040297] [Medline: 17941715]
13. Donaldson L, Ricciardi W, Sheridan S, Tartaglia R. Textbook of Patient Safety and Clinical Risk Management. Cham, Switzerland: Springer International Publishing; 2020.
14. Hicks JL, Althoff T, Sosic R, Kuhar P, Bostjancic B, King AC, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. NPJ Digit Med 2019 Jun 3;2:45 [FREE Full text] [doi: 10.1038/s41746-019-0121-1] [Medline: 31304391]
15. Gattrell WT, Logullo P, van Zuuren EJ, Price A, Hughes EL, Blazey P, et al. ACCORD (ACcurate COnsensus Reporting Document): a reporting guideline for consensus methods in biomedicine developed via a modified Delphi. PLoS Med 2024 Jan 23;21(1):e1004326 [FREE Full text] [doi: 10.1371/journal.pmed.1004326] [Medline: 38261576]
16. Doody O, Bailey ME. Setting a research question, aim and objective. Nurse Res 2016 Mar;23(4):19-23. [doi: 10.7748/nr.23.4.19.s5] [Medline: 26997231]
17. Creswell JW. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. 3rd ed. Thousand Oaks, CA: SAGE Publications; 2012.
18. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. BMJ 2009 Mar 31;338:b604. [doi: 10.1136/bmj.b604] [Medline: 19336487]
19. Lipowski EE. Developing great research questions. Am J Health Syst Pharm 2008 Sep 01;65(17):1667-1670. [doi: 10.2146/ajhp070276] [Medline: 18714115]
20. Aslam S, Emmanuel P. Formulating a researchable question: a critical step for facilitating good clinical research. Indian J Sex Transm Dis AIDS 2010 Jan;31(1):47-50 [FREE Full text] [doi: 10.4103/0253-7184.69003] [Medline: 21808439]
21. Tully MP. Research: articulating questions, generating hypotheses, and choosing study designs. Can J Hosp Pharm 2014 Jan;67(1):31-34 [FREE Full text] [doi: 10.4212/cjhp.v67i1.1320] [Medline: 24634524]
22. Fandino W. Formulating a good research question: pearls and pitfalls. Indian J Anaesth 2019 Aug;63(8):611-616 [FREE Full text] [doi: 10.4103/ija.IJA_198_19] [Medline: 31462805]
23. Abbade LP, Wang M, Sriganesh K, Jin Y, Mbuagbaw L, Thabane L. The framing of research questions using the PICOT format in randomized controlled trials of venous ulcer disease is suboptimal: a systematic survey. Wound Repair Regen 2017 Sep;25(5):892-900. [doi: 10.1111/wrr.12592] [Medline: 29080311]
24. Booth A. Clear and present questions: formulating questions for evidence based practice. Libr Hi Tech 2006;24(3):355-368. [doi: 10.1108/07378830610692127]
25. Mehta A, Malley B, Walkey A. Formulating the research question. In: Secondary Analysis of Electronic Health Records. Cham, Switzerland: Springer; 2016.
26. Sørensen MP, Ravn T, Marušić A, Elizondo AR, Kavouras P, Tijdink J, et al. Strengthening research integrity: which topic areas should organisations focus on? Humanit Soc Sci Commun 2021 Aug 12;8(1):1-15. [doi: 10.1057/s41599-021-00874-y]
27. Quiroga Gutierrez AC, Lindegger DJ, Taji Heravi A, Stojanov T, Sykora M, Elayan S, et al. Reproducibility and scientific integrity of big data research in urban public health and digital epidemiology: a call to action. Int J Environ Res Public Health 2023 Jan 13;20(2):1473 [FREE Full text] [doi: 10.3390/ijerph20021473] [Medline: 36674225]

XSL•FO
RenderX

28. Kandel S, Heer J, Plaisant C, Kennedy J, van Ham F, Riche NH, et al. Research directions in data wrangling: visualizations and transformations for usable and credible data. Inf Vis 2011 Sep 02;10(4):271-288. [doi: 10.1177/1473871611415994]
29. Munzner T. Visualization Analysis and Design. Boca Raton, FL: CRC Press; 2014.
30. Fekete JD. Visual analytics infrastructures: from data management to exploration. Computer 2013 Jul;46(7):22-29. [doi: 10.1109/MC.2013.120]
31. Arbesser C, Spechtenhauser F, Muhlbacher T, Piringer H. Visplause: visual data quality assessment of many time series using plausibility checks. IEEE Trans Visual Comput Graphics 2017 Jan;23(1):641-650. [doi: 10.1109/tvcg.2016.2598592]
32. Cerreta F, Ritzhaupt A, Metcalfe T, Askin S, Duarte J, Berntgen M, et al. Digital technologies for medicines: shaping a framework for success. Nat Rev Drug Discov 2020 Sep;19(9):573-574. [doi: 10.1038/d41573-020-00080-6] [Medline: 32398879]
33. Sedlmair M, Meyer M, Munzner T. Design study methodology: reflections from the trenches and the stacks. IEEE Trans Visual Comput Graphics 2012 Dec;18(12):2431-2440. [doi: 10.1109/tvcg.2012.213]
34. Muhlbacher T, Piringer H, Gratzl S, Sedlmair M, Streit M. Opening the black box: strategies for increased user involvement in existing algorithm implementations. IEEE Trans Visual Comput Graphics 2014 Dec 31;20(12):1643-1652. [doi: 10.1109/tvcg.2014.2346578]
35. Schreck T, Bernard J, von Landesberger T, Kohlhammer J. Visual cluster analysis of trajectory data with interactive Kohonen maps. Inf Vis 2009 Feb 12;8(1):14-29. [doi: 10.1057/ivs.2008.29]
36. Lipton ZC. The doctor just won't accept that!. arXiv Preprint posted online on November 20, 2017 [FREE Full text] [doi: 10.48550/arXiv.1711.08037]
37. Bernard J, von Landesberger T, Bremm S, Schreck T. Multi-scale visual quality assessment for cluster analysis with self-organizing maps. In: Proceedings of SPIE - The International Society for Optical Engineering. 2011 Jan Presented at: SPIE 2011; 2011; Bellingham, WA. [doi: 10.1117/12.872545]
38. Munzner T. A nested model for visualization design and validation. IEEE Trans Vis Comput Graph 2009 Nov;15(6):921-928. [doi: 10.1109/TVCG.2009.111] [Medline: 19834155]
39. de Hond AA, Leeuwenberg AM, Hooft L, Kant IM, Nijman SW, van Os HJ, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. NPJ Digit Med 2022 Jan 10;5(1):2 [FREE Full text] [doi: 10.1038/s41746-021-00549-7] [Medline: 35013569]
40. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ 2022 May 18;377:e070904 [FREE Full text] [doi: 10.1136/bmj-2022-070904] [Medline: 35584845]
41. Schwab S, Janiaud P, Dayan M, Amrhein V, Panczak R, Palagi PM, et al. Ten simple rules for good research practice. PLoS Comput Biol 2022 Jun 23;18(6):e1010139 [FREE Full text] [doi: 10.1371/journal.pcbi.1010139] [Medline: 35737655]
42. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016 Mar 15;3(1):160018 [FREE Full text] [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
43. Mazzocchi F. Could big data be the end of theory in science? A few remarks on the epistemology of data-driven science. EMBO Rep 2015 Oct 10;16(10):1250-1255 [FREE Full text] [doi: 10.15252/embr.201541001] [Medline: 26358953]
44. Lee EW, Yee AZ. Toward data sense-making in digital health communication research: why theory matters in the age of big data. Front Commun 2020 Feb 27;5. [doi: 10.3389/fcomm.2020.00011]
45. Ferretti A, Ienca M, Sheehan M, Blasimme A, Dove ES, Farsides B, et al. Ethics review of big data research: what should stay and what should be reformed? BMC Med Ethics 2021 Apr 30;22(1):51 [FREE Full text] [doi: 10.1186/s12910-021-00616-4] [Medline: 33931049]
46. Nebeker C, Torous J, Bartlett Ellis RJ. Building the case for actionable ethics in digital health research supported by artificial intelligence. BMC Med 2019 Jul 17;17(1):137 [FREE Full text] [doi: 10.1186/s12916-019-1377-7] [Medline: 31311535]
47. Facca D, Smith MJ, Shelley J, Lizotte D, Donelle L. Exploring the ethical issues in research using digital data collection strategies with minors: a scoping review. PLoS One 2020;15(8):e0237875 [FREE Full text] [doi: 10.1371/journal.pone.0237875] [Medline: 32853218]
48. Clark RA, Foote J, Versace VL, Brown A, Daniel M, Coffee NT, et al. The keeping on track study: exploring the activity levels and utilization of healthcare services of acute coronary syndrome (ACS) patients in the first 30-days after discharge from hospital. Med Sci (Basel) 2019 Apr 19;7(4):61 [FREE Full text] [doi: 10.3390/medsci7040061] [Medline: 31010168]
49. Arora A, Alderman JE, Palmer J, Ganapathi S, Laws E, McCradden MD, et al. The value of standards for health datasets in artificial intelligence-based applications. Nat Med 2023 Nov 26;29(11):2929-2938 [FREE Full text] [doi: 10.1038/s41591-023-02608-w] [Medline: 37884627]
50. Viberg Johansson J, Bentzen HB, Mascalzoni D. What ethical approaches are used by scientists when sharing health data? An interview study. BMC Med Ethics 2022 Apr 11;23(1):41 [FREE Full text] [doi: 10.1186/s12910-022-00779-8] [Medline: 35410285]
51. Wiltshire D, Alvanides S. Ensuring the ethical use of big data: lessons from secure data access. Heliyon 2022 Feb;8(2):e08981 [FREE Full text] [doi: 10.1016/j.heliyon.2022.e08981] [Medline: 35243099]

XSL•FO
RenderX

52. Bigdata-ethics-HeALth framework - BEHALF. Eidgenössische Technische Hochschule Zürich. URL: https://bioethics.ethz.ch/research/BEHALF.html [accessed 2022-11-28]
53. Young M, Varpio L, Uijtdehaage S, Paradis E. The spectrum of inductive and deductive research approaches using quantitative and qualitative data. Acad Med 2020 Jul;95(7):1122. [doi: 10.1097/ACM.0000000000003101] [Medline: 31833855]

Abbreviations

3RG: reproducibility, robustness, replicability, or generalizability
ACCORD: Accurate Consensus Reporting Document
CONSORT: Consolidated Standards of Reporting Trials
DSI: Digital Society Initiative
FAIR: findable, accessible, interoperable, and reusable
FINER: feasible, interesting, novel, ethical, and relevant
PICOT: population, intervention, comparison, outcome, and time
SMART: specific, measurable, achievable, realistic, and timely
SPIDER: sample, phenomenon of interest, design, evaluation, research type
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis

Edited by B Lesselroth; submitted 15.01.24; peer-reviewed by G Richter, M Mahmic Kaknjo; comments to author 12.02.24; revised version received 16.05.24; accepted 23.11.24; published 23.01.25.

Please cite as:
Sedlakova J, Stanikić M, Gille F, Bernard J, Horn AB, Wolf M, Haag C, Floris J, Morgenshtern G, Schneider G, Zumbrunn Wojczyńska A, Mouton Dorey C, Ettlin DA, Gero D, Friemel T, Lu Z, Papadopoulos K, Schläpfer S, Wang N, von Wyl V
Refining Established Practices for Research Question Definition to Foster Interdisciplinary Research Skills in a Digital Age: Consensus Study With Nominal Group Technique
JMIR Med Educ 2025;11:e56369
URL: https://mededu.jmir.org/2025/1/e56369
doi:10.2196/56369
PMID:

©Jana Sedlakova, Mina Stanikić, Felix Gille, Jürgen Bernard, Andrea B Horn, Markus Wolf, Christina Haag, Joel Floris, Gabriela Morgenshtern, Gerold Schneider, Aleksandra Zumbrunn Wojczyńska, Corine Mouton Dorey, Dominik Alois Ettlin, Daniel Gero, Thomas Friemel, Ziyuan Lu, Kimon Papadopoulos, Sonja Schläpfer, Ning Wang, Viktor von Wyl. Originally published in JMIR Medical Education (https://mededu.jmir.org), 23.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Global Health care Professionals' Perceptions of Large Language Model Use In Practice: Cross-Sectional Survey Study

Ecem Ozkan[1], MD; Aysun Tekin[2], MD; Mahmut Can Ozkan[1], MD; Daniel Cabrera[3], MD; Alexander Niven[4], MD; Yue Dong[2], MD

[1]Department of Medicine, Jersey Shore University Medical Center, 1945 NJ-33, Neptune, NJ, United States
[2]Department of Anesthesiology, Mayo Clinic College of Medicine, Rochester, MN, United States
[3]Department of Emergency Medicine, Mayo Clinic College of Medicine, Rochester, MN, United States
[4]Department of Pulmonary and Critical Care Medicine, Mayo Clinic College of Medicine, Rochester, MN, United States

Corresponding Author:
Ecem Ozkan, MD
Department of Medicine, Jersey Shore University Medical Center, 1945 NJ-33, Neptune, NJ, United States

Abstract

Background: ChatGPT is a large language model-based chatbot developed by OpenAI. ChatGPT has many potential applications to health care, including enhanced diagnostic accuracy and efficiency, improved treatment planning, and better patient outcomes. However, health care professionals' perceptions of ChatGPT and similar artificial intelligence tools are not well known. Understanding these attitudes is important to inform the best approaches to exploring their use in medicine.

Objective: Our aim was to evaluate the health care professionals' awareness and perceptions regarding potential applications of ChatGPT in the medical field, including potential benefits and challenges of adoption.

Methods: We designed a 33-question online survey that was distributed among health care professionals via targeted emails and professional Twitter and LinkedIn accounts. The survey included a range of questions to define respondents' demographic characteristics, familiarity with ChatGPT, perceptions of this tool's usefulness and reliability, and opinions on its potential to improve patient care, research, and education efforts.

Results: One hundred and fifteen health care professionals from 21 countries responded to the survey, including physicians, nurses, researchers, and educators. Of these, 101 (87.8%) had heard of ChatGPT, mainly from peers, social media, and news, and 77 (76.2%) had used ChatGPT at least once. Participants found ChatGPT to be helpful for writing manuscripts (n=31, 45.6%), emails (n=25, 36.8%), and grants (n=12, 17.6%); accessing the latest research and evidence-based guidelines (n=21, 30.9%); providing suggestions on diagnosis or treatment (n=15, 22.1%); and improving patient communication (n=12, 17.6%). Respondents also felt that the ability of ChatGPT to access and summarize research articles (n=22, 46.8%), provide quick answers to clinical questions (n=15, 31.9%), and generate patient education materials (n=10, 21.3%) was helpful. However, there are concerns regarding the use of ChatGPT, for example, the accuracy of responses (n=14, 29.8%), limited applicability in specific practices (n=18, 38.3%), and legal and ethical considerations (n=6, 12.8%), mainly related to plagiarism or copyright violations. Participants stated that safety protocols such as data encryption (n=63, 62.4%) and access control (n=52, 51.5%) could assist in ensuring patient privacy and data security.

Conclusions: Our findings show that ChatGPT use is widespread among health care professionals in daily clinical, research, and educational activities. The majority of our participants found ChatGPT to be useful; however, there are concerns about patient privacy, data security, and its legal and ethical issues as well as the accuracy of its information. Further studies are required to understand the impact of ChatGPT and other large language models on clinical, educational, and research outcomes, and the concerns regarding its use must be addressed systematically and through appropriate methods.

(JMIR Med Educ 2025;11:e58801) doi:10.2196/58801

KEYWORDS

ChatGPT; LLM; global; health care professionals; large language model; language model; chatbot; AI; diagnostic accuracy; efficiency; treatment planning; patient outcome; patient care; survey; physicians; nurses; educators; patient communication; clinical; educational; utilization; artificial intelligence

XSL•FO
RenderX

Introduction

Large language model (LLM) refers to advanced artificial intelligence (AI) models designed for natural language processing tasks. LLMs are trained on vast amounts of text data and use deep learning techniques to understand and generate human-like language. They helped transform various fields, including medicine [1]. Some examples of most popular LLMs are LlaMA by Meta, Orca and Phi-1 by Microsoft, BLOOM, PaLM2 by Google, and GPT by OpenAI. ChatGPT, a chatbot powered by GPT-3/4 was released by OpenAI in November 2022, incorporating billions of parameters that enable it to comprehend and generate human-like text with the capability of context creation. Its intuitive interface and capacity for prompt engineering have enabled diverse applications across domains [2].

In medicine, recent studies have demonstrated ChatGPT's potential to support clinical decision-making, summarize complex medical data, and streamline documentation processes. For instance, ChatGPT has been evaluated for its ability to generate discharge summaries, assist in developing differential diagnoses, and simplify patient communication [3-5]. Its role in medical education has also been explored, demonstrating its utility in preparing students for licensing exams like the United States Medical Licensing Examination (USMLE) and enhancing self-directed learning through case-based scenarios [5-7]. ChatGPT was also shown to be capable of defining and answering clinical vignettes and achieved >60% of the threshold on the USMLE, which is the passing score for all three exams [8,9]. Additionally, its ability to provide personalized health education and assist in chronic disease management has been highlighted as a promising avenue for improving patient outcomes [4,10].

The integration of ChatGPT into health care settings is accelerating, with a growing body of literature examining its applications. Despite these advancements, significant challenges remain. Concerns about data privacy, ethical implications, and the accuracy of AI-generated content persist as barriers to widespread adoption [4,5,10]. Additionally, little is known regarding global health care professionals' perspectives and the extent and impact of ChatGPT's integration in health care settings [11,12]. Most studies to date, have been limited to localized settings or specific subgroups. Yet, successful and ethical integration of ChatGPT into health care workflows depends heavily on end-user acceptance, awareness of limitations, and perceptions regarding safety, usability, and value [5-7].

This study aimed to evaluate health care professionals' awareness and perceptions of ChatGPT, with a focus on its applications, challenges, and utility across clinical, educational, and research settings. We surveyed a diverse group of health care professionals—including physicians, nurses, researchers, and educators—from multiple countries and practice settings. Using a cross-sectional survey design, we collected data on their familiarity with ChatGPT, how and why they used it, and their concerns about its integration. Our a priori hypothesis was that while many health care professionals would recognize ChatGPT's potential benefits, such as improving efficiency, communication, and access to knowledge, they would also express concerns regarding ethical, legal, and accuracy-related issues.

This study offers timely insights for health care leaders, educators, and policymakers considering the responsible adoption of generative AI tools. By reflecting on global perspectives from frontline users, our findings may help shape discussions on how to balance innovation with safety and trust in clinical AI applications.

Methods

This study was conducted as a cross-sectional survey between April 20 and July 3, 2023 (Multimedia Appendix 1).

Survey Instrument Development and Validation

The questionnaire used in this study was developed de novo by the research team. The design process was informed by the research team's multidisciplinary experience in medicine, education, and digital health, as well as the evolving discourse around AI in health care. To assist with rapid prototyping, the research team used ChatGPT (OpenAI) to generate the first draft of the questionnaire. This initial draft provided a foundation for question phrasing and thematic organization. The final survey was iteratively refined by the study investigators to ensure clinical and contextual relevance.

To enhance clarity and assess feasibility, the questionnaire was piloted informally among five health care researchers affiliated with our institution. Their feedback informed improvements in question wording, branching logic, and estimated completion time (approximately 5 minutes). No formal psychometric validation was conducted.

The final survey included 33 questions and was distributed electronically using Research Electronic Data Capture (REDCap) (version 13.1.30; Vanderbilt University) [13]. The questionnaire was structured around six thematic domains: (1) respondent demographics and work environment, (2) awareness and familiarity with ChatGPT, (3) frequency and purpose of use, (4) perceived benefits and challenges of ChatGPT in daily practice, (5) views on ethical, legal, and data security concerns, and (6) future expectations and training needs. The questionnaire incorporated branching logic to adapt follow-up questions based on initial responses—for example, only respondents who reported using ChatGPT were asked about specific applications or frequency of use. A visual summary of the questionnaire flow and branching logic is provided in Multimedia Appendix 2. The final instrument has been reported in Multimedia Appendix 1.

Participants and Sampling Strategy

We used a convenience sampling approach. The questionnaire was distributed to health care professionals via targeted emails, and professional Twitter, LinkedIn, and Instagram accounts using a snowball technique [14]. No predefined inclusion or exclusion criteria were applied beyond the requirement of being a health care professional (eg, physician, nurse, educator, researcher). There were no regional or institutional restrictions.

XSL•FO
RenderX

As the survey was open and anonymous, we did not estimate a denominator or calculate a response rate. For the purposes of this study, we defined the application of ChatGPT in the medical field broadly to include its use in clinical care, research, medical education, and health care–related administrative tasks. This inclusive definition reflects the multifaceted roles that health care professionals fulfill and acknowledges that tools such as ChatGPT may support a wide range of activities beyond direct patient care, such as writing grants, academic correspondence, and synthesizing medical literature. Survey items were designed to capture this broad spectrum of use across domains relevant to daily professional practice.

Demographic information of participants was summarized. Among those familiar with ChatGPT, opinions on the tool and potential dissemination resources were assessed. For those who had not used it, barriers to usage were examined (Multimedia Appendix 2). Participants with experience using the ChatGPT were also asked about perceived challenges and approaches for enhancing usability. Summary statistics were provided as numbers and frequencies. Comparative analyses were conducted using the χ^2 test, with a two-sided P value <.05 considered statistically significant. JMP Pro (version 14.1.0 software; SAS Institute Inc.) was used for the analyses.

Ethical Considerations

The study protocol was evaluated by the Mayo Clinic institutional review board and it was determined that it was exempted under 45 CFR 46.102 of the Code of Federal Regulations (2/28/2023). No personally identifying information was collected, and all data were fully anonymous. Study participation was voluntary and survey completion was considered as consent. All survey responses were stored on secure, access-restricted servers in compliance with institutional data protection policies.

Results

Main Findings

A total of 115 health care professionals from 21 countries responded to the survey. Table 1 displays a summary of their demographic information (Figures 1–2).

XSL•FO
RenderX

Table . Baseline characteristics.

Variables	Participants (N=115), n (%)
Age (years)	
20 - 29	30 (26.1)
30 - 39	27 (23.5)
40 - 49	26 (22.6)
50 - 59	10 (8.7)
>60	22 (19.1)
Sex[a]	
Female	45 (39.5)
Male	68 (59.6)
Profession[a]	
Educator	16 (14.0)
NP/PA[b]	5 (4.4)
Physician	62 (54.4)
Researcher	25 (21.9)
RN[c]	5 (4.4)
Area/ Unit	
Internal medicine	20 (17.4)
Surgery	15 (13)
Emergency medicine	10 (8.7)
Psychiatry and Neurology	8 (7)
Anesthesiology/ICU[d]	10 (8.6)
Obstetrics and Gynecology	7 (6.1)
Radiology	6 (5.2)
Others[e]	39 (33.9)
Years since graduation	
<5	43 (37.4)
5 - 10	27 (23.5)
11 - 20	16 (13.9)
>20	29 (25.2)
Work length in hospital (years)[a]	
<5	66 (57.9)
5 - 10	11 (9.6)
11 - 20	19 (16.7)
>20	18 (15.8)
Country of work	
United States	53 (46.1)
Turkey	24 (20.9)
Tanzania	7 (6.1)
China	6 (5.2)
Croatia	3 (2.6)
Russia	2 (1.7)

XSL•FO
RenderX

Variables	Participants (N=115), n (%)
France	2 (1.7)
Canada	2 (1.7)
Italy	2 (1.7)
Saudi Arabia	2 (1.7)
Others[e]	12 (10.4)
Native language	
English	28 (24.3)
Turkish	32 (27.8)
Spanish	10 (8.7)
Chinese (Mandarin)	9 (7.8)
Arabic	5 (4.3)
Others[e]	31 (26.8)
Place of employment[f]	
Academic hospitals and medical centers	72 (64.2)
Community hospitals	9 (8.0)
Private hospitals	13 (11.6)
Public hospitals	15 (13.4)
Free clinics	6 (5.4)
Others[e]	6 (5.3)
Frequency of ChatGPT usage (n=68)	
Multiple times per day	14 (20.6)
Once per day	3 (4.4)
Three to five times per week	14 (20.6)
Less than three times a week	13 (19.1)
Only tried it few times	24 (35.3)

[a]Due to lack of responses, missing data are not included in the reported totals; as a result, some category counts may not sum to the overall sample size.
[b]NP/PA: nurse practitioner/physician assistant.
[c]RN: registered nurse.
[d]ICU: intensive care unit.
[e]For Others see Multimedia Appendix 3.
[f]The subcategories are not mutually exclusive.

XSL•FO
RenderX

Figure 1. ChatGPT usage based on participants' age, gender, and country of work.

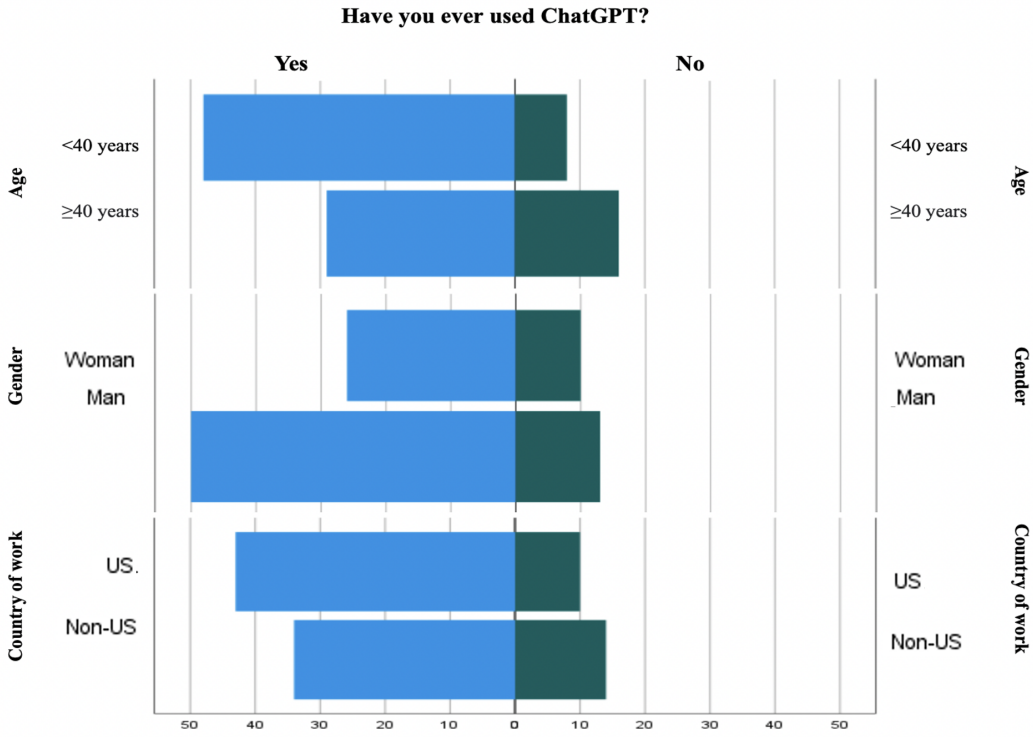
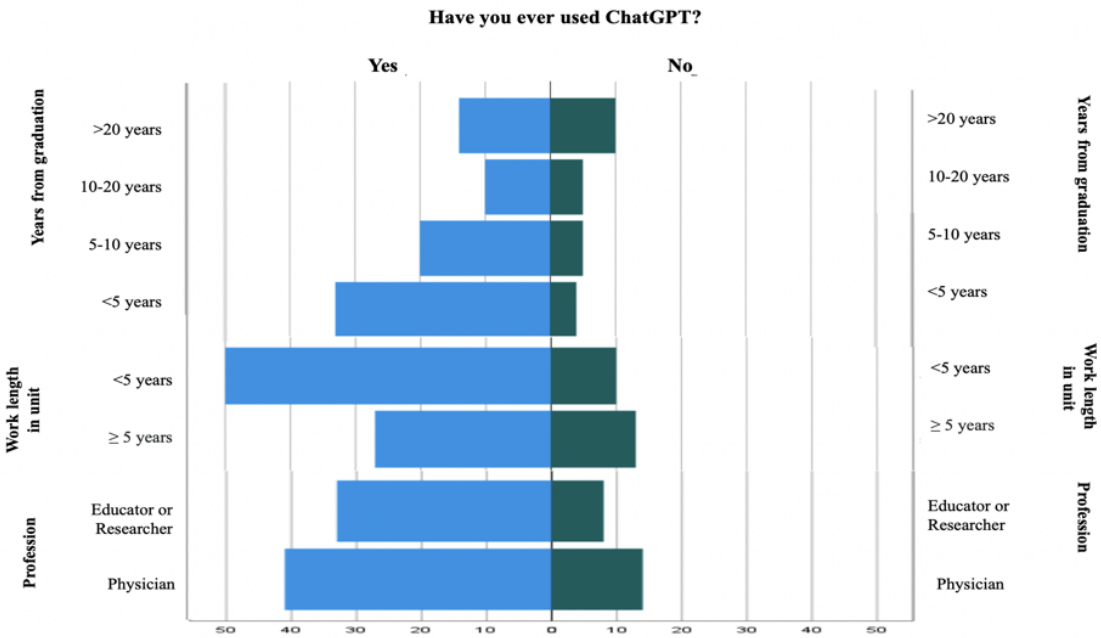


Figure 2. ChatGPT usage based on participants' years since graduation, length of work in the current unit, and profession.



Of the 115 participants, 101 (87.8%) had heard of ChatGPT, mainly from social media (n=33, 32.7%) and peers or colleagues (n=43, 42.6%). Of those, 77 (76.2%) had used ChatGPT before, with 18 (23.4%) using it multiple times per day and 23 (29.9%) having tried it only a few times. Moreover, 71 out of 77 (92.2%) participants used it in English. Among these, 50 were not native English speakers, and only 16/50 (32%) speakers used it both in English and their native language (Figure 3). Furthermore, variations in ChatGPT usage in daily practice were observed between participants using ChatGPT in English versus those who used it in their native language (Figure 4).

XSL•FO
RenderX

Figure 3. Ratio of native language use versus English use among participants while using ChatGPT.

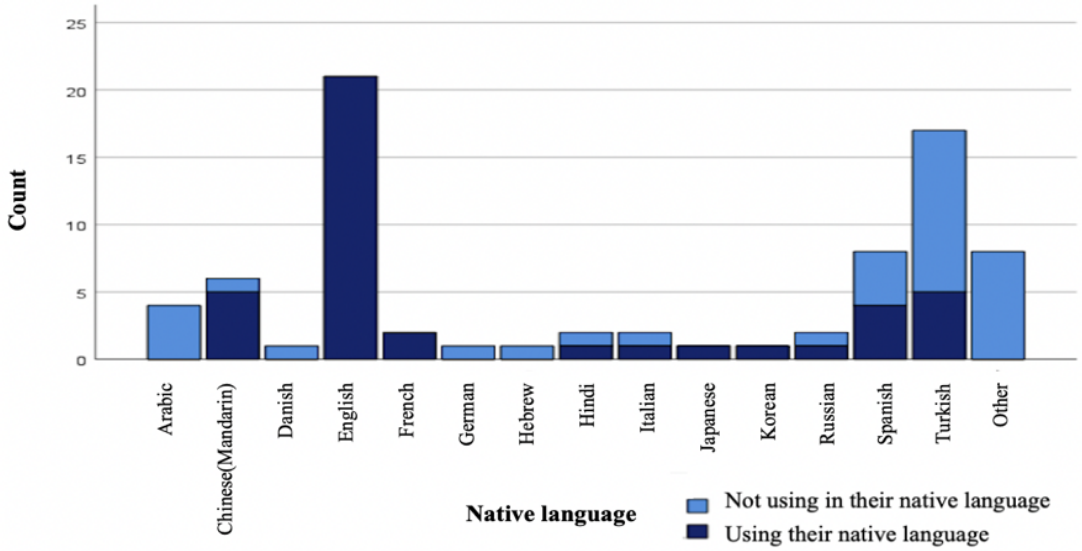
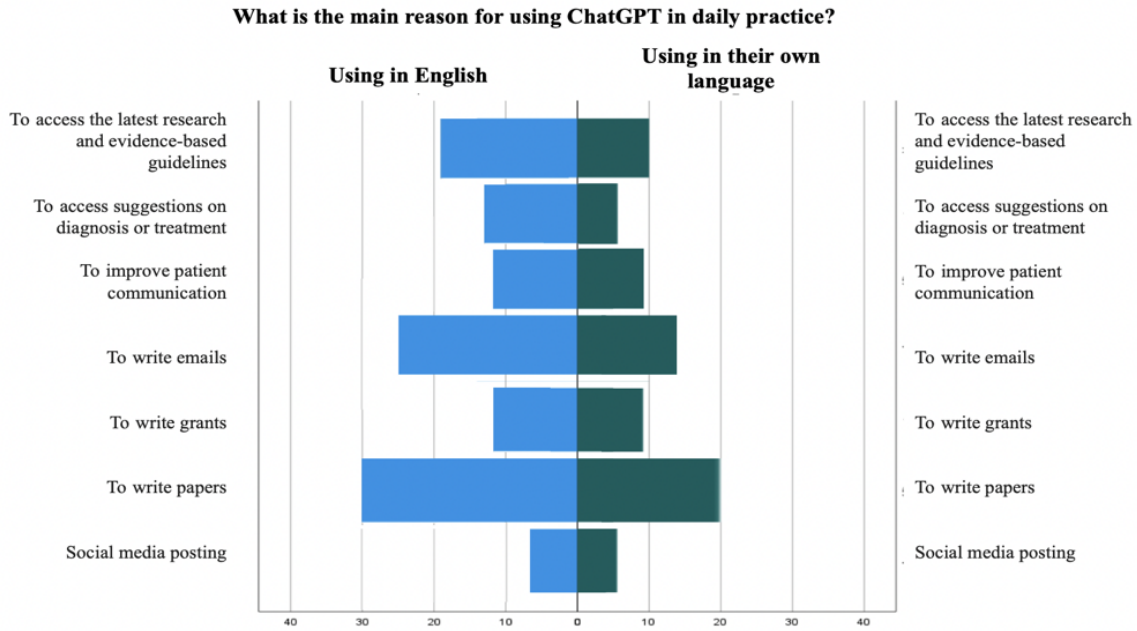


Figure 4. Main reasons for using ChatGPT in daily practice based on the language used by the participants.



The most common reasons to use ChatGPT included writing papers (n=29, 44.6%) and emails (n=25, 38.5%), and obtaining suggestions on diagnosis or treatment (n=14, 21.5%) (Table 2). Additional reasons for ChatGPT usage by health care professionals in daily practice are shared in Table 3.

XSL•FO
RenderX

Table . ChatGPT usefulness based on used features in daily practice.

ChatGPT features	Participants (n=68), n (%)
Usefulness in daily practice	
Not important	14 (20.6)
Slightly important	21 (30.9)
Moderately important	13 (19.1)
Important	13 (19.1)
Very important	7 (10.3)
ChatGPT's usefulness, 0 (most negative experience) to 10 (most positive experience)	
≥7	42 (61.8)
4-5-6	19 (27.9)
≤3	7 (10.3)
Most useful features	
To access and summarize research articles efficiently	22 (46.8)
To provide quick answers to clinical questions	15 (31.9)
To provide patient education materials	10 21.3
To write emails, grants, and papers	25 53.2

Table . Percentage of participants' main reasons for using ChatGPT in daily practice (multiple choice questions).

Main reason for using ChatGPT in daily practice	Participants (n=68), n (%)
Writing papers	31 (45.6)
Writing emails	25 (36.8)
To access the latest research and evidence-based guidelines	21 (30.9)
To access suggestions on diagnosis or treatment	15 (22.1)
To improve patient communication	12 (17.6)
To write grants	12 (17.6)

Incorporation of ChatGPT Into Daily Practice

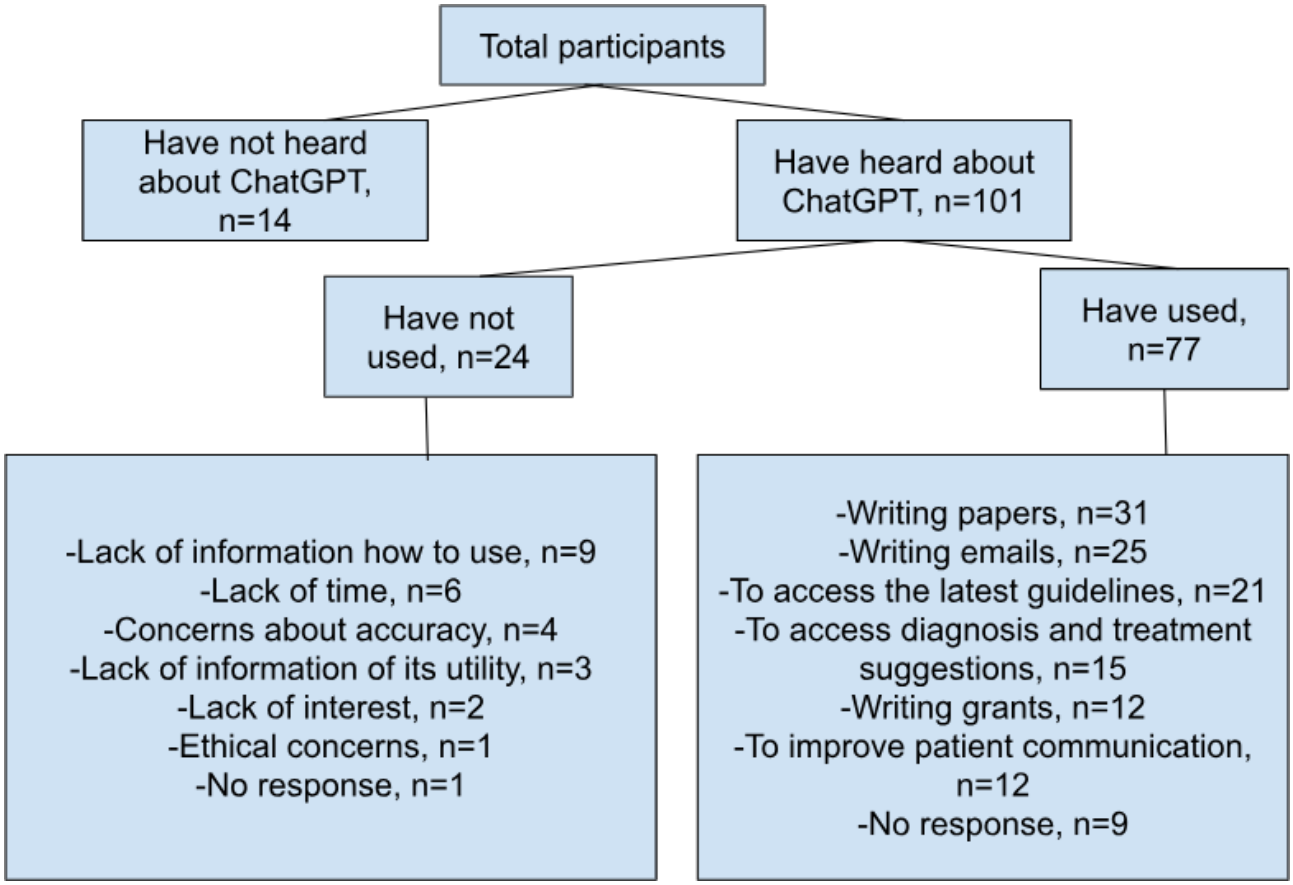
Of the 77 participants who used ChatGPT, 36 (46.8%) used ChatGPT in their clinical practice, 58 (75.3%) used it for research, and 56 out of 77 (72.7%) used it for educational activities (Figure 5).

Among all respondents, 42/101 (43.6%) participants agreed that they would not be concerned if their clinician used ChatGPT while providing care to them if they were the patient, whereas 32 (32.7%) disagreed and preferred that their clinician not use ChatGPT during care.

The majority (n=79, 78.2%) of participants agreed that ChatGPT could be useful for medical or health care professional education. In nonclinical settings, participants stated that ChatGPT could help to reduce workload (n=57, 73.1%), improve efficiency by automating certain tasks (n=51, 65.4%), offer greater access and efficiently summarize research articles (n=52, 66.7%), create patient educational materials (n=49, 62.8%), provide quick answers to questions (n=48, 61.5%), and enhance the ability to write papers (n=37, 47.4%).

XSL•FO
RenderX

Figure 5. Factors contributing to use and nonuse of ChatGPT. The activities are not mutually exclusive and therefore, the total number of participants may exceed 115.



Challenges for Integrating ChatGPT Into Daily Practice

The main reasons for respondents not using ChatGPT included concerns about the accuracy of ChatGPT responses (n=14, 29.8%), limited applicability to their practice (n=18, 38.3%), legal and ethical considerations (n=6, 12.8%), limited diagnostic capabilities (n=4, 8.5%), lack of time (n=3, 6.4%), and lack of interest (n=2, 4.3%).

As one of the significant barriers is legal and ethical considerations, participants were asked to define plagiarism or copyright violations. Participants defined it as copying text or ideas from ChatGPT and using it for another source without citation (n=64, 63.4%), paraphrasing or summarizing content from ChatGPT and using it for another source without citation (n=41, 40.6%), using images from ChatGPT without permission (n=36, 35.6%), reusing or repurposing content from ChatGPT that was previously created for another purpose without permission (n=44, 43.6%).

In response to the legal and ethical challenges, participants proposed several solutions for integrating ChatGPT into daily practice. Participants stated that data encryption (n=63, 62.4%), access control (n=52, 51.5%), user authentication such as two-factor authentication (n=48, 47.5%), compliance with regulations such as Health Insurance Portability and Accountability Act or General Data Protection Regulation (n=62, 61.4%), transparency and informed consent (n=53, 52.5%), and regular training and awareness for health care professionals (n=58, 57.4%) are necessary to ensure patient privacy and data security.

Views on ChatGPT's success and other possible uses

When asked whether the participants knew ChatGPT had performed with ≥60% accuracy on the USMLE, 52 (51.5%) participants indicated they had heard this before. Additionally, 76 (68.5%) participants reported that they had not used any other AI platform.

Participants stated that ChatGPT can improve patient outcomes through personalized health education by providing tailored information and support (n=76, 75.2%); assisting with medication management through reminders and refill prescriptions, and provide information on side effects and interactions (n=55, 54.5%); telemedicine support for health care professionals to conduct virtual consultations, collect patient data, and provide decision support (n=48, 50%); aiding in symptom triage for patients (n=49, 48.5%); and offering mental health support by providing guidance on self-management techniques and coping strategies (n=49, 48.5%).

The distribution of responses based on different levels of postgraduate experience is reported in Multimedia Appendix 4. This distribution was largely balanced between the participants with fewer than 10 years and those with 10 or more years of experience.

XSL•FO
RenderX

Discussion

Principal Findings

This study offers a global perspective on how health care professional perceive and use ChatGPT in clinical, research, and educational context. Our findings demonstrate that awareness and adoption of ChatGPT are already widespread, with 76.2% of respondents having used the tool at least once. Participants primarily reported using ChatGPT for manuscript and email writing, grant application preparation, accessing research articles, clinical guideline support, diagnostic suggestions, and improving patient communication. Notably, more than three-quarters of participants agreed that ChatGPT holds potential utility in medical education, highlighting its ability to enhance learning experiences and facilitate task automation. Moreover, our study indicates that health care professionals endorse its use among colleagues. However, concerns about data privacy, ethical risks such as plagiarism, and the accuracy of AI-generated content remained as significant barriers to broader adoption. Proposed solutions included implementing safety protocols such as data encryption, access control, and regulatory compliance. In exploratory analyses comparing ChatGPT use, we did not identify significant differences across professional experience levels, which might be due to the limited sample size. Due to the wide range and uneven distribution of medical subspecialties represented, we were not able to conduct a formal comparison across specialties.

Implications of Findings

Our findings highlight the broad and flexible potential of ChatGPT in health care workflows. In clinical practice, ChatGPT is perceived as a tool that can enhance efficiency by automating routine documentation tasks, such as generating draft discharge summaries and patient letters. It also supports decision-making by offering fast access to evidence summaries and aids communication through the creation of patient-friendly materials [5,15]. In medical education, participants identified ChatGPT as a valuable educational supplement—one that could be incorporated into curricula to simulate real-world clinical scenarios and assist in preparing students for standardized exams like the USMLE [5,16]. It can also support personalized learning experiences tailored to individual needs and self-directed learning pathways. In research, ChatGPT was valued for its ability in grant writing, literature synthesis, and ideation, especially in the early stages of manuscript development or protocol design [5].

These findings underscore the need for structured training programs and ethical guidelines to support responsible integration of AI tools. Implementing human-in-the-loop systems, in which clinicians oversee and validate AI outputs, may enhance safety, and build user confidence while mitigating risks associated with biases or inaccuracies in AI-generated content [17].

Comparison to the Literature

Our findings align with prior studies that underscore ChatGPT's potential in health care. Cascella et al [2] described ChatGPT's potential to reduce administrative burden and assist with clinical reasoning, which mirrors participants' reported use of ChatGPT for documentation and clinical queries. In medical education, Gilson et al [8] showed that ChatGPT achieved passing scores on all three components of the USMLE, highlighting its utility in medical education. Similarly, Kung et al [9] emphasized its role in creating standardized templates for patient education materials. These findings also align with our participants' views on its usefulness for both learners and patients alike. Sallam [18] highlighted ChatGPT's capacity to process and summarize complex medical data efficiently, which our participants also leveraged for research and evidence access.

However, our study adds unique insights by capturing global perspectives from diverse practice settings. Unlike prior reports focused on specific institutions or national populations, our results reflect a cross-disciplinary, international sample, offering a broader view of how generative AI is being perceived across diverse practice settings.

The main reasons behind the lack of use of ChatGPT in daily practice were mainly due to the nonapplicability to their practice, lack of information regarding its use, and concerns about the accuracy of ChatGPT's responses, and legal and ethical considerations. The reason behind not using ChatGPT due to lack of information may be partially attributed to insufficient training opportunities for health care professionals in the use of generative AI. Previous studies have also indicated similar concerns regarding its implementation [19]. For instance, the concern for the spread of wrong information is a major obstacle, and different languages may have inconsistent results [20,21]. Many studies have shown that up to 96.7% of users are concerned about ethical and legal obstacles [3,18], particularly plagiarism [21-23], and copyright issues [3,18]. In a study conducted by a university at Sweden, 62% of students considered the use of chatbots for assignments and exams as cheating [24]. Our study showed that 86 out of 101 participants defined copying from ChatGPT as plagiarism. These concerns show that the implementation of ChatGPT into clinical settings will require a transition period supported by extensive safety measures. Health care professional leaders need to work with technology experts to develop learning objectives, curricula, assessments and evaluations, and safety protocols for this emerging technology.

Regarding the accuracy of ChatGPT's responses, our study shows that health care professionals identified this as having a paramount importance. Similar studies have shown that ChatGPT should be used with caution due to potential biases of AI, which may lead to the generation of inaccurate information. When used in the health care system, this could potentially lead to harmful consequences [25].

Educational Implications

The educational relevance of our findings is especially important. Our study suggests several opportunities:

- Curriculum design: Educators can incorporate ChatGPT into simulation- and case-based learning modules to foster clinical reasoning and application of evidence-based medicine.

XSL•FO
RenderX

- Needs Assessment: Educators may use baseline familiarity and usage patterns to tailor AI training initiatives and address gaps in knowledge or ethical understanding.
- Institutional Strategies: ChatGPT may serve as a tool in flipped classrooms, interactive tutorials, and self-directed learning, offering real-time feedback and access to guideline-driven responses.
- Learner Outcomes: By providing immediate feedback and access to evidence-based guidelines, ChatGPT has the potential to improve learner performance on standardized assessments [16].

Additionally, ChatGPT's ability to generate accessible explanations for patients could enhance health literacy and improve communication between physicians and patients.

Strengths and Limitations

This study has several strengths. We examined ChatGPT adoption from a global perspective. By including participants from 21 countries and various clinical and academic backgrounds, the study provides a valuable overview of current usage patterns and attitudes toward generative AI tools in health care. The survey instrument was comprehensive, capturing a wide range of use cases and concerns across clinical, research, and educational domains.

However, several limitations must be acknowledged. Although participants were from diverse countries, they are unlikely to represent the full range of health care professionals within their regions. The sample was likely skewed toward individuals with greater access to technology and academic networks, especially in countries where access to ChatGPT or certain social media platforms may be restricted or limited. Therefore, findings should be interpreted with caution and may not be generalized to all health care professionals in low-resource or digitally restricted settings. The use of convenience and snowball sampling likely introduced self-selection bias, attracting participants with preexisting interest in technology or AI. Because of this sampling method, we could not calculate a response rate. Most respondents were from academic hospital settings in the United States, which may limit applicability to other regions or practice environments. Conducting the survey in English may have limited the global inclusivity. Given the swift pace of technological advancements, particularly in generative AI applications such as ChatGPT and the continuous process of learning and integration by health care professionals, the present survey may not accurately capture the current perceptions and attitudes of doctors and nurses toward these technologies [26], limiting the temporal relevance of our findings . Lastly, although our survey included open-ended questions, multiple-choice questions may have led participants to an available answer.

Future Directions

Further research is needed to address unanswered questions:

1. Long-term impact: Studies should evaluate how ChatGPT influences clinical outcomes, patient satisfaction, and educational performance over time.
2. Ethical frameworks: There is a pressing need for the development of institutional and regulatory guidelines governing AI use in health care [17].
3. Cross-language applications: Investigating how ChatGPT performs across different languages could help improve accessibility for non-English-speaking populations.
4. Training programs: Evidence-based strategies are needed to guide health care professionals in the ethical and effective use of generative AI technologies.

Conclusion

ChatGPT usage is expanding within health care settings due to its variety of capabilities, and the majority of health care professionals are likely aware of its availability. It can improve the caliber of writing papers, grants, and emails; help health care professionals in accessing the latest guidelines, diagnosis, and treatment suggestions; and possibly improve patient communication. There are several concerns related to the implementation of LLMs in clinical practice, including legal, ethical, and operational issues. Further research is necessary to clarify the role of ChatGPT and LLM-based generative AI tools in health care education, research, and clinical practice.

Acknowledgments

We thank Dr. Ognjen Gajic for critically reviewing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
ChatGPT Survey.
[PDF File, 59 KB - mededu_v11i1e58801_app1.pdf]

Multimedia Appendix 2
Diagram explaining survey flow.
[PNG File, 98 KB - mededu_v11i1e58801_app2.png]

Multimedia Appendix 3

XSL•FO
RenderX

Others within the demographic information table.
[DOCX File, 12 KB - mededu_v11i1e58801_app3.docx]

Multimedia Appendix 4
The distribution of answers to respondents with different levels of post-graduate experience.
[DOCX File, 24 KB - mededu_v11i1e58801_app4.docx]

References

1. Zhang K, Meng X, Yan X, et al. Revolutionizing health care: the transformative impact of large language models in medicine. J Med Internet Res 2025 Jan 7;27:e59069. [doi: 10.2196/59069] [Medline: 39773666]
2. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst 2023 Mar 4;47(1):33. [doi: 10.1007/s10916-023-01925-4] [Medline: 36869927]
3. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell 2023;6:1169595. [doi: 10.3389/frai.2023.1169595] [Medline: 37215063]
4. Mu Y, He D. The potential applications and challenges of chatgpt in the medical field. Int J Gen Med 2024;17:817-826. [doi: 10.2147/IJGM.S456659] [Medline: 38476626]
5. Tangsrivimol JA, Darzidehkalani E, Virk HUH, et al. Benefits, limits, and risks of ChatGPT in medicine. Front Artif Intell 2025;8:1518049. [doi: 10.3389/frai.2025.1518049] [Medline: 39949509]
6. Sallam M, Salim NA, Barakat M, Al-Tammemi AB. ChatGPT applications in medical, dental, pharmacy, and public health education: a descriptive study highlighting the advantages and limitations. Narra J 2023 Apr;3(1):e103. [doi: 10.52225/narra.v3i1.103] [Medline: 38450035]
7. Thomae AV, Witt CM, Barth J. Integration of ChatGPT into a course for medical students: explorative study on teaching scenarios, students' perception, and applications. JMIR Med Educ 2024 Aug 22;10:e50545. [doi: 10.2196/50545] [Medline: 39177012]
8. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 8;9:e45312. [doi: 10.2196/45312] [Medline: 36753318]
9. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for ai-assisted medical education using large language models. PLOS Digit Health 2023 Feb;2(2):e0000198. [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
10. Chen SY, Kuo HY, Chang SH. Perceptions of ChatGPT in healthcare: usefulness, trust, and risk. Front Public Health 2024;12:1457131. [doi: 10.3389/fpubh.2024.1457131] [Medline: 39346584]
11. Kim JK, Chua M, Rickard M, Lorenzo A. ChatGPT and large language model (LLM) chatbots: the current state of acceptability and a proposal for guidelines on utilization in academic medicine. J Pediatr Urol 2023 Oct;19(5):598-604. [doi: 10.1016/j.jpurol.2023.05.018] [Medline: 37328321]
12. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med 2023 Jul 6;6(1):120. [doi: 10.1038/s41746-023-00873-0] [Medline: 37414860]
13. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
14. Rossi R, Socci V, Pacitti F, et al. Mental health outcomes among frontline and second-line health care workers during the Coronavirus Disease 2019 (COVID-19) pandemic in Italy. JAMA Netw Open 2020 May 1;3(5):e2010185. [doi: 10.1001/jamanetworkopen.2020.10185] [Medline: 32463467]
15. Lu L, Zhu Y, Yang J, et al. Healthcare professionals and the public sentiment analysis of ChatGPT in clinical practice. Sci Rep 2025;15(1):1223. [doi: 10.1038/s41598-024-84512-y]
16. Khan AA, Khan AR, Munshi S, et al. Assessing the performance of ChatGPT in medical ethical decision-making: a comparative study with USMLE-based scenarios. J Med Ethics 2025 Jan 25:jme-2024-110240. [doi: 10.1136/jme-2024-110240] [Medline: 39863417]
17. Wang C, Liu S, Yang H, Guo J, Wu Y, Liu J. Ethical considerations of using ChatGPT in health care. J Med Internet Res 2023 Aug 11;25:e48009. [doi: 10.2196/48009] [Medline: 37566454]
18. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023 Mar 19;11(6):887. [doi: 10.3390/healthcare11060887] [Medline: 36981544]
19. De Angelis L, Baglivo F, Arzilli G, et al. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. Front Public Health 2023;11:1166120. [doi: 10.3389/fpubh.2023.1166120] [Medline: 37181697]
20. Chatterjee J, Dethlefs N. This new conversational AI model can be your friend, philosopher, and guide... and even your worst enemy. Patterns (N Y) 2023 Jan 13;4(1):100676. [doi: 10.1016/j.patter.2022.100676] [Medline: 36699746]
21. Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science. Nature New Biol 2023 Feb;614(7947):214-216. [doi: 10.1038/d41586-023-00340-6] [Medline: 36747115]

XSL•FO
RenderX

22. Gao CA, Howard FM, Markov NS, et al. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. NPJ Digit Med 2023 Apr 26;6(1):75. [doi: 10.1038/s41746-023-00819-6] [Medline: 37100871]
23. Rahimi F, Talebi Bezmin Abadi A. ChatGPT and publication ethics. Arch Med Res 2023 Apr;54(3):272-274. [doi: 10.1016/j.arcmed.2023.03.004] [Medline: 36990890]
24. Stöhr C, Ou AW, Malmström H. Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of study. Computers and Education: Artificial Intelligence 2024 Dec;7:100259. [doi: 10.1016/j.caeai.2024.100259] [Medline: 100259]
25. Sivarajah U, Wang Y, Olya H, Mathew S. Responsible artificial intelligence (AI) for digital health and medical analytics. Inf Syst Front 2023 Jun 5;2023:1-6. [doi: 10.1007/s10796-023-10412-7] [Medline: 37361886]
26. Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. Acad Med 2024 Jan 1;99(1):22-27. [doi: 10.1097/ACM.0000000000005439] [Medline: 37651677]

Abbreviations

AI: artificial intelligence
LLM: large language model
USMLE: United States Medical Licensing Examination

Edited by B Lesselroth; submitted 25.03.24; peer-reviewed by F Chen, K Achara, S Mao; revised version received 11.04.25; accepted 19.04.25; published 12.05.25.

Please cite as:
Ozkan E, Tekin A, Ozkan MC, Cabrera D, Niven A, Dong Y
Global Health care Professionals' Perceptions of Large Language Model Use In Practice: Cross-Sectional Survey Study
JMIR Med Educ 2025;11:e58801
URL: https://mededu.jmir.org/2025/1/e58801
doi:10.2196/58801

© Ecem Ozkan, Aysun Tekin, Mahmut Can Ozkan, Daniel Cabrera, Alexander Niven, Yue Dong. Originally published in JMIR Medical Education (https://mededu.jmir.org), 12.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Novel Blended Learning on Artificial Intelligence for Medical Students: Qualitative Interview Study

Zoe S Oftring[1,2*], MA, Dr med, MD; Kim Deutsch[3*], MA, Dr phil; Daniel Tolks[4,5], Dr rer biol hum; Florian Jungmann[6], PD, Dr med, MD; Sebastian Kuhn[1], Prof Dr, MME, MD

[1]Institute for Digital Medicine, Philipps University Marburg and University Clinic Giessen & Marburg, Baldingerstrasse 1, Marburg, Germany
[2]Department of Paediatrics, University Clinic Giessen & Marburg, Marburg, Germany
[3]Institute of Educational Science, Johannes Gutenberg University, Mainz, Germany
[4]Institute of Anatomy, Rostock University Medical Centre, Rostock, Germany
[5]Professorship in Health Management, International University of Applied Science, Hamburg, Germany
[6]Xcare Group Radiology, Nuclear Medicine and Radiotherapy, Saarlouis, Germany
[*]these authors contributed equally

Corresponding Author:
Sebastian Kuhn, Prof Dr, MME, MD
Institute for Digital Medicine, Philipps University Marburg and University Clinic Giessen & Marburg, Baldingerstrasse 1, Marburg, Germany

Abstract

Background: Artificial intelligence (AI) systems are becoming increasingly relevant in everyday clinical practice, with Food and Drug Administration–approved AI solutions now available in many specialties. This development has far-reaching implications for doctors and the future medical profession, highlighting the need for both practicing physicians and medical students to acquire the knowledge, skills, and attitudes necessary to effectively use and evaluate these technologies. Currently, however, there is limited experience with AI-focused curricular training and continuing education.

Objective: This paper first introduces a novel blended learning curriculum including one module on AI for medical students in Germany. Second, this paper presents findings from a qualitative postcourse evaluation of students' knowledge and attitudes toward AI and their overall perception of the course.

Methods: Clinical-year medical students can attend a 5-day elective course called "Medicine in the Digital Age," which includes one dedicated AI module alongside 4 others on digital doctor-patient communication; digital health applications and smart devices; telemedicine; and virtual/augmented reality and robotics. After course completion, participants were interviewed in semistructured small group interviews. The interview guide was developed deductively from existing evidence and research questions compiled by our group. A subset of interview questions focused on students' knowledge, skills, and attitudes regarding medical AI, and their overall course assessment. Responses were analyzed using Mayring's qualitative content analysis. This paper reports on the subset of students' statements about their perception and attitudes toward AI and the elective's general evaluation.

Results: We conducted a total of 18 group interviews, in which all 35 (100%) participants (female=11, male=24) from 3 consecutive course runs participated. This produced a total of 214 statements on AI, which were assigned to the 3 main categories "Areas of Application," "Future Work," and "Critical Reflection." The findings indicate that students have a nuanced and differentiated understanding of AI. Additionally, 610 statements concerned the elective's overall assessment, demonstrating great learning benefits and high levels of acceptance of the teaching concept. All 35 students would recommend the elective to peers.

Conclusions: The evaluation demonstrated that the AI module effectively generates competences regarding AI technology, fosters a critical perspective, and prepares medical students to engage with the technology in a differentiated manner. The curriculum is feasible, beneficial, and highly accepted among students, suggesting it could serve as a teaching model for other medical institutions. Given the growing number and impact of medical AI applications, there is a pressing need for more AI-focused curricula and further research on their educational impact.

(JMIR Med Educ 2025;11:e65220) doi:10.2196/65220

KEYWORDS

digital transformation; artificial intelligence; clinical AI; chatbot; digital literacy; medical education; medical students; medical curriculum; qualitative content analysis; medical studies

XSL•FO
RenderX

Introduction

Background

The digital transformation in the health care system represents a fundamental process of change and innovation that is altering the roles, competencies, and cooperation of doctors to a large extent [1]. Eric Topol [2] describes an increasing "super-convergence" of technologies that is transforming the existing health care system into a digital health care system. The key characteristics of this new system are individualization, precision, and prevention. It is expected that this will result in data-based health care that will be characterized by a pronounced intensification of interdisciplinary cooperation and a stronger participatory role for patients. Every patient is increasingly becoming a "big data" challenge, with huge amounts of information about previous illnesses and conditions. At the same time, existing medical knowledge is growing exponentially. These two facts cumulate in increasingly complex decision-making processes in patient care. One recent digital transformative technology that can help bridge this complexity gap by preparing, analyzing, and organizing large amounts of data is artificial intelligence (AI). In health care, AI is becoming increasingly important for extracting and interpreting clinically useful information from large volumes of digital data and information sources and, in some cases, deriving recommendations for therapeutic action.

In the following section, the term "AI applications in medicine" refers to medical software, devices, and technologies such as apps whose analytical processes are AI-based and which are used in the health care sector by patients and/or practitioners.

AI Applications in Medicine

Integrating AI applications into medical processes can automate repetitive tasks currently handled by humans. This hybrid working model improves human performance through technology. In 2012, the US Food and Drug Administration (FDA) certified a medical AI application for the first time [3]. Currently, the FDA database comprises 950 applications (as of the last FDA update on August 7, 2024), predominantly in radiology and the cardiovascular field [4]. Clinical AI systems have already demonstrated expert-level performance in radiology [5-7] and equaled the diagnostic performance of health care professionals in medical imaging [6]. Beyond radiology, there are numerous publications on the clinical application of AI [8-16] and large language models such as ChatGPT [17-19]. A scoping review by Han et al [20] generated an overview of all published randomized controlled trials on clinical AI as of November 2023 and found 84 studies. Their review underpins the growing evidence for the use of AI-supported tools in health care. However, from a populational and thus patient perspective, attitudes toward AI in health care are still fluid and demonstrate varying levels of knowledge, acceptance, and skepticism across different countries and demographic groups [21-24].

The Need for Curricular Training About AI in Medicine

This development has far-reaching implications for doctors and requires a fundamental examination of AI systems [25,26]. At present, neither medical professionals already practicing nor the generation currently studying is adequately prepared for the integration of AI in medicine. At the same time, both groups will—or are already—encountering actionable AI in their day-to-day work that is or will be able to predict, diagnose and, if necessary, treat diseases [2]. At a clinical level, doctors require the competencies to critically assess AI applications to use only those tools that have an evidence-based effect on improving clinical workflows or patient outcomes. At the development level, it is also important to ensure that doctors are actively involved in the development and scientific testing of new AI applications. This raises the question of the extent to which these systems can be effectively integrated into the diagnosis and treatment process, as well as how limitations of the systems can be recognized by medical users and how a fallback level can be ensured. In rapidly changing health care systems, it is therefore essential to ensure that doctors have the knowledge, skills, and attitudes to both master current challenges and be prepared for future challenges [1].

The basic competencies required for this must be learned by medical students during studies and continuously developed throughout their careers [27]. There are already various international ideas for this qualification mandate. For example, the Standing Committee of European Physicians addresses this goal in its Policy on Digital Competencies for Doctors and defines digital core competencies [28]. The EU Health Policy Platform has formulated specific instructions for achieving these core competencies [29]. According to these policymakers, educators should consider including content about the following skill sets into their curricula: (1) general digital skills (data and software security, ethical and legal implications), (2) technical digital skills (telemedicine, AI, health apps, smart devices, robotics, virtual reality/augmented reality, data literacy), and (3) the patient-doctor relationship (digital communication and collaboration, digital health literacy). Seth et al [30] created a theoretical framework of topics related to AI that need to be taught to train medical students in this technology. Laupichler et al [31] emphasize the need to assess medical students' AI literacy and attitudes in order to hone medical curricula to the AI educational needs of the next generation.

According to a recent review by Gordon et al [32], a growing number of medical schools are addressing AI throughout medical studies, but this is limited by the fact that only 2 of the 278 included studies focused on educational competencies in AI. For the German landscape, a study on national course programs found that the majority (72%) of surveyed medical schools stated that they offer AI-related learning opportunities [33]. In contrast to this, 70% of German medical students indicated in a survey conducted at the same time that they had never received any education in digital topics [34]. This surprising discrepancy can be explained by the fact that, although most German medical schools report offering such opportunities, they are mostly part of elective or extracurricular courses, with only 2 institutions including a separate subject specifically on AI in the core curriculum. As a result, existing AI curricula are currently only available to a very limited number of students, and large-scale AI education is still lacking. Recent studies consistently highlight knowledge gaps in AI

XSL•FO
RenderX

education from the perspectives of both international medical students [31,35-37] and German medical students [31,34,38-41]. Students displayed low familiarity with AI and limited awareness of its potential applications in health care; they also reported limited or uncertain access to AI education in medical training. However, they believed AI training would be beneficial and showed great interest in working with it. Results differed regarding attitudes. Although Alkhaaldi et al [36], Moldt et al [42], and Laupichler et al [31] found students to be more optimistic and accepting about AI applications, in Boillat et al's [37] survey there was more skepticism among students regarding the potential harm of AI for patients and job safety within the medical profession.

Despite the increasing number of medical schools that include AI-related teaching according to recent literature, current medical curricula struggle to meet the demands of students to equip them with a strong competency base to interact with, integrate, and critically evaluate AI tools in their clinical practice. Integrating education on core AI competencies into the general curriculum on a broader scale could significantly improve students' experience levels with AI, enhance their attitudes toward the technology, and better prepare them to navigate medical AI effectively in clinical practice.

The objective of this paper is twofold. The first part introduces the concept of a novel, multisession elective course, "Medicine in the Digital Age," which integrates AI teaching in the context of digital transformation into the medical curriculum at a German university. The second part presents findings from a qualitative interview evaluation of participants' feedback on the AI module as well as their overall experience of the elective. Our aim was to conduct an explorative prospective study using a semistructured qualitative interview approach to generate a multidimensional insight into students' knowledge, skills, and attitudes in dealing with AI in medical practice and their overall learning experience. For this, we asked students to comment on the following a priori deductive dimensions of interest: "Areas of AI Application," "Future Work," and "Critical Reflection." The interview findings supported the iterative refinement of our teaching concept, as well as the ongoing educational reform processes.

Methods

Ethical Considerations

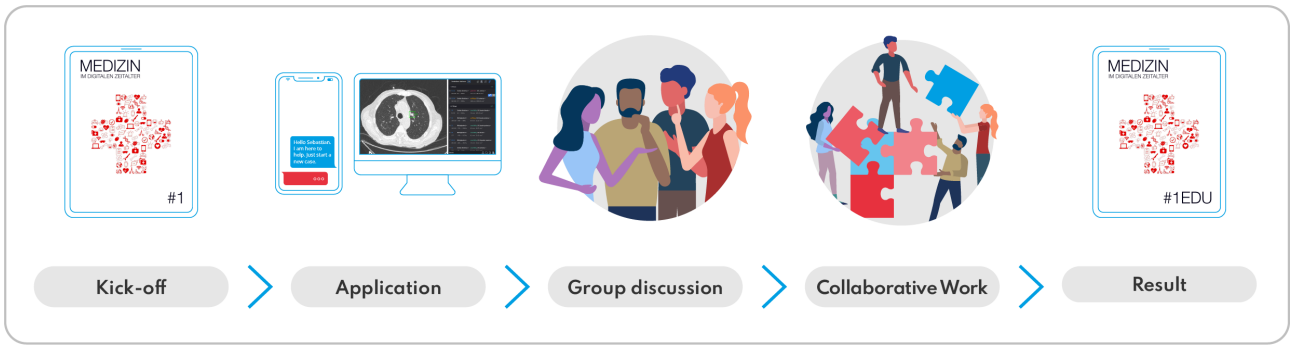
The local ethics committee was consulted during the development of the teaching evaluation for the curriculum presented. Following the consultation, the committee confirmed that the teaching evaluation constituted an additional quality assurance measure for teaching and curriculum development. This evaluation complements the existing concept for quality assurance at the Mainz University Medical Centre [43]. In accordance with the committee's recommendations, it was determined that an ethics vote was not considered necessary. However, participation in the evaluation required informed written consent from all students involved and was strictly voluntary. No compensation was provided to participants. To ensure confidentiality, all responses were pseudonymized prior to analysis.

Structure and Rationale Behind the Module on AI

The teaching module "Artificial Intelligence" is part of the competency-based multisession elective course entitled "Medicine in the Digital Age." The overall course aim is to equip students with digital skills that can be applied in a medically sound, technically feasible, legally compliant, data protection–compliant and ethically responsible manner and thus prepare them for the working environments of the future [2]. It was the first curriculum of its kind in Germany [1].

The "Medicine in the Digital Age" course consists of 5 modules: AI; digital doctor-patient communication; digital health applications and smart devices; telemedicine; and virtual/augmented reality and robotics. The course is offered to medical students in their clinical years as a 5-day elective, with each module comprising an 8-hour face-to-face course day. The didactic concept follows a flipped classroom and blended learning format by combining e-learning (e-book), face-to-face teaching (hands-on workshops, practical exercises, discussion and reflection formats), coproduction, and transfer projects (Figure 1). The different formats alternate over the course and build on each other.

Figure 1. Didactic concept consisting of e-learning, workshops, discussion/reflection formats, and transfer. The results of the learning process are incorporated into the e-learning e-book.



Using an interactive e-book, the participants deal with topics of digital transformation in the preliminary stages of the course. The e-book was created by our working group consisting of experts from medicine, medical education, ethics, media education, data science, and data protection, and also included patient perspectives. Its content mirrors the program of the elective course and contains a dedicated chapter for each module, combining theoretical background, reflective articles,

XSL•FO
RenderX

and stakeholder or patient interviews. Students are expected to read the corresponding chapter in preparation for each course day to independently develop the basics of digital medicine. The entire e-book follows the collaborative concept of "Do it by the book, but be the author" [44]. By incorporating all student transfer projects into an iterative version of the e-book, students are encouraged to actively interact with the course and become the "authors" of their elective course's e-book after completion of the elective.

The thematic breadth and interconnectedness of medical specialties necessitates an interdisciplinary team of lecturers. Therefore, onsite teaching is carried out by various medical disciplines (anesthetists, surgeons, medical informaticians, psychologists, pediatricians, psychosomatics, radiologists, orthopedic and trauma surgeons). In addition, computer scientists, representatives of federal state data protection and medical ethics, and patients complement the team of lecturers in the spirit of a transdisciplinary approach.

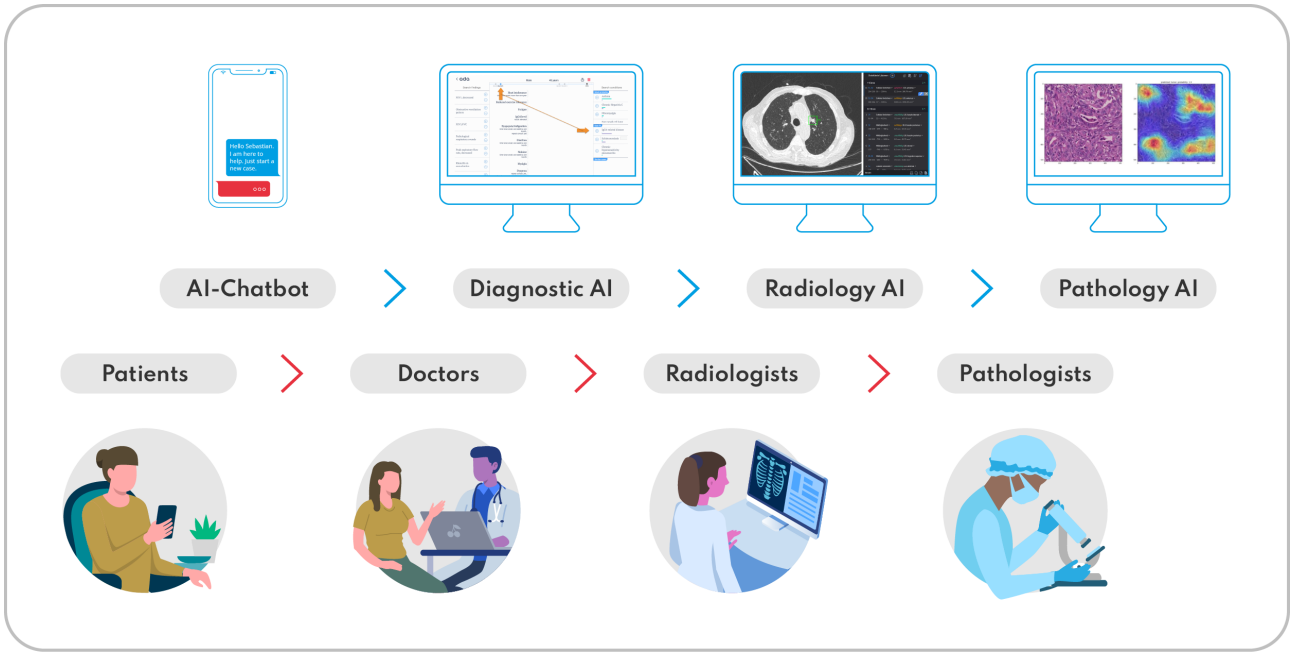
For the AI module, e-learning combined with face-to-face teaching and transfer tasks results in a total of 20 hours of teaching on AI. In accordance with the KSAVE model (Knowledge, Skills, Attitudes, Values, and Ethics) [45,46], the AI module aims at teaching the following overarching competences:

1. The student can describe various areas of application and programs that work with AI and is able to categorize clinical AI assistance systems in medical treatment in an evidence-based manner.
2. The student is able to explain examples of AI-assisted anamnesis, clinical examination, diagnosis, and therapy.
3. The student is able to name limitations of AI applications in current clinical practice and to evaluate the use and benefits of AI within the complex interplay of technical, legal, and ethical principles as well as under sociopolitical framework conditions and to place them in a medical context.
4. The student is able to reflect on how roles in the medical profession will change or evolve in the light of integrated AI assistance systems.

To give an example: a student demonstrates competence mastery by reflecting on the integration of AI-based systems into clinical workflows in oncological imaging in radiology, emphasizing their role in improving early identification of curative versus palliative needs and complementing clinical decision-making. They critically evaluate practical applications across anamnesis, diagnosis, and therapy, addressing limitations such as data quality, clinician acceptance, and ethical concerns. Through this analysis, the student links theoretical knowledge to real-world challenges, demonstrating readiness to apply AI to improve patient care and health care processes.

In the onsite teaching of the AI module, the focus is on practical workshops (Figure 2). These are designed to illustrate the integration of AI into medical treatment processes, followed by discussion and reflection sessions to promote the transfer to the students' own actions.

Figure 2. Workshops within the AI curriculum demonstrate hybrid workflows between humans and AI. The human workflow (lower section) is enhanced by the integration of AI-based narrow intelligence (upper section) along the patient care continuum by various medical specialists. AI: artificial intelligence.



Workshop on Medical History/AI Chatbot

To address the relevant technologies (ie, natural language processing, large language models, and chatbots), students are introduced to an AI-based smartphone app (Ada Health), which acts as a chatbot to take a symptom-based clinical history and make a suspected diagnosis [47-49]. Students then take a clinical history in groups of two from one of the lecturers, who takes on the role of a patient based on a predefined case vignette. First, one of the students takes a classic medical history and

XSL•FO
RenderX

formulates a suspected diagnosis. Thereafter, the second student obtains a medical history of the same patient case by reading out the chatbot's questions. Subsequently, the independently formulated suspected diagnoses are compared with the suspected diagnoses of the chatbot in the entire group. The students discuss which anamnesis questions and diagnoses they did not consider, and which questions and diagnoses the chatbot did not list. The usefulness of the different suspected diagnoses is then discussed and differences in the clinical histories are explored to determine the advantages and disadvantages of each method.

Workshop on Radiology/AI-Supported Radiology

Together with a radiologist, students learn how digitalization is changing the way radiologists work (eg, Picture Archiving and Communication System, radiology information system, speech recognition). Radiological AI applications are demonstrated as examples. Specifically, an AI for the automatic detection of tumor-specific lung foci is demonstrated. The software (InferVision, InferRead CT Lung) provides the user with the size and localization of the lesion as well as an estimate of the malignancy as a percentage. Additionally, an AI application for the automatic diagnosis of conventional X-ray examinations of the thorax is presented (Oxipit, ChestEye). The students learn that a number of published papers have already shown that various AI applications are equivalent to radiologists in individual subtasks [6].

Workshop on Pathology/AI-Supported Pathology

Students are introduced to the influence of digitalization on the field of pathology. For this purpose, a pathologist demonstrates the use of AI as a supporting tool in the diagnosis and detection of malignant changes in histopathological tissue sections [16]. Both the technical and informative background, as well as the process of developing and scientifically evaluating an AI application (AI development life cycle) and its practical application (integration into patient care), are illustrated. Students learn more about the future potential of AI in pathology and can ask questions and contribute their own thoughts.

Discussion and Reflection Formats

For reflection, students and lecturers discuss the following questions together in fishbowl discussions:

- What are the opportunities and risks of using AI in the context of patient treatment?
- How do we deal with probabilities calculated by an AI?
- What will your day-to-day work look like in 2025?
- What new skills will you need in the future?

Transfer Projects

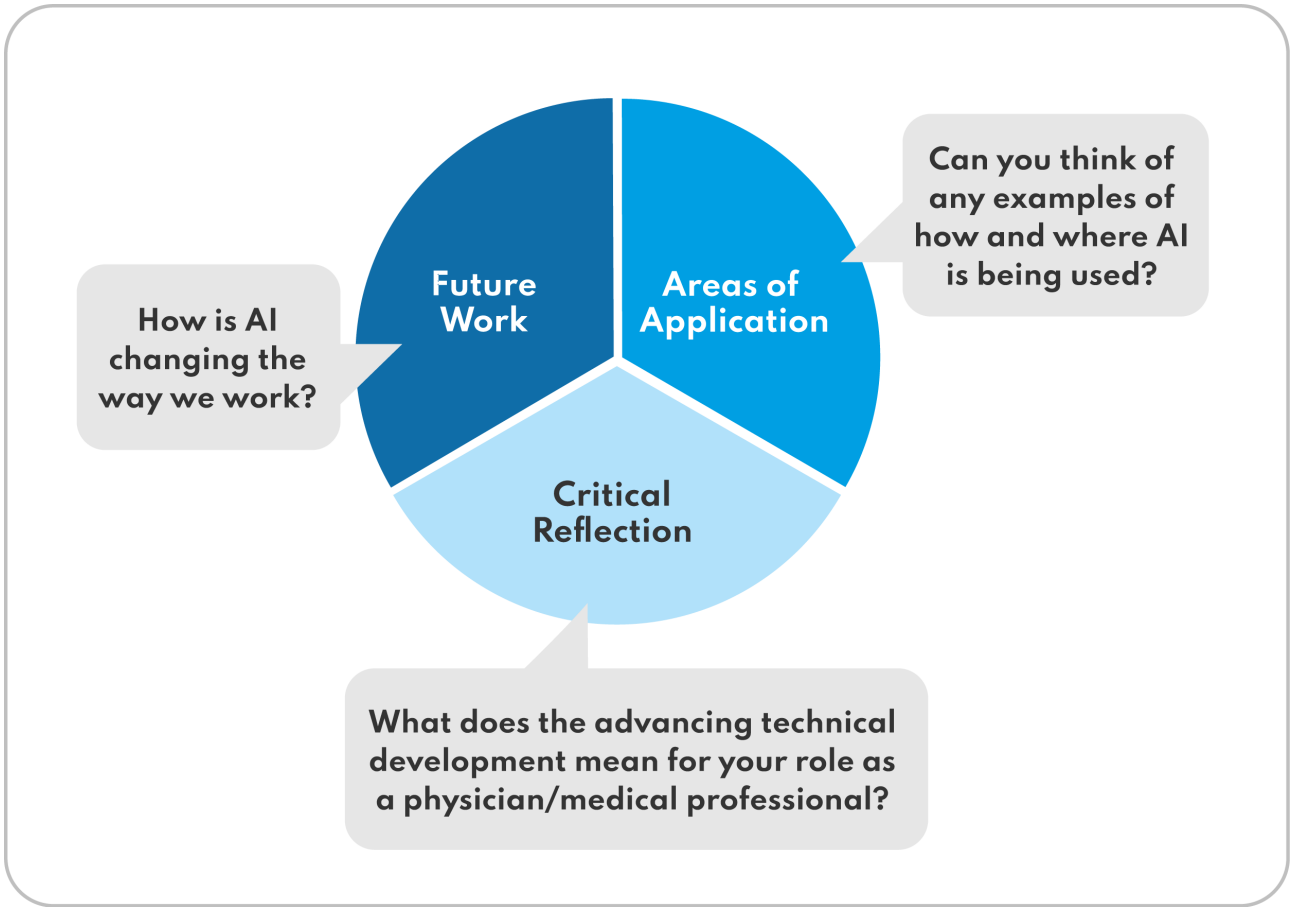
Throughout the course, students work in self-selected small groups (groups of 4) on the overarching task of researching a useful medical AI application and presenting it in plenary on the last day of the course. The group presentations are followed by 15-minute discussion rounds with the plenum. The transfer projects students chose reflect the wide range of AI in terms of technology (language, imaging, data procession), medical use cases (conservative medicine, surgical medicine), and different age groups (from AI solutions for pediatrics to palliative care). In addition to their research, part of students' transfer performances also lies in critically analyzing and presenting their various solutions. Among others, students addressed topics such as an AI-supported ultrasound image navigation for regional anesthesia [50], an AI algorithm supporting clinicians in sepsis management [13], or an AI algorithm predicting the end of life developed by Stanford University [51].

Evaluation

The elective course on Medicine in the Digital Age including the presented AI module was introduced to the medical curriculum at the University Medical Centre of the Johannes Gutenberg-University Mainz. The evaluation of the 3 groups consisting of medical students in their second and third clinical year (ie, years 4 and 5 of the 6-year medical program) was carried out using semistructured, focused, guided group interviews consisting of open and targeted questions based on Merton, Fiske, and Kendall [52,53]. The interview questions were formulated based on the KSAVE model. The interview guide aimed to ascertain the participants' existing competencies in the areas of knowledge, skills, and attitudes in dealing with AI in physicians' practice. This theoretical background resulted in 3 main interview topics: "Areas of Application," "Future Work," and "Critical Reflection" (Figure 3). Additionally, the last section of the interview addressed students' overall assessment of the course. The interview guide was used to evaluate the entire course and is provided in the appendix (Multimedia Appendix 1).

XSL•FO
RenderX

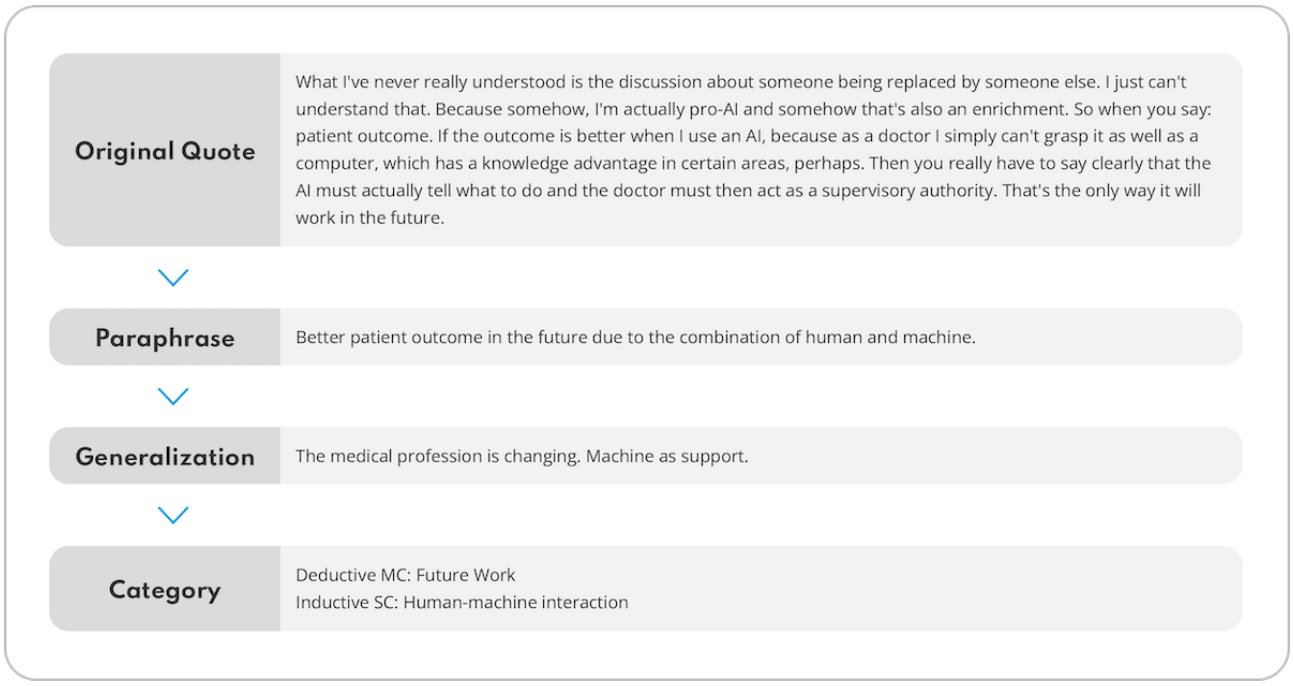
Figure 3. Outline of the main points of the interview guide questions on AI (circle) including guide questions (boxes outside). AI: artificial intelligence.



The focus group interviews took place within 2 weeks of the comprehensive course. Participants were informed about recording, transcription, data usage, storage, and privacy, and consent was obtained beforehand. The interviews were conducted and recorded by a researcher with expertise in qualitative research. The audio files were then transcribed for further analysis by 3 student research assistants (KD, LU, and EK). The interview transcripts were subsequently analyzed using content-structuring qualitative content analysis according to Mayring [54]. This is a text analysis method that follows a logical, systematic pattern and aims at transferring raw text data into structured categories (Figure 4). Categories can be formed deductively based on previous knowledge or hypotheses, or inductively based on new, text-immanent findings. For a structured evaluation of the results, the categories formed in the category system were organized hierarchically into main categories (MCs) and subcategories (SCs). The MCs were deductively derived from the research questions prior to the survey phase. During the analysis process, additional inductive SCs were formed from the interview statements. Inductive coding was used and codings were discussed and agreed upon by the coding research assistants. Saturation was achieved by interviewing all participants and subsequently coding all interview material.

XSL•FO
RenderX

Figure 4. Process of category development through the continuous comparison of the content of the deductive categories with the compiled material. Through the steps of paraphrasing, generalization, and reduction of the content-structuring content analysis, additional inductive categories can be developed, and the statements can then be assigned to an explicit category. MC: main category; SC: subcategory.



The Results section presents the parts of the overall evaluation results that explicitly relate to the AI module and the overall course assessment.

Results

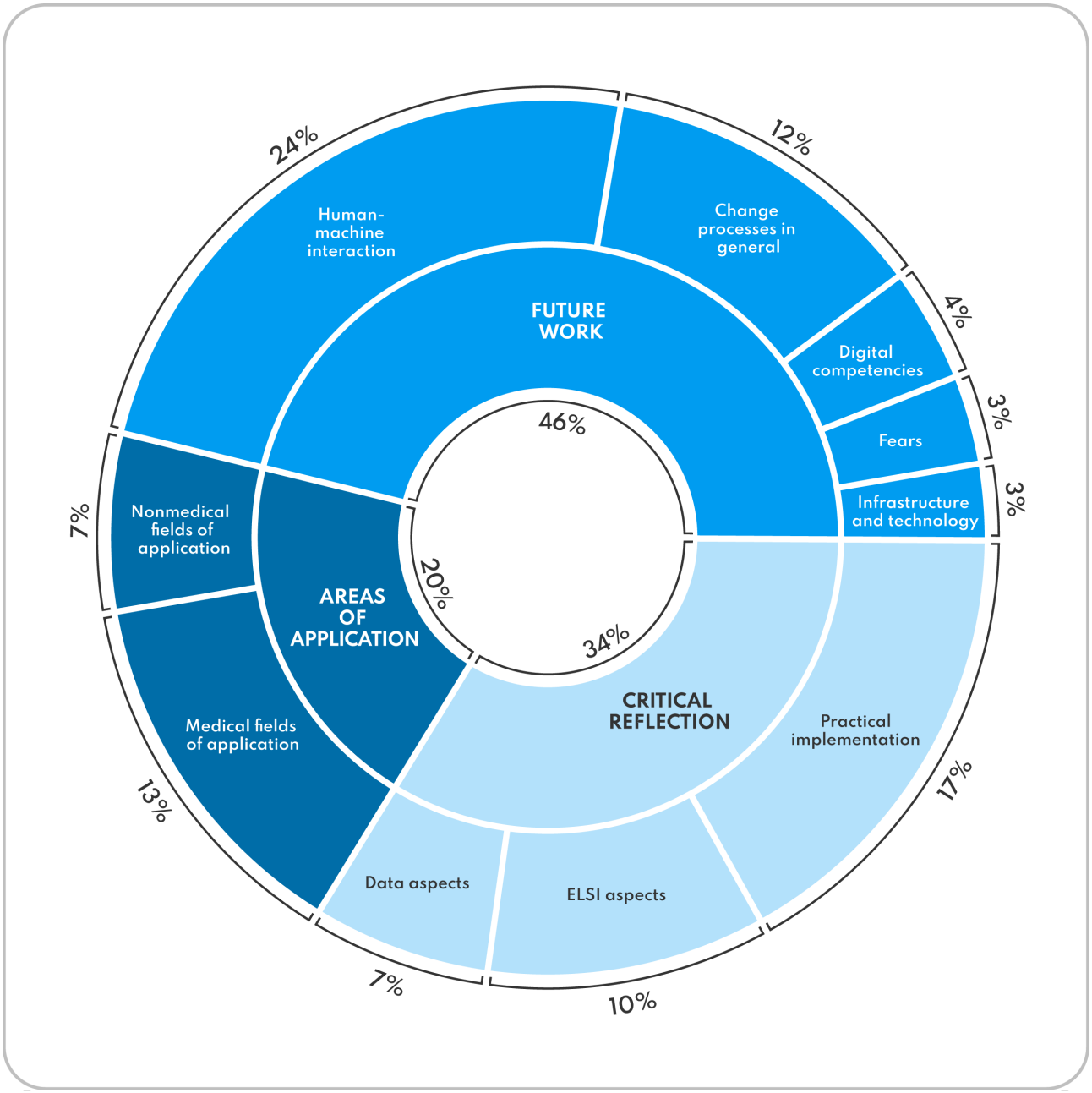
Evaluation Outcomes

From 3 group cycles, 18 semistructured, focused, guided interviews were conducted with all 35 participants (female=11, male=24) from the 3 consecutive courses, which formed the basis for the qualitative evaluation of the course concept. The interviews lasted 24:36 minutes on average. In all interviews, a total of 214 statements were made that could be assigned to the area of “Artificial Intelligence” and 610 statements related to “Overall Course Assessment.”

Following the analysis steps outlined above, the 3 deductive MCs related to AI—“Areas of Application,” “Future Work,” and “Critical Reflection”—were assigned further concrete inductive SCs derived from the content of the text during the evaluation process (Figure 5). Statements in the MC “Overall Course Assessment” were divided into the 6 SCs “Learning Experience,” “Learning Success,” “Structure,” “Content,” “Methods,” and “Conclusion.” For quality assurance purposes, the results report was written in accordance with the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist [55]. The following section details the qualitative results for each main category. Anchor quotes support the result report for each category. For this, quotes were translated into English by the authors and minor changes were made to improve readability. An extensive overview of anchor quotes for all SCs is provided in the appendix (Multimedia Appendix 2). Identification codes in parentheses accompany each quote to allow allocation to individual participants (the ID code reads as follows: course run:interview number:speaker ID).

XSL•FO
RenderX

Figure 5. Graphical depiction of the qualitative research results on the AI teaching module broken down into main and subcategories, based on 214 coding units (student statements). Percentages are relative to the overall sample of 214 statements on AI. All percentages are rounded. AI: artificial intelligence; ELSI: Ethical, Legal, and Social Implications.



Main Category "Areas of Application"

This category investigated students' knowledge about the existence of different AI applications in health care and beyond and their attitudes toward them after completing the elective course. A total of 20% (43/214) of all statements on AI fall into the category "Areas of Application." Of these, 29 fall into the SC "Medical fields of application" and 14 into the SC "Nonmedical fields of application."

For the SC "Medical fields of application," students listed a variety of medical use cases. The topics that were covered in the course dominated, namely diagnosis in general as well as image diagnosis in radiology, pathology, and dermatology.

> *During the week, for example, I learnt that the radiologists can have the lung round foci assessed by AI during the CT scan. With different probabilities. […] Then we learned in the pathology department that in the future AI will also calculate […] how high the probability is whether it is a tumor or not. Or […] this dermatology AI, which can show whether it's a melanoma or a benign mole. […] So I've definitely learnt a lot, I could go on listing all the examples.* [3:6:F]

In terms of diagnosis, participants rated the use of AI in triage and preliminary anamnesis in the outpatient sector as useful. This could reduce waiting times and counteract missing information. Regarding image diagnosis, the students refer to AI as a "safety net" backing up their own findings with a digital second opinion.

XSL•FO
RenderX

Interviewees expressed disbelief about the status quo on several levels. They discussed the limited usage of AI applications in everyday clinical practice and their own lack of knowledge before attending the course. They identified a general absence of curriculum focused on AI in medicine within their own course of study. Overall, students were surprised by the variety of possible applications, as well as by the rapidly increasing number of market-ready AI medical products. They expressed great interest in the use of AI in their medical practice and see the meaningful and context-specific application of AI as an urgent task of the present.

For the SC "Nonmedical fields of application," at the nonmedical level, students mentioned various possible or already existing application scenarios for AI, some of which have been adopted into everyday life without reflection (eg, voice assistants, navigation, purchase recommendations). The most attention was paid to autonomous driving and transportation.

Main Category "Future Work"

A total of 46% (99/214) of all statements on AI fall into the category "Future Work." Of these, 51 fall into the SC "Human-machine interaction," 26 into the SC "Change process in general," 9 into the SC "Digital competencies," 7 into the SC "Fears" and 6 into the SC "Infrastructure and technology."

For the SC "Human-machine interaction," students focus on the question of how AI can contribute to becoming a better doctor. The students state that they thought intensively about the combination of humans and machines during the course week and consider this combination to be the optimum for the future practice of medicine.

> *Just like the chess player and the computer together (Centaur Chess computer), they are unbeatable. […] and hopefully it will be the same with the doctor [and AI].* [2:1:3]

"Shared decision making" and digital second opinions can counteract incorrect treatment or improve treatment outcomes in the sense of assistive systems and thus increase patient safety. The students describe their experiences with the anamnesis chatbot and characterize the comparison of the two forms of anamnesis as very informative. The differences between the chatbot anamnesis and their own anamnesis practice were thus revealed. Emphasis was placed on the aspect of social anamnesis, which the students carried out more intensively than the chatbot. However, the students rated the chatbot's anamnesis procedure as more structured and systematic than their own. Students predominantly regard AI as an opportunity.

For the SC "Change processes in general," students did not consider the medical profession to be threatened by the use of AI, but the practice of the profession and the subdivision of specialties may change. Students stated that the rapid increase in knowledge means that doctors are even more obliged than before to undergo continuous further training. The preinformed patient will become more of a discussion partner at eye level. The interviewees see this as a great opportunity to improve the doctor-patient relationship, as the inclusion of AI in routine medical procedures could lead to an increase in personnel and time resources, allowing doctors to focus on the traditional core medical activities of consultation, treatment, and care.

For the SC "Digital competencies," the competency profile and requirements for the medical profession are also changing as a result of the transformation. Here, the interviewees speak of a lack of or inadequately trained digital skills, which are also not considered in the standard curriculum. Students would like a safe framework for trying out new technology. Furthermore, the use of AI in everyday medical work requires clear quality criteria that are similar to a review process for evaluating technologies. The ability to correctly classify AI-generated information and to critically question the statements made by the AI is regarded as particularly relevant. According to the students, it is and remains the task of the doctor to assess which application is appropriate for the individual patient and when treating physicians need to actively decide against its use.

> *Where does the AI get the data from? How is it analyzed? And above all, how is it checked to make sure it really is a sensible AI? You need to know that. In a way just like we learn how to read scientific publications. And decide whether they are good or bad.* [2:1:3]

For the SC "Fears," the biggest problem addressed is general ignorance and the resulting fear of, for example, the threat of job losses. Students suspected that this fear is the reason why AI applications are often not developed by health care experts, but by fast-moving commercial companies. Students also considered the combination of human and machine to be problematic if it is not possible for the doctor to understand how the AI operates and reaches decisions.

For the SC "Infrastructure and technology," the students note that the technical change in everyday working life is particularly noticeable through deficits in (technical) equipment.

Main Category "Critical Reflection"

A total of 34% (72/214) of all statements on AI fall into the category "Critical Reflection." Of those, 36 fall into the SC "Practical implementation," 22 into the SC "ELSI aspects" (Ethical, Legal, Social Implications), and 14 into the SC "Data aspects."

For the SC "Practical implementation," students see opportunities for larger-scale, international cooperation. This requires openness and investment in progress and research. They take their practical experience from the "clinical anamnesis" workshop as an example of low-threshold contact, which they want to take out of the course to raise awareness. The limitations of the chatbot, for example, making a misdiagnosis, are critically questioned. In such cases, the intended time saving backfires and becomes extra work.

> *That would unsettle me […] if something completely different comes out as a treatment suggestion or what I see in an MRI image or something like that, then it would make me very insecure and then I would want to make sure. Be it through the senior physician or that I can just have a look: How does this AI come up with this? And if that doesn't work, then it's*

XSL•FO
RenderX

> *unfavorable for the procedure. Then you have to rely on the senior consultant again and maybe the head physician […] and then I'm back where I was before without AI.* [2:1:3]

One problem discussed is that current AIs can only act as "narrow intelligence" in very specific settings, meaning that anomalies that do not correspond to the AI's specific field of action remain undetected. This results in the risk of unconsidered use. In general, all interviewees share the opinion that not everything that is technically possible should also be used in practice and that each use case must be considered individually.

For the SC "ELSI aspects," the question of whether increasing digitalization will reduce or intensify doctor-patient contact is viewed critically. On the one hand, digitalization represents an opportunity to relieve doctors and invest the freed-up capacities in the doctor-patient relationship. On the other hand, there is a risk that AI will impact the interpersonal interaction and thus the patient's individuality.

The possibility of consciously influencing AI is the subject of intense ethical debate. Specifically, it is questioned at what point it is unethical not to use the advancing technology, as this would deliberately deny the patient the best possible treatment.

> *At some point it becomes unethical not to use such things. […] That's actually the point. Why are we always so afraid that we're not important enough? At some point, the doctor is no longer the all-knowing person.* [2:3:1]

The interviewees see a further ethical dilemma in the case of a discrepancy between the diagnosis provided by the doctor and the AI. The right not to know and the handling of probabilities play a decisive role in sensitive areas, such as prenatal or genetic diagnostics and palliative medicine. Students also discussed the unclear legal situation regarding liability issues as a possible cause of rejection of AI applications.

For the SC "Data aspects," regarding data protection, too little regulation violates personal rights. Too much regulation makes it difficult or, in the worst case, prevents access to data for clinical research. In general, students also question the lack of traceability of AI results. They critically note that convenience or lack of time can lead to the results not being checked over time.

Main Category "Overall Course Assessment"

Of all 610 "Overall Course Assessment" statements, 134 fell into the SC "Learning Experience" and 108 into the SC "Learning Success." The remaining statements were categorized into the SCs "Structure" (n=61), "Content" (n=126), "Methods" (n=142), and "Conclusion" (n=39).

For the SCs "Learning Experience" and "Learning Success," students highlighted engaging with AI and digitalization as a significant learning success, given the absence of such topics in the standard curriculum.

> *I'm just glad that I had this week, because it really showed me what we don't learn at university. And how big the topic actually is for us.* [2:2:2]

They described the hands-on interaction with various technologies as "eye-opening" (2:2:4) and the group work on human-AI comparisons as "impressive" (2:1:4). Many students, initially skeptical or ambivalent about AI, reported increased knowledge and awareness of AI technologies and a deeper understanding of their impact as a result of the elective. They felt better prepared for their future careers regarding questioning and categorizing digital tools such as apps or AI, and underlined the gain in competences:

> *I think everyone left with a gain in expertise. Be it in the form of medical expertise, technical expertise, or simply that you've thought about things like data protection and apps and so you've also gained absolute everyday expertise.* [1:3:B1]

For the SCs "Structure," "Content," and "Methods," students appreciated the involvement of diverse experts, valuing the variety of perspectives on the technology.

> *What was outstanding […] was that the input came from the legal side, from the ethical side, from the technical side somehow every time.* [3:8:A]

They praised the active and innovative learning format of the elective, noting that it encouraged reflection and engagement rather than the rote learning typical of other subjects.

> *It's often the case that you're told things and then you have to memorize them. And here it was more the case that you were given information but then had to think about it yourself, for example to discuss it or draw a picture or whatever. And that's a completely different kind of learning, which unfortunately we don't usually do that much of in our degree programs. So I thought it was really good. Because these are actually skills that you should have and not that you can somehow memorize a book.* [2:1:3]

The discussion formats were highlighted as a distinctive feature in comparison with previous teaching experiences, with critical reflection helping students develop a more nuanced understanding of the topic.

> *I think you learnt an incredible amount, especially in the discussions, and you were actually forced to really think about certain theses. I also found this kind of discussion extremely productive.* [2:2:4]

For the SC "Conclusion," students almost unanimously agreed that the elective had broadened their horizons and appreciated the opportunity to participate. They wished that the course would be expanded so that more students could participate. Some expressed a wish for more breaks or even longer discussion sessions. Although students felt that the scope and time commitment of the elective was appropriate, many would have liked it to last longer:

> *"I think the biggest minus is actually the time. It's rare that you leave a course saying: 'Hey, I wish I'd stayed longer.' But […] Tuesday and Wednesday were actually days when I thought: 'Okay. I could have stayed two hours longer'.* [2:2:5]

XSL•FO
RenderX

In summary, the qualitative evaluation showed a high level of acceptance of the course concept and differentiated attitude toward AI among students. The course participants emphasized the increase in their knowledge and competences about the technology as well as the appreciation they felt as a result of the intensive and varied collaboration with each other and with the lecturers. The opportunity for critical discussion, practical interaction, and application was rated particularly positively. All 35 students stated that they would recommend the elective course to peers.

Discussion

Principal Findings

The digitalization of medicine and the use of AI applications is a fundamental process of change that will have a major impact on the future job profile of doctors to an extent that cannot yet be foreseen. What is certain, however, is that we are transitioning from the "information age" to the "age of artificial intelligence" [26] and that the integration of AI into medical treatment processes will redefine human-machine interaction. It is therefore essential to prepare future doctors to use AI in daily practice [27]. At present, although curricula are beginning to change, structured teaching concepts are lacking in terms of curricular mapping, although educators and practitioners emphasize the need to impart such competencies both nationally and internationally. Most students also advocate for AI education in their studies and report limited or no exposure to AI technologies and learning resources [35-37,41]. This does not mean that students must be able to program themselves but they must learn the practical application of AI in line with ELSI principles, data science, biostatistics, and evidence-based medicine during their studies [30,56].

The qualitative results of the AI module show that the embedding of curricular teaching about AI is generally feasible and sensible, that the added value of such a teaching module is recognized by students and acknowledged with great interest and acceptance, and that it leads to an increase in competence among students and promotes a critical and reflective attitude toward new technologies. Regarding the core aspects reflected in the main categories of the analysis (Areas of Application, Future Work, Critical Reflection), several key points can be learned, as detailed in the following sections.

Areas of Application

At present, it is not sufficiently clear how and when AI should be used in clinical diagnostics and therapy. Regarding the "how," students demonstrate a forward-thinking and nuanced examination to potential AI applications in clinical settings.

With regard to the "when," clarification is needed on the specific areas and questions where AI can assist in the clinical workflow [30,57]. Here, students express ambivalence about its integration, acknowledging both benefits and risks.

Future Work

Regarding patient care, students highlighted AI's potential to enhance care through personalized application and resource optimization, aligning with its reported ability to save time and personnel resources amid health care resource scarcity [58]. Students expressed some apprehension about the future impact of AI on the medical profession, though concerns about career choices were less prominent. This aligns with a survey in which 83% of medical students disagreed that AI would render radiologists obsolete [41]. Nonetheless, a minority of participants expressed fears about career prospects. Although AI's full impact remains unpredictable, it is undeniable that medical professions will change. Wartmann and Combs [26] speak of a "reboot" of the health care system and postulate the need to skillfully manage the interface between medicine and machines, as AI will surpass human capabilities in certain tasks [26]. Reflecting this, students stressed the importance of human-machine interaction and corresponding digital skills.

Critical Reflection

Most students underlined the potential of AI for their future career while maintaining a critical perspective, avoiding blind enthusiasm. They emphasized the risk of AI manipulation and its consequences for patient care, underscoring the need for doctors to retain ultimate decision-making authority over AI recommendations. The evaluation presented here thus indicates students' development of a critical attitude due to the module. These findings underscore the importance of future medical curricula teaching students to integrate AI assistance into their decision-making processes [2].

At the industry and developer level, students also acknowledged the need to design AI applications with ELSI aspects in mind. Such recommendations already exist. For example, the multisociety statement on the ethics of AI in radiology [59] and a white paper from the European Society of Radiology outline key ethical and practical considerations for the responsible use of AI in clinical practice [60].

In summary, students acknowledged the evolving nature of AI in health care as well as the necessity for skillful management of the interface between medicine and AI. They emphasized the importance of human-machine interaction as well as the need to develop digital skills while maintaining a reflective mindset toward technology.

Implications

Current literature on medical students' evaluation of their AI competencies demonstrates a relevant knowledge gap and the need for rapid-employment curricula solutions to change this. Overall, both the Medicine in the Digital Age elective as well as its AI teaching module were demonstrated to be feasible and reasonable teaching concepts, which supported maintaining the blended learning approach and the basic content of the modules. Nevertheless, valuable insights for iteration were drawn from the evaluation. First, the course has been updated to reflect technological and regulatory developments, such as the AI Act. Second, insights from students' transfer projects and reflective discussions informed an "agility by design" [61] approach, incorporating noteworthy projects or themes identified through students' input into the subsequent course iterations.

With the didactic framework, course design, and content outlined, this teaching concept can serve as a transferable model for implementation and adaptation in other universities or

XSL•FO
RenderX

training settings. Adjustments may be required to address specific target groups or local circumstances.

Evaluating the AI module's learning objectives—knowledge, skills, and reflection—is critical in both simulated (eg, Observed Structured Clinical Examination exams) and real-world settings. Complex educative interventions like this require robust assessment of efficacy and sustainability. Future research should explore whether a single elective suffices, if refresher courses are needed, or if phased AI education is beneficial. To validate this teaching course, prospective longitudinal trials comparing students who attended the AI module and untrained students are essential.

Methodological Strengths

The "Medicine in the Digital Age" curriculum described here addresses the digital transformation of medicine in an interdisciplinary and interactive way for medical students. AI is one of the 5 teaching modules and the rapid development and adoption of AI technologies in health care requires students and professionals to familiarize themselves with it and develop an attitude toward it. Standard quantitative methods can only inadequately depict the development of a professional attitude. The potential of the qualitative methodology used in teaching research should therefore be emphasized. Qualitative approaches provide insights into the learners' assessment of individual learning success, including gains in the areas of knowledge, skills, and attitude, as well as the content design and methodological structure. They are therefore ideal for the iterative further development of teaching concepts and the evaluation of attitude-oriented teaching content. The application of qualitative methodology represents a distinctive strength and unique contribution of this study. Although most research on medical students' perceptions relies on quantitative questionnaire surveys, this study uses qualitative survey instruments. The 2 existing qualitative studies in Germany are limited by their focus on analyzing free-text survey responses [34] and by their narrow scope, specifically examining students' attitudes toward mental health chatbots [42]. To the best of our knowledge, this study is the first to offer comprehensive, in-depth qualitative insights into German medical students' perceptions and attitudes toward AI.

Limitations

A common limiting factor in qualitative research is the small sample size. Helfferich [62] cites a sample size of between 6 and 120 respondents as appropriate. This means that our sample of 35 students can be assumed to have sufficient result validity. A second limitation could be that the results present a retrospective evaluation. Incorporating a qualitative pre-post analysis might have drawn a more concise picture of students' changes in knowledge, attitudes, and reflection on AI as a result of the course. Third, the findings might not be generalizable to other medical training programs, student attitudes, countries, or demographics. Lastly, group dynamics in the focus groups might have influenced the outcome by introducing social desirability bias.

Conclusions

Digitalization will continue to fundamentally change medicine. Therefore, in line with international appeals, today's education and training curricula must teach students and practicing physicians the basic competencies for using digital tools such as AI applications. It is not enough to simply integrate context-specific AI solutions as teaching examples into existing curricula. The aim of future curricula must be to equip students with the key competencies for their future day-to-day work in the age of AI and enable them to internalize knowledge, skills, and attitudes toward these tools from the beginning of their training. As an outlook for the AI curriculum presented here, it can be said that it addresses this need in a unique way. The qualitative teaching evaluation showed that students were able to deal with the topic in a very differentiated way after the AI teaching unit. The transferability of the curriculum to other university locations can be assumed in principle. The curriculum could therefore serve as an exemplary teaching concept for other universities and contribute to training medical students in two future-oriented skills: AI literacy and its transfer to medical human-machine interaction.

Acknowledgments

Above all, the authors would like to thank all participants of this study for their valuable insights. Furthermore, they would like to thank Björn Hirte for his great support in designing the illustrations in this paper. Open Access funding was provided by the Open Access Publishing Fund of Philipps-Universität Marburg.

Authors' Contributions

All authors were responsible for drafting, conceptualization, methodology, validation, and visualization. KD was responsible for interview curation and analysis. SK was responsible for funding acquisition, project administration, resources, software, and supervision. SK, FJ, KD, and ZSO were responsible for curricular development. DT supported the methodology portion through his expertise as a medical educator. KD, ZSO, and SK wrote the original draft. ZSO edited and finalized the initial draft. All authors reviewed and edited the manuscript.

Conflicts of Interest

SK is the founder and a shareholder of MED.digital. All other authors declare no conflicts of interest.

Multimedia Appendix 1

XSL•FO
RenderX

Interview guide for artificial intelligence in medicine.
[DOC File, 25 KB - mededu_v11i1e65220_app1.doc]

Multimedia Appendix 2
Qualitative results. Anchor quotes for all categories.
[DOC File, 60 KB - mededu_v11i1e65220_app2.doc]

References

1. Kuhn S. Medizin im digitalen Zeitalter: Transformation durch Bildung [Article in German]. Deutsches Ärzteblatt 2018 Jun;115(14):633-638.
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
3. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med 2020;3(1):118. [doi: 10.1038/s41746-020-00324-0] [Medline: 32984550]
4. Artificial intelligence and machine learning (AI/ML)-enabled medical devices. US Food and Drug Administration. URL: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices?utm_medium=email&utm_source=govdelivery [accessed 2025-02-26]
5. Ting DSW, Liu Y, Burlina P, Xu X, Bressler NM, Wong TY. AI for medical imaging goes deep. Nat Med 2018 May;24(5):539-540. [doi: 10.1038/s41591-018-0029-3] [Medline: 29736024]
6. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health 2019 Oct;1(6):e271-e297. [doi: 10.1016/S2589-7500(19)30123-2] [Medline: 33323251]
7. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
8. Annarumma M, Withey SJ, Bakewell RJ, Pesce E, Goh V, Montana G. Automated triaging of adult chest radiographs with deep artificial neural networks. Radiology 2019 Apr;291(1):196-202. [doi: 10.1148/radiol.2018180921] [Medline: 30667333]
9. Arbabshirani MR, Fornwalt BK, Mongelluzzo GJ, et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. NPJ Digit Med 2018;1(1):9. [doi: 10.1038/s41746-017-0015-z] [Medline: 31304294]
10. Murphy A, Skalski M, Gaillard F. The utilisation of convolutional neural networks in detecting pulmonary nodules: a review. Br J Radiol 2018 Oct;91(1090):20180028. [doi: 10.1259/bjr.20180028] [Medline: 29869919]
11. Lindsey R, Daluiski A, Chopra S, et al. Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci U S A 2018 Nov 6;115(45):11591-11596. [doi: 10.1073/pnas.1806905115] [Medline: 30348771]
12. Kitamura G, Chung CY, Moore BE. Ankle fracture detection utilizing a convolutional neural network ensemble implemented with a small sample, de novo training, and multiview incorporation. J Digit Imaging 2019 Aug;32(4):672-677. [doi: 10.1007/s10278-018-0167-7] [Medline: 31001713]
13. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 2018 Nov;24(11):1716-1720. [doi: 10.1038/s41591-018-0213-5] [Medline: 30349085]
14. Schaffert D, Bibi I, Blauth M, et al. Using automated machine learning to predict necessary upcoming therapy changes in patients with psoriasis vulgaris and psoriatic arthritis and uncover new influences on disease progression: retrospective study. JMIR Form Res 2024 Jun 27;8:e55855. [doi: 10.2196/55855] [Medline: 38738977]
15. Chatterji S, Niehues JM, van Treeck M, et al. Prediction models for hormone receptor status in female breast cancer do not extend to males: further evidence of sex-based disparity in breast cancer. NPJ Breast Cancer 2023 Nov 8;9(1):91. [doi: 10.1038/s41523-023-00599-y] [Medline: 37940649]
16. Försch S, Klauschen F, Hufnagl P, Roth W. Artificial intelligence in pathology. Dtsch Arztebl Int 2021 Mar 26;118(12):194-204. [doi: 10.3238/arztebl.m2021.0011] [Medline: 34024323]
17. Truhn D, Weber CD, Braun BJ, et al. A pilot study on the efficacy of GPT-4 in providing orthopedic treatment recommendations from MRI reports. Sci Rep 2023 Nov 17;13(1):20159. [doi: 10.1038/s41598-023-47500-2] [Medline: 37978240]
18. Griewing S, Gremke N, Wagner U, Lingenfelder M, Kuhn S, Boekhoff J. Challenging ChatGPT 3.5 in senology-an assessment of concordance with breast cancer tumor board decision making. J Pers Med 2023 Oct 16;13(10):1502. [doi: 10.3390/jpm13101502] [Medline: 37888113]
19. Mesko B. The ChatGPT (generative artificial intelligence) revolution has made artificial intelligence approachable for medical professionals. J Med Internet Res 2023 Jun 22;25:e48392. [doi: 10.2196/48392] [Medline: 37347508]

XSL•FO
RenderX

20. Han R, Acosta JN, Shakeri Z, Ioannidis JPA, Topol EJ, Rajpurkar P. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. Lancet Digit Health 2024 May;6(5):e367-e373. [doi: 10.1016/S2589-7500(24)00047-5] [Medline: 38670745]
21. Miró Catalina Q, Femenia J, Fuster-Casanovas A, et al. Knowledge and perception of the use of AI and its implementation in the field of radiology: cross-sectional study. J Med Internet Res 2023 Oct 13;25:e50728. [doi: 10.2196/50728] [Medline: 37831495]
22. Artificial intelligence: public awareness survey. GOV.UK. 2019. URL: https://www.gov.uk/government/publications/artificial-intelligence-public-awareness-survey [accessed 2025-02-26]
23. Rainie L, Funk C, Anderson M, Tyson A. How Americans think about artificial intelligence. Pew Research Center. 2022. URL: https://www.pewresearch.org/internet/2022/03/17/how-americans-think-about-artificial-intelligence/ [accessed 2025-05-19]
24. Fischer S, Petersen T, Bertelsmann Stiftung. Was Deutschland Über Algorithmen Weiß Und Denkt [Book in German]: Bertelsmann Stiftung; 2018. [doi: 10.11586/2018022]
25. Wartman SA. The empirical challenge of 21st-century medical education. Acad Med 2019 Oct;94(10):1412-1415. [doi: 10.1097/ACM.0000000000002866] [Medline: 31299675]
26. Wartman SA, Combs CD. Medical education must move from the information age to the age of artificial intelligence. Acad Med 2018 Aug;93(8):1107-1109. [doi: 10.1097/ACM.0000000000002044] [Medline: 29095704]
27. Masters K. Artificial intelligence in medical education. Med Teach 2019 Sep;41(9):976-980. [doi: 10.1080/0142159X.2019.1595557] [Medline: 31007106]
28. CPME policy on digital competencies for doctors. Standing Committee of European Doctors (CPME). 2020. URL: https://www.cpme.eu/api/documents/adopted/2020/11/CPME_AD_Board_21112020_100.FINAL_.CPME_.Policy.Digital.Competencies.for_.Doctors.pdf [accessed 2025-02-26]
29. European Commission's Thematic Network on Digital skills for future-proof doctors (Digital Doc). Training future-proof doctors for the digital society. European Junior Doctors. 2020. URL: https://www.juniordoctors.eu/sites/default/files/2021-01/Digital%20Doc_Training%20future-proof%20doctors%20for%20the%20digital%20society.docx.pdf [accessed 2025-02-26]
30. Seth P, Hueppchen N, Miller SD, et al. Data science as a core competency in undergraduate medical education in the age of artificial intelligence in health care. JMIR Med Educ 2023 Jul 11;9:e46344. [doi: 10.2196/46344] [Medline: 37432728]
31. Laupichler MC, Aster A, Meyerheim M, Raupach T, Mergen M. Medical students' AI literacy and attitudes towards AI: a cross-sectional two-center study using pre-validated assessment instruments. BMC Med Educ 2024 Apr 10;24(1):401. [doi: 10.1186/s12909-024-05400-7] [Medline: 38600457]
32. Gordon M, Daniel M, Ajiboye A, et al. A scoping review of artificial intelligence in medical education: BEME Guide No. 84. Med Teach 2024 Apr;46(4):446-470. [doi: 10.1080/0142159X.2024.2314198] [Medline: 38423127]
33. Mosch L, Back A, Balzer F, et al. Lernangebote zu künstlicher intelligenz in der medizin [Report in German]. : Zenodo; 2021. [doi: 10.5281/ZENODO.5497668]
34. Sorg H, Ehlers JP, Sorg CGG. Digitalization in medicine: are German medical students well prepared for the future? Int J Environ Res Public Health 2022 Jul 7;19(14):8308. [doi: 10.3390/ijerph19148308] [Medline: 35886156]
35. Liu DS, Sawyer J, Luna A, et al. Perceptions of US medical students on artificial intelligence in medicine: mixed methods survey study. JMIR Med Educ 2022 Oct 21;8(4):e38325. [doi: 10.2196/38325] [Medline: 36269641]
36. Alkhaaldi SMI, Kassab CH, Dimassi Z, et al. Medical student experiences and perceptions of ChatGPT and artificial intelligence: cross-sectional study. JMIR Med Educ 2023 Dec 22;9:e51302. [doi: 10.2196/51302] [Medline: 38133911]
37. Boillat T, Nawaz FA, Rivas H. Readiness to embrace artificial intelligence among medical doctors and students: questionnaire-based study. JMIR Med Educ 2022 Apr 12;8(2):e34973. [doi: 10.2196/34973] [Medline: 35412463]
38. Gillissen A, Kochanek T, Zupanic M, Ehlers J. Medical students' perceptions towards digitization and artificial intelligence: a mixed-methods study. Healthcare (Basel) 2022 Apr 13;10(4):723. [doi: 10.3390/healthcare10040723] [Medline: 35455898]
39. Weidener L, Fischer M. Artificial intelligence in medicine: cross-sectional study among medical students on application, education, and ethical aspects. JMIR Med Educ 2024 Jan 5;10(1):e51247. [doi: 10.2196/51247] [Medline: 38180787]
40. Busch F, Hoffmann L, Truhn D, et al. Medical students' perceptions towards artificial intelligence in education and practice: a multinational, multicenter cross-sectional study. medRxiv. Preprint posted online on 2023. [doi: 10.1101/2023.12.09.23299744]
41. Pinto Dos Santos D, Giese D, Brodehl S, et al. Medical students' attitude towards artificial intelligence: a multicentre survey. Eur Radiol 2019 Apr;29(4):1640-1646. [doi: 10.1007/s00330-018-5601-1] [Medline: 29980928]
42. Moldt JA, Festl-Wietek T, Madany Mamlouk A, Nieselt K, Fuhl W, Herrmann-Werner A. Chatbots for future docs: exploring medical students' attitudes and knowledge towards artificial intelligence and medical chatbots. Med Educ Online 2023 Dec;28(1):2182659. [doi: 10.1080/10872981.2023.2182659] [Medline: 36855245]
43. Qualitätssicherung [Website in German]. Universitätsmedizin Mainz. URL: https://www.um-mainz.de/rfl/studium-lehre/informationen-fuer-lehrende-und-einrichtungen/evaluation-qualitaetssicherung.html [accessed 2025-05-19]

XSL•FO
RenderX

44. Kuhn S, Kirchgässner E, Deutsch K. Medizin im digitalen Zeitalter – 'Do it by the book … but be the author!' [Website in German].: Synergie; 2017. URL: https://www.synergie.uni-hamburg.de/de/media/ausgabe04/synergie04-beitrag06-kuhn-kirchgaessner-deutsch.pdf [accessed 2025-02-26]
45. Binkley M, Erstad O, Herman J, et al. Defining twenty-first century skills. In: Griffin P, McGaw B, Care E, editors. Assessment and Teaching of 21st Century Skills: Springer Netherlands; 2012:17-66. [doi: 10.1007/978-94-007-2324-5_2]
46. Seidl T. (Wert-)haltung, als wichtiger bestandteil der entwicklung von 21st century skills an hochschulen. AG curriculum 4.0. diskussionspapier nr. 3 [Report in German]. : Zenodo; 2018. [doi: 10.5281/ZENODO.2634975]
47. Kuhn S, Jungmann SM, Jungmann F. Künstliche Intelligenz für Ärzte und Patienten: “Googeln” war gestern [Article in German]. Dtsch Arztebl Int 2018;115(26):A-1262. [Medline: 30135007]
48. Jungmann SM, Klan T, Kuhn S, Jungmann F. Accuracy of a chatbot (Ada) in the diagnosis of mental disorders: comparative case study with lay and expert users. JMIR Form Res 2019 Oct 29;3(4):e13863. [doi: 10.2196/13863] [Medline: 31663858]
49. Ada Health. URL: https://ada.com/de/ [accessed 2025-02-26]
50. Magoon R, Suresh V. A novel recognition of artificial intelligence in regional anaesthesia. Digit Med 2023 Jun;9(2):e00003. [doi: 10.1097/DM-2023-00003]
51. Hsu J. Stanford’s AI predicts death for better end-of-life care. IEEE Spectrum. 2018. URL: https://spectrum.ieee.org/stanfords-ai-predicts-death-for-better-end-of-life-care [accessed 2025-02-26]
52. Merton RK, Fiske M, Kendall P. The Focused Interview: A Manual of Problems and Procedures, 1st edition: Free Press; 1956.
53. Friebertshäuser B, Langer A, Prengel A, Boller H, Richter S, editors. Handbuch Qualitative Forschungsmethoden in Der Erziehungswissenschaft 3, Vollständig Überarb Aufl, Neuausg [Book in German]: Juventa-Verl; 2010.
54. Mayring P. Qualitative Inhaltsanalyse [Book in German]: Beltz; 2015.
55. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
56. Park SH, Do KH, Kim S, Park JH, Lim YS. What should medical students know about artificial intelligence in medicine? J Educ Eval Health Prof 2019;16:18. [doi: 10.3352/jeehp.2019.16.18] [Medline: 31319450]
57. Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. NPJ Digit Med 2018;1(1):40. [doi: 10.1038/s41746-018-0048-y] [Medline: 31304321]
58. Topol EJ, Verghese A. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again, 1st edition: Basic Books; 2019.
59. Geis JR, Brady AP, Wu CC, et al. Ethics of artificial intelligence in radiology: summary of the Joint European and North American Multisociety Statement. Radiology 2019 Nov;293(2):436-440. [doi: 10.1148/radiol.2019191586] [Medline: 31573399]
60. European Society of Radiology (ESR). What the radiologist should know about artificial intelligence - an ESR white paper. Insights Imaging 2019 Apr 4;10(1):44. [doi: 10.1186/s13244-019-0738-2] [Medline: 30949865]
61. Kuhn S, Jungmann F, Deutsch K, Drees P, Rommens PM. Digitale Transformation der Medizin [Book in German]. OUP 2018;7:453-458. [doi: 10.3238/oup.2018.0453–0458]
62. Helfferich C. Die Qualität Qualitativer Daten [Book in German], 4th edition: VS Verlag für Sozialwissenschaften; 2011.

Abbreviations

AI: artificial intelligence
COREQ: Consolidated Criteria for Reporting Qualitative Research
ELSI: Ethical, Legal, and Social Implications
FDA: US Food and Drug Administration
KSAVE: Knowledge, Skills, Attitudes, Values and Ethics
MC: main category
SC: subcategory

Edited by B Lesselroth; submitted 09.08.24; peer-reviewed by J Aulenkamp, M Laupichler, N Schlicker; revised version received 03.03.25; accepted 06.04.25; published 26.05.25.

Please cite as:
Oftring ZS, Deutsch K, Tolks D, Jungmann F, Kuhn S
Novel Blended Learning on Artificial Intelligence for Medical Students: Qualitative Interview Study
JMIR Med Educ 2025;11:e65220
URL: https://mededu.jmir.org/2025/1/e65220
doi:10.2196/65220

XSL•FO
RenderX

© Zoe S Oftring, Kim Deutsch, Daniel Tolks, Florian Jungmann, Sebastian Kuhn. Originally published in JMIR Medical Education (https://mededu.jmir.org), 26.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Chatbots' Role in Generating Single Best Answer Questions for Undergraduate Medical Student Assessment: Comparative Analysis

Enjy Abouzeid*, MBChB, MSc, PhD; Rita Wassef*, MBBCh, MSc, MD; Ayesha Jawwad*, BDS, MPH; Patricia Harris*, BSc(Hons), PhD

School of Medicine, University of Ulster, Northland Road, Derry-Londonderry, United Kingdom

*all authors contributed equally

Corresponding Author:
Enjy Abouzeid, MBChB, MSc, PhD
School of Medicine, University of Ulster, Northland Road, Derry-Londonderry, United Kingdom

Abstract

Background: Programmatic assessment supports flexible learning and individual progression but challenges educators to develop frequent assessments reflecting different competencies. The continuous creation of large volumes of assessment items, in a consistent format and comparatively restricted time, is laborious. The application of technological innovations, including artificial intelligence (AI), has been tried to address this challenge. A major concern raised is the validity of the information produced by AI tools, and if not properly verified, it can produce inaccurate and therefore inappropriate assessments.

Objective: This study was designed to examine the content validity and consistency of different AI chatbots in creating single best answer (SBA) questions, a refined format of multiple choice questions better suited to assess higher levels of knowledge, for undergraduate medical students.

Methods: This study followed 3 steps. First, 3 researchers used a unified prompt script to generate 10 SBA questions across 4 chatbot platforms. Second, assessors evaluated the chatbot outputs for consistency by identifying similarities and differences between users and across chatbots. With 3 assessors and 10 learning objectives, the maximum possible score for any individual chatbot was 30. Third, 7 assessors internally moderated the questions using a rating scale developed by the research team to evaluate scientific accuracy and educational quality.

Results: In response to the prompts, all chatbots generated 10 questions each, except Bing, which failed to respond to 1 prompt. ChatGPT-4 exhibited the highest variation in question generation but did not fully satisfy the "cover test." Gemini performed well across most evaluation criteria, except for item balance, and relied heavily on the vignette for answers but showed a preference for one answer option. Bing scored low in most evaluation areas but generated appropriately structured lead-in questions. SBA questions from GPT-3.5, Gemini, and ChatGPT-4 had similar Item Content Validity Index and Scale Level Content Validity Index values, while the Krippendorff alpha coefficient was low (0.016). Bing performed poorly in content clarity, overall validity, and item construction accuracy. A 2-way ANOVA without replication revealed statistically significant differences among chatbots and domains ($P<.05$). However, the Tukey-Kramer HSD (honestly significant difference) post hoc test showed no significant pairwise differences between individual chatbots, as all comparisons had P values >.05 and overlapping CIs.

Conclusions: AI chatbots can aid the production of questions aligned with learning objectives, and individual chatbots have their own strengths and weaknesses. Nevertheless, all require expert evaluation to ensure their suitability for use. Using AI to generate SBA prompts us to reconsider Bloom's taxonomy of the cognitive domain, which traditionally positions creation as the highest level of cognition.

(JMIR Med Educ 2025;11:e69521) doi:10.2196/69521

KEYWORDS

artificial intelligence; assessment; Bing; ChatGPT; Gemini; medical education; single best answer

Introduction

Across disciplines of education, including medical education, programmatic assessment offers flexible learning modalities that pave the road for individual progression. However, it represents a challenge to educators, as they are required to develop frequent assessments that reflect different competencies, thus necessitating the continuous creation of examination content in a comparatively restricted time [1]. For many years, multiple choice questions (MCQs) have been adopted in medical education for assessing knowledge and clinical reasoning skills

XSL•FO
RenderX

in high-stakes undergraduate and postgraduate medical exams. MCQs are reliable, objective, standardized, equitable, and efficient formats for testing large volumes of content in a limited time. A main problem with MCQs is that producing high-quality questions is time-consuming, from drafting the question that includes a clinical vignette or stem, a lead-in question, a correct answer, and distractors to validation of content and detection of potential flaws [1,2]. To tackle this dilemma, the application of many technological innovations, including artificial intelligence (AI), has been tried [3].

AI refers to machines mimicking the human brain in performing intellectual tasks. This originates from the imitation game developed by the British mathematician Alan Turing, who posed the universally famous question "Can machines think?" [4]. Since then, many AI research laboratories have invested time, effort, and money to answer this question. One particular AI research laboratory known as OpenAI, based in California, United States, has revolutionized our world at the end of 2022 by launching an AI-based large language model (LLM) software (GPT-3.5) that uses natural language processing to engage in human-like conversations and making it freely available for the public [5]. Within a few weeks after its release, the OpenAI chatbot, known as ChatGPT, had gained much attention in many fields, including medical education. It became the fastest-growing app of all time with more than 120 million users in just a few months after its launch [6]. This led competitors to develop and launch other chatbots. Microsoft launched Bing Chat AI in February 2023, followed by Google releasing Gemini in March 2023 [7]. A newer, improved version of ChatGPT (ChatGPT Plus), which uses the GPT-4 Turbo language model, has been developed by OpenAI and launched as a paid subscription version by the end of 2023 [6]

In terms of assessment in medical education, ChatGPT has been the most extensively studied chatbot. It was found to be able to quickly and accurately apply known concepts in medicine to novel problems, including reflection prompts and examination questions, and to mimic human writing styles, introducing a potential threat to the validity of traditional forms of medical student assessment including short answer assessment [8], it even successfully passed the USMLE (United States Medical Licensing Examination) [9]. Similarly, ChatGPT-4 was able to achieve a mean of more than 75% in the newly derived undergraduate medical exit examination: UKMLA (United Kingdom Medical Licensing Assessment) [10]. Its application has been described across multiple areas of academic assessment, for example, developing innovative assessments, grading submitted work, and providing feedback [11]. Nevertheless, concerns persist around the validity of the information provided by all AI tools. Sample [12] argued that if the chatbot response is not properly verified, it can be misleading and result in "junk science."

Additionally, the broad availability of LLMs such as ChatGPT, Gemini, and Bing has facilitated extensive comparative studies across various domains. For example, 1 study evaluated these models using case vignettes in physiology and found that ChatGPT-3.5 outperformed Bing and Google Bard (an old version of Gemini), indicating its superior effectiveness in case-based learning [13]. Another study, using the clinicopathological conferences method, compared the ability of AI chatbots to infer neuropathological diagnoses from clinical summaries. The findings revealed that Google Bard and ChatGPT-3.5 correctly diagnosed 76% of cases, while ChatGPT-4 achieved a higher accuracy rate, correctly identifying 84% of cases [14]. Similarly, a comparison of ChatGPT-3.5, Google Bard, and Microsoft Bing in hematology cases highlighted significant performance differences, with ChatGPT achieving the highest accuracy [15].

Recent studies have explored the use of AI in generating MCQs and single best answer (SBA) questions for medical examinations, highlighting its potential applications and limitations. For instance, Zuckerman et al [16] examined ChatGPT's role in assessment writing, while Kıyak et al [17] and Mistry et al [18] investigated AI-generated MCQs in pharmacotherapy and radiology board exams, respectively.

Despite these contributions, the ability of AI to generate valid SBA questions, an assessment format that better evaluates higher-order cognitive skills such as data interpretation, problem-solving, and decision-making [19], remains an area requiring further exploration. Additionally, a critical consideration is the variation in AI-generated outputs and the potential for examination candidates to predict examination items based on curriculum learning objectives (LOBs). Given the significance of these issues, this study aims to examine the content validity and consistency of different chatbots in generating SBAs for undergraduate medical education.

Methods

Study Context

The Graduate Entry Medical Programme at Ulster University's School of Medicine is a 4-year program. Similar to most UK medical schools, students undergo assessment through a series of SBA papers comprising over 1500 questions across the program. Managing this extensive assessment requirement has prompted the exploration of innovative solutions to support the assessment team.

To ensure assessment standards, the school has implemented a rigorous quality assurance process. Questions are first created by designated clinical or academic authors who have been trained and provided with a "house style" to follow. Questions then undergo internal review by other clinical or academic staff before external review by external examiners to ensure they meet rigorous requirements. Post hoc psychometric analysis of question performance is also used to drive evidence-based review and enhancement. This meticulous review process aims to uphold the integrity and effectiveness of assessments used to make high-stakes progression decisions and forms part of a wider suite of quality processes to deliver against the assessment strategy.

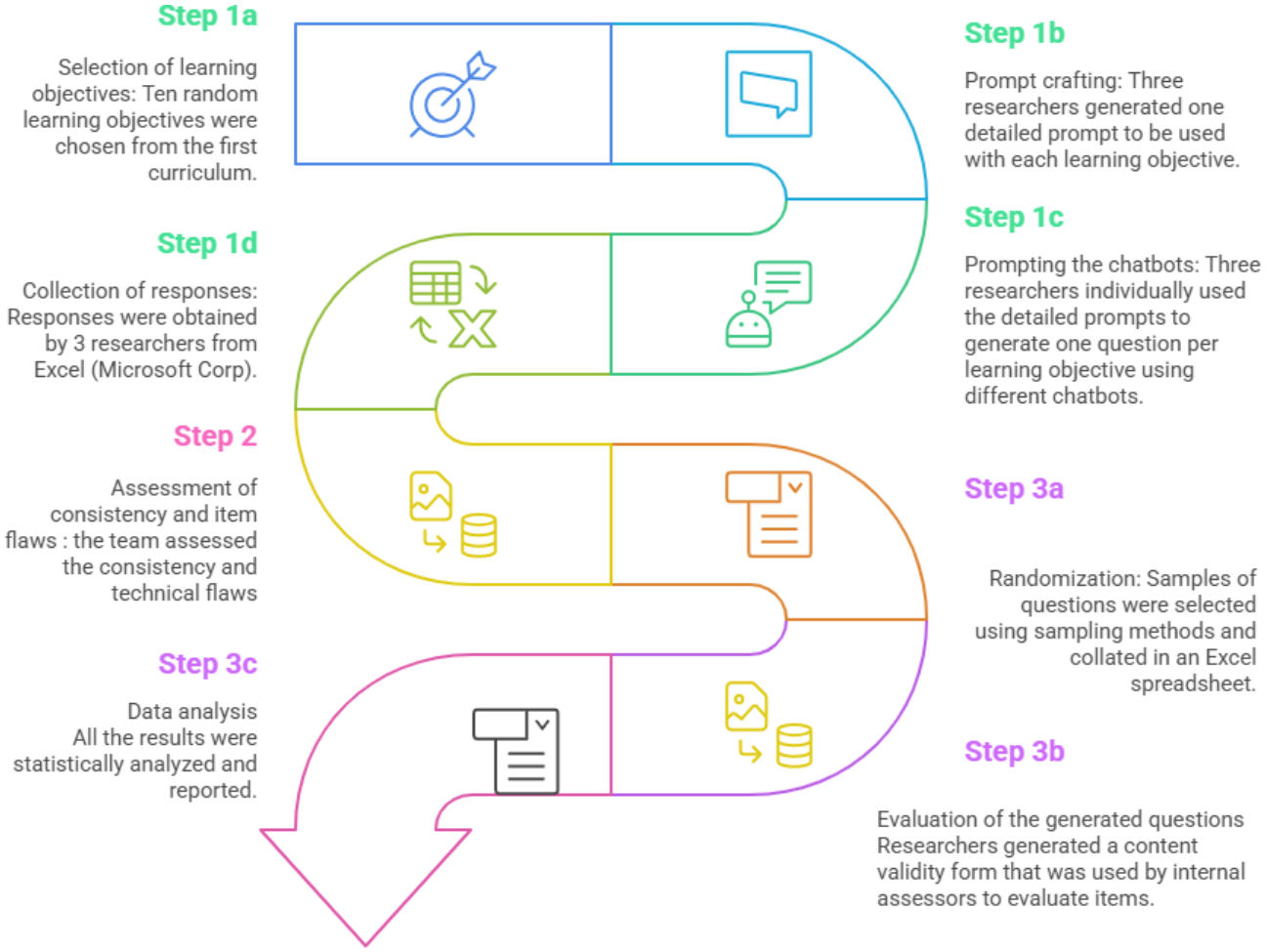
Study Design

This exploratory comparative study was conducted between December 2023 and May 2024; we continued to follow the school's established quality assurance process, but the designated first authors of the questions were AI chatbots. This includes 3 versions of AI chatbots: ChatGPT which will be

XSL•FO
RenderX

referred to as ChatGPT-3.5 in this study, Google Gemini, and Microsoft Bing AI, in addition to the subscription-only version of OpenAI: ChatGPT-4 that provides access to GPT-4 Turbo, which is advertised as a more powerful and faster version of GPT-4. During this study, Google changed the name of its platform from Bard to Gemini. For consistency, this paper will refer to the current name: Gemini. Figure 1 depicts the full study design, which included three main phases: (1) Generation of questions using various AI chatbots, (2) Assessment of the consistency of the chatbot outputs, and (3) Evaluation of the quality of the questions generated.

Figure 1. The study design. SBA: single best answer.



Generation of Questions Using Various AI Chatbots

In phase one, the research team randomly selected year 1 curriculum LOBs (n=10) to create SBA questions for. These objectives were selected using stratified random sampling from the official list of LOBs for second-semester educational units. Three researchers were involved, and each one created a new account for each of the 4 chatbot platforms. All researchers used the same predefined prompts (see below) around the same time (end of December 2023) to request 10 questions from each chatbot, one for each LOB. The 10 prompts were entered one by one in the same conversation with each chatbot. All the questions were compiled into a shared Microsoft Excel (Microsoft Corp) spreadsheet for analysis in steps 2 and 3.

To allow a fair comparison, the same prompt was used in each chatbot, which specified SBA features:

- You are a university lecturer in a UK medical school. Generate an MCQ on "the learning objective," with the following criteria:
  - The question is in a clinical vignette format.
  - The question is designed to assess the knowledge (±clinical judgment) of undergraduate medical students.
  - The question meets the standard for a medical graduate examination.
  - Five choices are allowed for each question.
  - Only 1 correct answer
  - Tag the correct answer.
  - Justify the correct answer.

Assessment of the Consistency and Quality (Item Flaws) of the Chatbot Outputs

In the second phase, researchers involved in the previous step assessed each chatbot's output consistency and technical flaws. Consistency was evaluated based on the similarity between the outputs generated across the 3 researchers, including any bias in the correct answer allocation (eg, favoring option "A" as the correct answer). Similarity was evaluated based on specific elements of the output and accordingly classified into one of three categories: (1) exact questions: when the outputs contain the same wording, condition, and lead-in question; (2) similar

XSL•FO
RenderX

questions: when the outputs share common elements such as patient characteristics, age, condition, presentation, or lead-in question; (3) different questions: when the outputs do not have any content in common.

Technical item flaws assessed the overall construct and structure of the questions produced by the chatbots using 7 previously published criteria for determining the quality of SBAs [20]. The 7 criteria include judgments on whether the questions: follow the SBA structural format, satisfy the "cover test" rule where the question should be answerable solely from the vignette or stem and lead-in (with the answers "covered"), test the application of knowledge rather than recall isolated facts, have item balance (which ensures a balance in information between the stem, lead-in, and options), tests 1 idea, are dependent on the vignette to reach the correct answer, and have appropriate lead-ins length. The researchers used a defined scale to evaluate how often or to what extent each criterion was met across the 3 researchers' outputs. Each criterion was scored on a scale from 0 to 3 for each of the 10 LOB prompts. In this scale, 0 meant none, 1 meant 1 SBA, 2 meant 2 SBAs, and 3 meant all 3 SBAs, representing the number of questions produced by each chatbot that met the criterion. With 3 assessors and 10 LOBs, the maximum possible score for any individual chatbot was 30.

Assessment of the Content Validity and Accuracy of the Questions Generated

In phase 3, samples of questions generated by the chatbots were distributed to various internal assessors as per our normal quality review process. The questions were selected using stratified random sampling to select 1 of the 3 questions generated by each chatbot for each LOB, yielding a total of 39 questions. Alongside this, a content validation evaluation form, developed by the research team, was used to ensure consistent review between assessors, providing assessors with clear expectations and an understanding of the task. The assessors are faculty members with expertise in the curriculum content. Each question was evaluated by 7 assessors.

Considering published recommendations for content validation [21,22], 20 internal assessors were invited, of which 7 consented to participate. The internal assessors critically reviewed the questions based on several criteria to ensure their quality and alignment with educational objectives. This includes content clarity and validity; accuracy of information, answers, and justification; and educational accuracy. Each of these elements was scored on a Likert scale of 1 to 4 (with 1 representing the lowest level of construct and 4 the highest level of the construct; Multimedia Appendix 1).

Statistical Analysis

Quantitative data was analyzed through scores obtained from the rating scale using IBM SPSS Statistics (version 26; IBM Corp). Subsequently, 2 content validity indexes were computed: the Item Content Validity Index (I-CVI) and the Scale Level Content Validity Index (S-CVI). Percentages and frequencies were calculated for the questions' scores to provide further insights into the data. A 2-way ANOVA without replication was conducted to assess differences in chatbot performance across 6 domains. Post hoc comparisons were performed using the Tukey-Kramer HSD (honestly significant difference) test to identify specific group differences. The average ratings provided by 7 evaluators were used for each chatbot and each criterion. The Krippendorff alpha [23] was used to assess interrater reliability, using the K-Alpha Calculator [24]. A coefficient value of 0.8 is considered satisfactory [23]. However, the low Krippendorff alpha suggested a need for further refinement of the rating scheme or additional training for raters to improve reliability.

Ethical Considerations

Participants were informed that their responses would be anonymized and that they could withdraw from this study at any point without penalty. Informed consent was obtained from all participants before data collection. Only those who provided explicit consent were included in this study. This study received ethical approval from the Ulster University Centre for Higher Education Research and Practice Ethics Committee and the Learning Enhancement Directorate Ethics Filter Committee (LEDEC; formerly CHERP; LEDEC-24-004). All data were anonymized during the analysis phase to ensure confidentiality and to protect participants' identities. Staff members who chose not to participate experienced no disadvantage or impact on their professional standing. No financial or material compensation was offered to participants for their involvement in this research.

Results

Generation of Questions

In response to the predefined prompts provided to the chatbots, 3 of them (free ChatGPT, ChatGPT Plus, and Gemini) generated 10 questions each, for a total of 30 across the 3 researchers. Bing could not respond to the prompt for LOB9 and thus generated 9 questions, for a total of 27 across the 3 researchers. Thus, 117 questions were generated (Multimedia Appendix 2).

Assessment of Consistency Within Chatbots and Technical Item Flaws Among the Outputs

Consistency within chatbots was evaluated based on the similarity of outputs between the 3 researchers and any bias in the allocation of the correct answer option. Bing had the highest degree of similarity between items generated by multiple users (4 exact question matches and 20 similar ones), while ChatGPT-4 had the highest degree of variation (Table 1).

Table . Similarity between the questions generated by different chatbots.

	Gemini (N=30), n (%)	Bing (N=27), n (%)	ChatGPT-3.5 (N=30), n (%)	ChatGPT-4 (N=30), n (%)
Exact questions	0 (0)	4 (14.81)	2 (6.67)	0 (0)
Similar questions	24 (80)	20 (74.07)	22 (73.33)	22 (73.33)
Different questions	6 (20)	3 (11.11)	6 (20)	8 (26.67)

XSL•FO
RenderX

The original predefined prompt did not request answer options to be given in any particular order. Therefore, for assessing potential bias in the correct answer allocation, 3 scenarios were modeled (Table 2):

- Any bias or preference in the correct answer allocation based on the raw chatbot output.
- Any bias or preference in the correct answer allocation based on the chatbot output when the researchers manually ordered answers into alphabetical order.
- Any bias or preference in the correct answer allocation based on a new output, where each chatbot was prompted to produce 30 new SBA questions with answers alphabetically.

Table . Assessment of possible bias or preference in correct answer allocation.

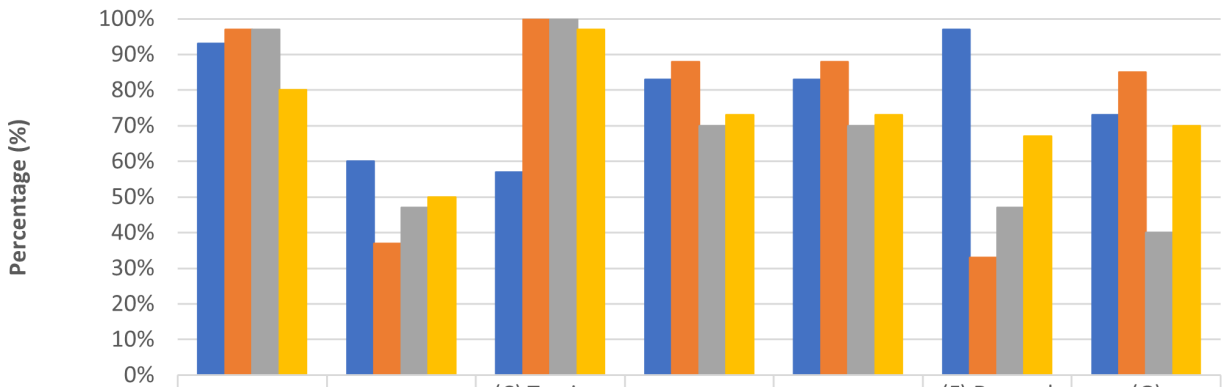
Options	Gemini (N=30), n (%)	Bing (N=27), n (%)	ChatGPT-3.5 (N=30), n (%)	ChatGPT-4 (N=30), n (%)
Original chatbot output				
A	5 (16.67)	6 (22.22)	9 (30)	11 (36.67)
B	12 (40)	4 (14.81)	10 (33.33)	10 (33.33)
C	6 (20)	10 (37.04)	7 (23.33)	4 (13.33)
D	5 (16.67)	6 (22.22)	3 (10)	4 (13.33)
E	2 (6.67)	1 (3.7)	1 (3.33)	1 (3.33)
Manual reordering of chatbot output into alphabetical order				
A	4 (13.33)	8 (29.63)	8 (26.67)	6 (20)
B	10 (33.33)	3 (11.11)	3 (10)	7 (23.33)
C	3 (10)	5 (18.52)	7 (23.33)	5 (16.67)
D	9 (30)	4 (14.81)	6 (20)	5 (16.67)
E	4 (13.33)	7 (25.93)	6 (20)	7 (23.33)

Gemini, ChatGPT-3.5, and ChatGPT-4 occasionally provided answer options in alphabetical order when not specifically prompted. Gemini consistently demonstrated a preference for the correct answer to be listed as option B. The ChatGPT-3.5 and ChatGPT-4 appeared to favor options A, B, and C. Bing appeared to favor options A and E.

Regarding the technical item flaws among the outputs, the chatbots performed similarly in terms of following an SBA format (Figure 2A) and achieving the “cover test” satisfaction (Figure 2B), although ChatGPT-4 scored slightly lower on satisfying the cover test. Overall, Gemini performed well across most items, except for item balance. Notably, Gemini stood out by creating questions with a lead-in that relied heavily on the vignette for the answer (Figure 2F). Bing scored low across most evaluation items but performed well in generating a lead-in question of appropriate length (Figure 2G). ChatGPT Plus, which required a paid subscription, did not outperform the other chatbots in any item. The evaluation item “questions test the application of knowledge rather than recall of isolated facts” received the lowest scores across all the chatbots (Figure 2C), with Gemini achieving the highest score among them.

XSL•FO
RenderX

Figure 2. Shows technical item flaws among the chatbots: (**A**) single best answer format, (**B**) satisfy the "cover test" rule, (**C**) test the application of knowledge rather than recall isolated facts, (**D**) questions were balanced, (**E**) lead-in question tests one idea, (**F**) questions depend on the vignette to reach an answer, and (**G**) appropriate lead-in question length. The total number of questions generated by Bing was 27.



	(A) Using single-best answer format	(B) Satisfy the "cover test" rule	(C) Testing the application of knowledge	(D) Balanced structure	(E) Test a single concept	(F) Depend on the vignette to reach an answer	(G) Appropriate lead-in question length
Gemini	93%	60%	57%	83%	83%	97%	73%
Bing	97%	37%	100%	88%	88%	33%	85%
ChatGPT-3.5	97%	47%	100%	70%	70%	47%	40%
ChatGPT-4	80%	50%	97%	73%	73%	67%	70%

Assessment of Content Validity and Accuracy

Seven internal assessors evaluated item clarity and relevance, deriving the I-CVI for individual SBA items and the S-CVI (following the Universal Agreement method) to assess the overall content validity for questions from each chatbot (Table 3). Items with I-CVI>0.79 and scales with S-CVI/UA>0.8 can be interpreted as acceptable [20].

Assessors also evaluated items for content clarity and 4 elements of accuracy: vignette information, answers, justifications, and educational accuracy, on a scale from 1 to 4 (Tables 4 and 5). The Krippendorff alpha coefficient was low, 0.016, with a 95% bootstrap CI of −0.066 to 0.116.

As depicted in Tables 3 and 4, SBA questions from 3 chatbots (ChatGPT, Gemini, and ChatGPT Plus) had similar content clarity and S-CVI values. In comparison to the other chatbots, Bing performed worst in content clarity, overall (scale) validity, and all elements of item accuracy. ChatGPT Plus, which required a paid subscription, did not outperform the other chatbots except in the measure of educational accuracy. Further statistical analysis was performed using the 2-way ANOVA without replication, which showed statistically significant differences among chatbots and domains ($P<.05$). However, the Tukey-Kramer HSD post hoc test revealed no significant pairwise differences between individual chatbots, as all comparisons had P values>.05 and overlapping CIs. Thus, although the chatbots' performance varied overall, specific chatbot differences were not statistically significant.

XSL•FO
RenderX

Table . Item-content validity and scale-content validity across the chatbots.

Item number	Gemini	Bing	ChatGPT-3.5	ChatGPT-4
I-CVI[a]				
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1
5	0.85	0.85	0.85	0.83
6	0.85	0.85	0.71	0.85
7	0.85	0.85	0.85	0.85
8	0.85	0.85	0.85	0.85
9	0.85	—[b]	0.85	0.85
10	0.85	0.85	0.85	0.85
S-CVI/UA[c]	0.91	0.83	0.9	0.91

[a]I-CVI: Item Content Validity Index.

[b]Not applicable.

[c]S-CVI/UA: Scale Level Content Validity Index.

Table . Average score for content clarity and accuracy of items across the chatbots.

	Content clarity[a]	Accuracy of information[b]	Accuracy of answers[c]	Accuracy of justification[d]	Educational accuracy[e]
Gemini	3.68	3.71	3.8	3.91	3.49
Bing	3.41	3.3	3.49	3.47	3.2
ChatGPT-3.5	3.75	3.71	3.84	3.9	3.5
ChatGPT-4	3.71	3.66	3.81	3.82	3.56

[a]Content clarity refers to the extent to which the question is clearly written, free of ambiguity, and easily understood by the intended audience.

[b]Accuracy of information verifies that the facts, concepts, and explanations presented are scientifically and contextually correct.

[c]Accuracy of answers ensures that the correct response is indeed accurate, while the distractors remain plausible yet distinguishable.

[d]Accuracy of justification evaluates whether the rationale provided for correct and incorrect answers is logically sound, evidence-based, and supports a deeper understanding of the topic.

[e]Educational accuracy assesses whether the question is appropriately challenging to the student level, measures higher cognitive levels (such as application or analysis), and adheres to best practices in assessment design.

Table . Two-way ANOVA table.

Source of variation	Sum of squares due to the source	*df*	Mean sum of squares due to the source	*F* test	*P* value
Average content clarity and accuracy scores	0.304357	2	0.152178	24.26587	<.001
Chatbots	17.9744	4	4.493601	716.5349	<.001
Error	0.05017	8	0.006271	—[a]	—
Total	18.32893	14	—	—	—

[a]Not applicable.

Discussion

Interpretation of Findings

This study was designed to examine the content validity and consistency of SBA questions generated by different chatbots in the context of undergraduate medical education. The findings revealed that no single chatbot excelled in all studied domains nor demonstrated a universal superiority over other chatbots, but rather showed unique strengths of some chatbots in specific areas and highlighted their notable limitations in other ones. This emphasizes the importance of critically assessing the output

XSL•FO
RenderX

of chatbots in a context-sensitive manner. Bing produced items that were least suitable for inclusion in medical student assessment. These findings echo previous studies, which also show Bing to generate less valid MCQs in comparison to other chatbots [25]. ChatGPT-4 showed the greatest variation in responses across users (suggesting higher protection against examination candidates predicting potential assessment items), and had strong performance in content clarity and accuracy, though it also exhibited some less effective question design practices, such as poorer performance in the "cover test" rule. These findings align with the results of Doughty et al [26], who found that GPT-4's ability to generate effective MCQs was nearly on par with human performance, in which 81.7% of the generated MCQs met all evaluation criteria, suggesting that fewer than 1 in 5 questions would need revision by instructors. However, in cases where ChatGPT-4 failed to meet a quality standard, this was typically the only issue with the question. Gemini performed well across all evaluations, matching ChatGPT Plus's strong index score for content validity, and excelled in creating questions where the lead-in tested 1 item and relied heavily on the vignette for the answer. Although slightly behind both ChatGPT versions in content clarity, Gemini scored the highest in providing accurate justifications for the correct answer.

This variation across chatbots is consistent with results from studies where chatbots were asked to answer questions. Kumari et al [15] found significant differences in solving hematology case vignettes using LLMs. ChatGPT achieved the highest score, followed by Google Gemini and then Microsoft Bing. In line with this, Dhanvijay et al [13] reported that ChatGPT-3.5 scored the highest, Bing the lowest, and Bard (Gemini) ranked in the middle when solving case vignettes in physiology. When chatbots were tested on their ability to answer SBA questions, ChatGPT-4 and Microsoft Copilot (Bing) outperformed Google Gemini [27]. Overall, these results suggest that OpenAI's ChatGPT shows strong potential in the medical education field. However, it is worth noting that none of the models were able to answer all questions correctly, and in our study, all platforms had some flaws when generating SBAs.

Additionally, this study's results reveal several key insights and revelations concerning SBA questions produced by AI chatbots. First, we observed that chatbots often exhibit a correct answer bias toward particular options. Recent studies have identified that LLMs tend to display positional bias when handling MCQs [28,29]. Radford et al [30] and Li and Gao [31] found that this susceptibility to positional bias is pronounced in the GPT-2 family however a more recent technical report for GPT-4 suggests AI's performance in MCQ remains susceptible to the position of the correct answer among the choices [32], a pattern referred to as "anchored bias." To minimize this inherent bias that appears to occur across AI platforms, when using AI to generate MCQ or SBA, we would recommend not stipulating an order for answer options in the prompt.

Furthermore, assessment literature emphasizes that high-quality SBA questions should assess the higher levels of Bloom's taxonomy to encourage students' critical thinking and complex problem-solving [33]. Our study revealed that chatbots were not always successful in crafting questions that engaged these advanced cognitive levels, and this was an area of relative weakness when evaluating items. Gemini scored highest, followed by ChatGPT Plus, ChatGPT-3.5, and then Bing. Similar findings regarding ChatGPT's limitations were reported by Herrmann-Werner et al [34]. Likewise, studies by Klang et al [35] and Liu et al [36] also emphasized GPT-4's limited ability to integrate knowledge and apply clinical reasoning, highlighting challenges in logical reasoning, which could limit AI's ability to generate questions that test this concept. However, it should be noted that while human-written questions were rated higher in direct comparisons, the score gap was narrow and largely insignificant, suggesting that AI tools still hold potential as educational aids [2].

Our analysis also revealed some technical flaws, variations, and inconsistencies in item construction within all chatbots. These flaws highlight instances of overconfidence and inadequacies in question design, suggesting an inability of the chatbots to evaluate their output's consistency, relevance, and complexity. Flawed MCQs hinder the accurate and meaningful interpretation of test scores and negatively impact student pass rates. Therefore, identifying and addressing technical flaws in MCQs can enhance their quality and reliability [37]. Similarly, Klang et al [35] reported that approximately 15% of questions generated using detailed prompts required corrections, primarily due to content inaccuracies or methodological shortcomings. These revisions often involved addressing a lack of sensitivity in certain topics, such as failing to include specific details such age, gender, or geographical context in the questions or answers.

Most of the questions tested recall and comprehension levels, but Gemini included some that assessed the application of knowledge. In contrast, Bing struggled to generate questions on specific topics. These findings can be explained as critical thinking at higher levels involves considering evidence, context, conceptualization, methods, and the criteria required for judgment [38]. AI models are trained on large datasets of text, but they may not fully understand the context or underlying concepts behind the content. Higher-order thinking skills, such as application, analysis, and synthesis, require deeper comprehension and reasoning that AI might not be able to simulate effectively.

Thus, using AI to generate SBAs encourages us to reconsider Bloom's taxonomy of the cognitive domains [39,40], which traditionally positions "creation" as the highest level of cognition. In the era of AI, evaluation might be considered the most critical level of cognition [41]. While AI chatbots can often produce well-written questions aligned with LOBs, they still require expert evaluation to ensure their suitability for use. Future research should compare AI-generated outputs with those from subject matter experts to assess accuracy and relevance. Evaluating AI's ability to test higher-order cognition in Bloom's taxonomy is also crucial. As AI evolves, ongoing validation is essential to ensure reliability and effectiveness in assessments.

Despite the methodological rigor and innovative approach of this study, some limitations need to be highlighted to improve the interpretation of the findings presented here. First, the researchers or assessors generated or evaluated only 30 questions per chatbot. Variation was observed in the content validity and

XSL•FO
RenderX

accuracy between the SBAs produced by an individual chatbot. Therefore, this sample may not sufficiently represent the wide range of possible outputs, potentially limiting the generalizability and robustness of the findings. Second, the accuracy of the chatbots' responses may have been compromised by the absence of reference materials, which could have negatively affected their performance. Finally, this study is limited by low interrater reliability and the use of measures are not specifically designed to assess MCQ quality. Future research should consider using validated tools to enhance evaluation accuracy.

Conclusions

Chatbot platforms varied in their ability to generate educational questions. ChatGPT models produced the most variable outputs, reducing predictability while maintaining strong content clarity and accuracy with minimal answer bias. Gemini performed similarly but showed a strong preference for 1 option, while Bing had the least variation and the lowest content clarity and accuracy. ChatGPT-4 did not significantly improve question quality but maximized variability. Technical flaws were present across all platforms, with many questions poorly linked to vignettes. Most tested recall and comprehension, though Gemini included some application-level items, whereas Bing struggled with specific topics.

These findings highlight AI's limitations in generating higher-order thinking questions, reinforcing the need for expert evaluation. This challenges Bloom's taxonomy's traditional cognitive hierarchy, suggesting that "evaluation" may be more critical than "creation" in AI-assisted assessments.

Acknowledgments

The authors extend their gratitude to the internal assessors from the School of Medicine, Ulster University, who dedicated their time to evaluating the questions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Further data on the assessment of questions generated
[XLSX File, 35 KB - mededu_v11i1e69521_app1.xlsx]

Multimedia Appendix 2
Questions generated.
[XLSX File, 85 KB - mededu_v11i1e69521_app2.xlsx]

References

1. Pugh D, De Champlain A, Gierl M, Lai H, Touchie C. Using cognitive models to develop quality multiple-choice questions. Med Teach 2016 Aug;38(8):838-843. [doi: 10.3109/0142159X.2016.1150989] [Medline: 26998566]
2. Cheung BHH, Lau GKK, Wong GTC, et al. ChatGPT versus human in generating medical graduate exam multiple choice questions-a multinational prospective study (Hong Kong S.A.R., Singapore, Ireland, and the United Kingdom). PLoS One 2023;18(8):e0290691. [doi: 10.1371/journal.pone.0290691] [Medline: 37643186]
3. Rodriguez-Torrealba R, Garcia-Lopez E, Garcia-Cabot A. End-to-end generation of multiple-choice questions using text-to-text transfer transformer models. Expert Syst Appl 2022 Dec;208:118258. [doi: 10.1016/j.eswa.2022.118258]
4. Turing AM. I.—computing machinery and intelligence. Mind 1950 Oct 1;LIX(236):433-460. [doi: 10.1093/mind/LIX.236.433]
5. Rudolph J, Tan S, Tan S. ChatGPT: bullshit spewer or the end of traditional assessments in higher education? JALT 2023;6(1):342-363 [FREE Full text] [doi: 10.37074/jalt.2023.6.1.9]
6. Rudolph J, Tan S, Tan S. War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. JALT 2023;6(1):364-389 [FREE Full text] [doi: 10.37074/jalt.2023.6.1.23]
7. Giannakopoulos K, Kavadella A, Stamatopoulos V, Kaklamanos EG, Salim AA. Evaluation of the performance of generative AI large language models ChatGPT, Google Bard, and Microsoft Bing chat in supporting evidence-based dentistry: comparative mixed methods study. J Med Internet Res 2023 Dec 28;25:e51580. [doi: 10.2196/51580] [Medline: 38009003]
8. Morjaria L, Burns L, Bracken K, et al. Examining the threat of ChatGPT to the validity of short answer assessments in an undergraduate medical program. J Med Educ Curric Dev 2023;10:23821205231204178. [doi: 10.1177/23821205231204178] [Medline: 37780034]
9. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 8;9:e45312. [doi: 10.2196/45312] [Medline: 36753318]
10. Lai UH, Wu KS, Hsu TY, Kan JKC. Evaluating the performance of ChatGPT-4 on the United Kingdom Medical Licensing Assessment. Front Med (Lausanne) 2023;10:1240915. [doi: 10.3389/fmed.2023.1240915] [Medline: 37795422]

XSL•FO
RenderX

11. O'Connor S. Open artificial intelligence platforms in nursing education: tools for academic progress or abuse? Nurse Educ Pract 2023 Jan;66:103537. [doi: 10.1016/j.nepr.2022.103537] [Medline: 36549229]
12. Sample I. Science journals ban listing of ChatGPT as co-author on papers. The Guardian. 2023. URL: https://www.theguardian.com/science/2023/jan/26/science-journals-ban-listing-of-chatgpt-as-co-author-on-papers [accessed 2025-05-14]
13. Dhanvijay AKD, Pinjar MJ, Dhokane N, Sorte SR, Kumari A, Mondal H. Performance of large language models (ChatGPT, Bing Search, and Google Bard) in solving case vignettes in physiology. Cureus 2023 Aug;15(8):e42972. [doi: 10.7759/cureus.42972] [Medline: 37671207]
14. Koga S, Martin NB, Dickson DW. Evaluating the performance of large language models: ChatGPT and Google Bard in generating differential diagnoses in clinicopathological conferences of neurodegenerative disorders. Brain Pathol 2024 May;34(3):e13207. [doi: 10.1111/bpa.13207] [Medline: 37553205]
15. Kumari A, Kumari A, Singh A, et al. Large language models in hematology case solving: a comparative study of ChatGPT-3.5, Google Bard, and Microsoft Bing. Cureus 2023 Aug;15(8):e43861. [doi: 10.7759/cureus.43861] [Medline: 37736448]
16. Zuckerman M, Flood R, Tan RJB, et al. ChatGPT for assessment writing. Med Teach 2023 Nov;45(11):1224-1227. [doi: 10.1080/0142159X.2023.2249239] [Medline: 37789636]
17. Kıyak YS, Coşkun Ö, Budakoğlu I, Uluoğlu C. ChatGPT for generating multiple-choice questions: evidence on the use of artificial intelligence in automatic item generation for a rational pharmacotherapy exam. Eur J Clin Pharmacol 2024 May;80(5):729-735. [doi: 10.1007/s00228-024-03649-x] [Medline: 38353690]
18. Mistry NP, Saeed H, Rafique S, Le T, Obaid H, Adams SJ. Large language models as tools to generate radiology board-style multiple-choice questions. Acad Radiol 2024 Sep;31(9):3872-3878. [doi: 10.1016/j.acra.2024.06.046] [Medline: 39013736]
19. Tan LT, McAleer JJA, Final FRCR Examination Board. The introduction of single best answer questions as a test of knowledge in the final examination for the fellowship of the Royal College of Radiologists in Clinical Oncology. Clin Oncol (R Coll Radiol) 2008 Oct;20(8):571-576. [doi: 10.1016/j.clon.2008.05.010] [Medline: 18585017]
20. Case SM, Swanson DB. Writing one-best-answer questions for the basic and clinical sciences. In: Constructing Written Test Questions for the Basic and Clinical Sciences: National Board of Medical Examiners; 2016:31-66.
21. Yusoff MSB. ABC of content validation and content validity index calculation. EIMJ 2019;11(2):49-54 [FREE Full text] [doi: 10.21315/eimj2019.11.2.6]
22. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med 2006 Feb;119(2):166. [doi: 10.1016/j.amjmed.2005.10.036] [Medline: 16443422]
23. Krippendorff K. Content Analysis: An Introduction to Its Methodology, 4th edition: SAGE Publications; 2019. [doi: 10.4135/9781071878781]
24. Marzi G, Balzano M, Marchiori D. K-Alpha Calculator-Krippendorff's Alpha Calculator: a user-friendly tool for computing Krippendorff's Alpha inter-rater reliability coefficient. MethodsX 2024 Jun;12:102545. [doi: 10.1016/j.mex.2023.102545] [Medline: 39669968]
25. Agarwal M, Sharma P, Goswami A. Analysing the applicability of ChatGPT, Bard, and Bing to generate reasoning-based multiple-choice questions in medical physiology. Cureus 2023 Jun;15(6):e40977. [doi: 10.7759/cureus.40977] [Medline: 37519497]
26. Doughty J, Wan Z, Bompelli A, et al. A comparative study of AI-generated (GPT-4) and human-crafted MCQs in programming education. Presented at: ACE 2024; Jan 29 to Feb 2, 2024; Sydney, New South Wales, Australia p. 114-123 URL: https://dl.acm.org/doi/proceedings/10.1145/3636243 [accessed 2025-05-14] [doi: 10.1145/3636243.3636256]
27. Rossettini G, Rodeghiero L, Corradi F, et al. Comparative accuracy of ChatGPT-4, Microsoft Copilot and Google Gemini in the Italian entrance test for healthcare sciences degrees: a cross-sectional study. BMC Med Educ 2024 Jun 26;24(1):694. [doi: 10.1186/s12909-024-05630-9] [Medline: 38926809]
28. Pezeshkpour P, Hruschka E. Positional bias in large language models when handling multiple-choice questions. arXiv. Preprint posted online on Aug 22, 2023 URL: https://arxiv.org/abs/2308.11483 [accessed 2025-05-14] [doi: 10.48550/arXiv.2308.11483]
29. Zheng J, Li X, Wang R. Investigating option position biases in large language models. arXiv. Preprint posted online on Sep 7, 2024 URL: https://arxiv.org/abs/2309.03882 [accessed 2025-05-14] [doi: 10.48550/arXiv.2309.03882]
30. Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners. OpenAI Blog 2019;1(8):9 [FREE Full text]
31. Li R, Gao Y. Anchored answers: unravelling positional bias in GPT-2's multiple-choice questions. arXiv. Preprint posted online on May 6, 2024 URL: https://arxiv.org/abs/2405.03205 [accessed 2025-05-14] [doi: 10.48550/arXiv.2405.03205]
32. Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, Aleman FL, et al. GPT-4 technical report. arXiv. Preprint posted online on Mar 15, 2023 URL: https://arxiv.org/abs/2303.08774 [accessed 2025-05-14] [doi: 10.48550/arXiv.2303.08774]
33. Walsh JL, Harris BHL, Smith PE. Single best answer question-writing tips for clinicians. Postgrad Med J 2017 Feb 1;93(1096):76-81. [doi: 10.1136/postgradmedj-2015-133893]
34. Herrmann-Werner A, Festl-Wietek T, Holderried F, et al. Assessing ChatGPT's mastery of Bloom's taxonomy using psychosomatic medicine exam questions: mixed-methods study. J Med Internet Res 2024 Jan 23;26:e52113. [doi: 10.2196/52113] [Medline: 38261378]

XSL•FO
RenderX

35. Klang E, Portugez P, Gross R, et al. Advantages and pitfalls in utilizing artificial intelligence for crafting medical examinations: a medical education pilot study with GPT-4. BMC Med Educ 2023 Oct 17;23(1):772. [doi: 10.1186/s12909-023-04752-w] [Medline: 37848913]
36. Liu H, Ning R, Teng Z, Liu J, Zhou Q, Zhang Y. Evaluating the logical reasoning ability of ChatGPT and GPT-4. arXiv. Preprint posted online on Apr 7, 2023 URL: https://arxiv.org/abs/2304.03439 [accessed 2025-05-14] [doi: 10.48550/arXiv.2304.03439]
37. Khan HF, Danish KF, Awan AS, Anwar M. Identification of technical item flaws leads to improvement of the quality of single best multiple choice questions. Pak J Med Sci 2013 May;29(3):715-718. [doi: 10.12669/pjms.293.2993] [Medline: 24353614]
38. Dergaa I, Chamari K, Zmijewski P, Ben Saad H. From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. Biol Sport 2023 Apr;40(2):615-622. [doi: 10.5114/biolsport.2023.125623] [Medline: 37077800]
39. Krathwohl DR. A revision of Bloom's taxonomy: an overview. Theory Pract 2002 Nov 1;41(4):212-218. [doi: 10.1207/s15430421tip4104_2]
40. Tutkun OF, Güzel G, Köroğlu M, Ilhan H. Bloom's revised taxonomy and critics on it. Online J Couns Educ 2012;1(3):23-30.
41. Scheuer-Larsen C, Lauridsen PS. Bloom's taxonomy in the interaction between artificial intelligence and human learning. Viden.AI. URL: https://viden.ai/en/blooms-taxonomy-and-ai [accessed 2025-05-14]

Abbreviations

AI: artificial intelligence
HSD: honestly significant difference
I-CVI: Item Content Validity Index
LLM: large language model
LOB: learning objective
MCQ: multiple choice question
S-CVI: Scale Level Content Validity Index
SBA: single best answer
UKMLA: United Kingdom Medical Licensing Assessment
USMLE: United States Medical Licensing Examination

Edited by A Bahattab, B Lesselroth; submitted 02.12.24; peer-reviewed by U Hin Lai, YS Kiyak; revised version received 22.04.25; accepted 30.04.25; published 30.05.25.

Please cite as:
Abouzeid E, Wassef R, Jawwad A, Harris P
Chatbots' Role in Generating Single Best Answer Questions for Undergraduate Medical Student Assessment: Comparative Analysis
JMIR Med Educ 2025;11:e69521
URL: https://mededu.jmir.org/2025/1/e69521
doi:10.2196/69521

© Enjy Abouzeid, Rita Wassef, Ayesha Jawwad, Patricia Harris. Originally published in JMIR Medical Education (https://mededu.jmir.org), 30.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

ChatGPT's Performance on Portuguese Medical Examination Questions: Comparative Analysis of ChatGPT-3.5 Turbo and ChatGPT-4o Mini

Filipe Prazeres[1,2,3], MD, MSc, PhD

[1]Faculty of Health Sciences, University of Beira Interior, Av. Infante D. Henrique, Covilhã, Portugal

[2]Family Health Unit Beira Ria, Gafanha da Nazaré, Portugal

[3]CINTESIS@RISE, Department of Community Medicine, Information and Health Decision Sciences, Faculty of Medicine of the University of Porto, Porto, Portugal

Corresponding Author:
Filipe Prazeres, MD, MSc, PhD
Faculty of Health Sciences, University of Beira Interior, Av. Infante D. Henrique, Covilhã, Portugal

Abstract

Background: Advancements in ChatGPT are transforming medical education by providing new tools for assessment and learning, potentially enhancing evaluations for doctors and improving instructional effectiveness.

Objective: This study evaluates the performance and consistency of ChatGPT-3.5 Turbo and ChatGPT-4o mini in solving European Portuguese medical examination questions (2023 National Examination for Access to Specialized Training; Prova Nacional de Acesso à Formação Especializada [PNA]) and compares their performance to human candidates.

Methods: ChatGPT-3.5 Turbo was tested on the first part of the examination (74 questions) on July 18, 2024, and ChatGPT-4o mini on the second part (74 questions) on July 19, 2024. Each model generated an answer using its natural language processing capabilities. To test consistency, each model was asked, "Are you sure?" after providing an answer. Differences between the first and second responses of each model were analyzed using the McNemar test with continuity correction. A single-parameter *t* test compared the models' performance to human candidates. Frequencies and percentages were used for categorical variables, and means and CIs for numerical variables. Statistical significance was set at $P<.05$.

Results: ChatGPT-4o mini achieved an accuracy rate of 65% (48/74) on the 2023 PNA examination, surpassing ChatGPT-3.5 Turbo. ChatGPT-4o mini outperformed medical candidates, while ChatGPT-3.5 Turbo had a more moderate performance.

Conclusions: This study highlights the advancements and potential of ChatGPT models in medical education, emphasizing the need for careful implementation with teacher oversight and further research.

(JMIR Med Educ 2025;11:e65108) doi:10.2196/65108

KEYWORDS

ChatGPT-3.5 Turbo; ChatGPT-4o mini; medical examination; European Portuguese; AI performance evaluation; Portuguese; evaluation; medical examination questions; examination question; chatbot; ChatGPT; model; artificial intelligence; AI; GPT; LLM; NLP; natural language processing; machine learning; large language model

Introduction

Generative artificial intelligence (AI) represents a branch of AI dedicated to the development of systems that can autonomously generate high-quality digital content on demand, and it can do so across various modalities, such as written text, images, audio, and video [1-3]. Generative AI tools are trained on large datasets, enabling them to produce work that mirrors human-created content [2]. Nowadays, there are several examples of generative AI tools, including ChatGPT (OpenAI Inc), Runway, Gemini (Google Inc), DALL-E (OpenAI Inc), Copilot (Microsoft Inc), Midjourney, NovelAI (Anlatan), Claude (Anthropic), and Jasper AI, among others. ChatGPT, the large language model (LLM) chatbot, developed by OpenAI [4], that started the AI boom in November 2022, became the most popular AI tool of 2023, accounting for over 60.2% of visits between September 2022 and August 2023, with a total of 14.6 billion website visits [5]. ChatGPT's availability as a free-to-use, low-bandwidth service may reduce disparities compared to paid versions or models by making advanced AI technology accessible to a broader and more diverse global population [6], contributing to making it the most popular generative AI tool [7].

Recent literature reviews regarding AI have shown that this type of technology has potential applications in several fields, spanning from the architecture, engineering, and construction industry to health care [8-11]. The possible applications in

XSL•FO
RenderX

medicine are substantial, ranging from diagnostic and treatment support (eg, clinical imaging improvement, classification of diseases, prediction of disease onset, development of treatment, and medication prescriptions) [12] to facilitate communication and engagement between medical professionals and their patients [13], and also improving medical education and its accessibility [10,14,15]. For example, ChatGPT can be used as a study tool to clearly explain complex medical concepts [16,17] (eg, radiology reports [18]), create memory aids for challenging topics, clarify medical practice questions, summarize research articles, compile lists of differential diagnoses [17], generate medical examination questions [19], and simulate physician-patient interactions [14].

Medical written examinations are an important part in evaluating the competence and knowledge of medical students and graduates (eg, access of physicians to specialized training, such is the case in Portugal). These examinations not only test factual knowledge but also evaluate the critical thinking and problem-solving skills of the candidates. With the recent growing interest in AI, an important question arises: Can AI, specifically ChatGPT, perform at a level comparable to human candidates in medical written examinations? By evaluating ChatGPT's ability to correctly answer medical questions, its medical proficiency and its potential role as an educational tool can be assessed. Successfully completing this task can demonstrate ChatGPT's capability to serve as a resource for medical students by providing continuous access to information, particularly benefiting students in remote or under-resourced areas [6].

ChatGPT is known for having the capability of performing near the passing threshold of 60% accuracy of the United States Medical Licensing Examination (USMLE) [20] and for approximately having the knowledge equivalent to a third-year medical student [21]. ChatGPT's performance on medical examinations has been analyzed across different countries and questions. A 2023 systematic review with a meta-analysis of 19 articles found a mean performance of ChatGPT of around 61% [22], and a more recent review published in 2024 concluded that, despite ChatGPT's satisfactory performance in examinations, further studies are necessary to fully explore its potential in medical education [23].

Furthermore, ChatGPT struggles with non-English language assessments possibly due to a limited understanding of linguistic nuances and Western-centric internet data, which may not fully represent the clinical and disease differences in some countries, like African and Asian populations [24], warranting more studies in other languages to ensure better understanding of ChatGPT's accuracy in diverse cultural contexts. For example, ChatGPT performed considerably lower on a medical examination in Chinese (45.8% correct answers on the Chinese National Medical Licensing Examination) [25], and even worse in the French examination with 22% correct answers [26].

In July 2024, OpenAI launched GPT-4o mini, a smaller version of its latest GPT-4o ("o" for "omni") AI language model. This new model replaced GPT-3.5 Turbo in ChatGPT, making this an ideal time to study the performance of both free models in resolving written medical examinations.

This study aims to evaluate the performance and consistency of 2 AI models, ChatGPT-3.5 Turbo and ChatGPT-4o mini, in solving the questions of a non-English language (European Portuguese) written medical examination, with a format of multiple-choice with one best answer—the 2023 National Examination for Access to Specialized Training (Prova Nacional de Acesso à Formação Especializada [PNA])—and compare their performance to that of human candidates.

Methods

Study Design

The PNA examination is part of the requirements for entering specialized medical training in Portugal. Its purpose is to rank candidates for accessing specialized training vacancies, so no minimum passing grade is needed [27].

The PNA questions used in this study were from the actual 2023 Portuguese PNA examination, which is publicly available on the web [27]. This examination includes 150 questions with 5 multiple-choice answers each, with only a single best answer, similar to the USMLE. The questions are based on clinical vignettes and divided into 2 parts with 75 questions each. The examination emphasizes clinical reasoning and the application and integration of clinical knowledge and is scored on a scale from 0 to 150 points, with no penalties for blank or incorrect answers. It covers various medical disciplines, including medicine, surgery, pediatrics, gynecology and obstetrics, and psychiatry. The examination duration is 240 minutes, divided into 2 parts of 120 minutes each [27].

ChatGPT-3.5 Turbo was provided with the first part of the examination (74 no image-based multiple-choice questions [MCQs]) on July 18, 2024, and ChatGPT-4o mini with the second part of the examination (74 no image-based MCQs) on July 19, 2024. The questions were entered into the models in European Portuguese and in a format similar to how they are presented to human candidates, and each model was requested to provide a single-letter answer, just like human candidates. For each question, the models generated an answer using their natural language processing capabilities. Following each model's response, a follow-up question, "Are you sure?" was asked to test for consistency—this technique was previously used by Brin et al [28]. An example of the input format of the questions and the respective responses by ChatGPT in European Portuguese is depicted in Table 1, with corresponding translations to English performed by ChatGPT-4o mini. Each question was addressed in a new chat session to reduce the potential influence of memory retention bias of ChatGPT.

XSL•FO
RenderX

Table . Example of the input format of the questions and the respective responses by ChatGPT.

	Examination question in European Portuguese	Examination question translated to English (performed by ChatGPT-4o mini)
Question	Um homem de 73 anos vem à consulta hospitalar para reavaliação de doença pulmonar obstrutiva crónica. Na consulta prévia, há seis meses, apresentava-se em estadio GOLD B. Refere agora, desde há três meses, agravamento da dispneia para esforços médios, sem alteração do padrão habitual de tosse ou de expetoração. Nega febre, perda de apetite ou outras queixas de novo. A história médica revela ainda obesidade. A medicação habitual inclui brometo de tiotrópio e salmeterol. É ex-fumador de 40 UMA desde há 10 anos. Os sinais vitais são temperatura 36 °C, frequência respiratória 18/min, frequência cardíaca 78/min e pressão arterial 115/89 mm Hg; SpO2 94% (ar ambiente). Ao exame físico apresenta cianose labial, com aparência confortável e atrofia muscular na área temporal. A auscultação pulmonar revela crepitações raras dispersas, sem prolongamento do tempo expiratório. O restante exame físico, incluindo a auscultação cardíaca, encontra-se dentro dos parâmetros de normalidade.	A 73-year-old man comes to the hospital for reevaluation of chronic obstructive pulmonary disease (COPD). At the previous visit six months ago, he was in GOLD stage B. He now reports worsening shortness of breath with moderate exertion over the past three months, without changes in his usual cough or sputum production. He denies fever, loss of appetite, or any new complaints. His medical history also reveals obesity. His usual medications include tiotropium bromide and salmeterol. He is a former smoker with a 40 pack-year history, quit 10 years ago. Vital signs are temperature 36 °C, respiratory rate 18/min, heart rate 78/min, and blood pressure 115/89 mm Hg; SpO2 94% (room air). On physical examination, he shows lip cyanosis, appears comfortable, and has muscle atrophy in the temporal area. Lung auscultation reveals scattered rare crackles, with no prolonged expiratory phase. The rest of the physical examination, including cardiac auscultation, is within normal limits.
	Qual das seguintes alternativas mais provavelmente reduz o risco de morte relacionada com esta doença? A. Azitromicina. B. Montelucaste. C. Oxigenoterapia. D. Prednisolona. E. Teofilina. Indique apenas a letra da resposta, sem justificação.	Which of the following is most likely to reduce the risk of death related to this condition? A. Azithromycin. B. Montelukast. C. Oxygen therapy. D. Prednisolone. E. Theophylline. Please indicate only the letter of the answer, without justification.
ChatGPT-3.5 Turbo answer	A resposta correta é: (C) Oxigenoterapia.	The correct answer is: (C) Oxygen therapy.
Follow-up question	Tem a certeza?	Are you sure?
ChatGPT-3.5 Turbo answer	Peço desculpa pelo erro anterior. Vamos rever a situação:	I apologize for the previous error. Let's review the situation:
	Dadas as características do paciente com DPOC grave e hipoxemia (SpO2 de 94% em ar ambiente), a intervenção que mais provavelmente reduzirá o risco de morte relacionada com esta doença é a oxigenoterapia. Portanto, a resposta correta é: (C) Oxigenoterapia.	Given the characteristics of the patient with severe COPD and hypoxemia (SpO2 of 94% on room air), the intervention that is most likely to reduce the risk of death related to this condition is oxygen therapy. Therefore, the correct answer is: (C) Oxygen therapy.

Human candidates' scores were obtained from the official records of the PNA 2023 examination, which are also publicly available on the web [27].

To compare the mean scores of human candidates to those of the language models, all questions of the PNA 2023 examination had to be answered. Since the examination included 2 questions using images (one in the first part and another one in the second part; both with electrocardiogram strips), these questions were answered by GPT-4o, as it can handle images in addition to text.

Ethical Considerations

This study exclusively used data that had been previously published online and did not involve direct interaction with human participants. As a result, ethical guidelines pertaining to human participants are not applicable.

Statistical Analysis

Analyses were performed using IBM SPSS Statistics (Version 21). The McNemar test [29] with continuity correction [30] was used to determine differences between the first and second responses of ChatGPT-3.5 Turbo and ChatGPT-4o mini. Single-parameter *t* test was used to compare the performance of ChatGPT-3.5 Turbo and ChatGPT-4o mini with that of human candidates. Frequencies and percentages were used for

XSL•FO
RenderX

categorical variables and means and CIs for numerical variables. Statistical significance was considered at $P<.05$.

Results

Overall Performance and Consistency

In the initial response with ChatGPT-3.5 Turbo, of the 74 questions, 40 (54%) answers were correct and 34 (46%) answers were incorrect. After the follow-up question, "Are you sure?," the number of correct answers decreased to 28 (38%), while the number of incorrect answers increased to 46 (62%). This change occurred because ChatGPT-3.5 Turbo corrected 12 originally incorrect answers, but also changed 24 originally correct answers to incorrect. This pattern of change approached, but did not reach, significance ($\chi^2_1=3.361$, $P=.067$).

Initially, of the 74 questions, ChatGPT-4o mini produced 48 (65%) correct answers and 26 (35%) incorrect answers. After being asked, "Are you sure?," the correct answers dropped to 42 (57%), while incorrect answers rose to 32 (43%). This change occurred because ChatGPT-4o mini fixed 12 previously wrong answers but also changed 18 previously correct answers to incorrect. This pattern of change was not statistically significant ($\chi^2_1=0.833$, $P=.361$).

The 2 questions using images (one in the first part and another one in the second part) were answered correctly by GPT-4o.

LLM Chatbot Versus Human

When evaluating AI capabilities in relation to human abilities, LLM responses in part 1 of PNA (74 questions resolved by ChatGPT-3.5 Turbo plus 1 by GPT-4o) showed lower accuracy than human respondents. The human mean score was statistically significantly higher by 6.04 (95% CI 5.65-6.43) than the LLM score of 41 ($P<.001$).

In part 2 of PNA (74 questions resolved by ChatGPT-4o mini added to 1 question by GPT-4o), the LLM score showed higher accuracy than human respondents. The human mean score was statistically significantly lower by 5.58 (95% CI 5.25-5.9) than the LLM score of 49 ($P<.001$).

Discussion

Principal Findings

This study analyzes the performance of 2 ChatGPT models (ChatGPT-3.5 Turbo and ChatGPT-4o mini) on the Portuguese medical written examination: 2023 National Examination for Access to Specialized Training, revealing important differences in accuracy and consistency. Although, both ChatGPT-3.5 Turbo and ChatGPT-4o mini answered correctly in the majority of the questions, ChatGPT-4o mini achieved a higher accuracy rate of 65% (48/74) compared to ChatGPT-3.5 Turbo's 54% (40/74), demonstrating a superior capability in handling medical questions. Additionally, ChatGPT-4o mini showed greater consistency in confirming answers, highlighting its reliability. When evaluated against human respondents, ChatGPT-4o mini outperformed the average human accuracy, while ChatGPT-3.5 Turbo fell short.

Strengths

This study stands out for its innovative approach in analyzing the performance of ChatGPT-3.5 Turbo and ChatGPT-4o mini in a medical examination context. It is the first to evaluate these models using an examination conducted in a less commonly studied language, Portuguese, thereby broadening the scope of language-specific AI assessments. By incorporating the actual scores of human candidates for comparison, the study provides a robust benchmark against real-world performance. Furthermore, the research examines the stability of the AI's answers by repeatedly asking "Are you sure?," offering valuable insights into the consistency of the responses.

Comparison to Prior Work

A recent study evaluated ChatGPT's performance on medical licensing examinations across multiple countries (United States, Italy, France, Spain, United Kingdom, and India) and determined a variable accuracy, ranging from 22% on the French examination to 73% on the Italian examination [26]. In this study, ChatGPT answered correctly in more than 50% of the Portuguese medical examination questions, positioning it next to the countries with better performance. For example, in a Turkish study, ChatGPT reached 70.9% accuracy in the medical specialty examination [31]. In the Iranian medical licensing examination, ChatGPT performed with 68.5% of the questions answered correctly [32]. And in Poland, ChatGPT achieved a 67.1% correct response rate on the Polish medical specialization licensing examination [33].

When analyzing the differences between the 2 ChatGPT versions, ChatGPT-4o mini outperformed ChatGPT-3.5 Turbo in this study: 65% (48/74) vs 54% (40/74) correct response rate. This suggests that advancements in the underlying architecture and training data of ChatGPT-4o mini (knowledge up to October 2023) have improved its capability to understand and respond to medical questions with more accuracy. Previous studies evaluating the performance of different ChatGPT models found that ChatGPT-4 consistently performed better compared to ChatGPT-3.5. For example, ChatGPT-4 outperformed ChatGPT-3.5 on the Polish Medical Final Examination [34], the Spanish Medical Residency Entrance Examination (Médico Interno Residente) [35], the 2023 Japanese Nursing Examination [36], the Peruvian National Licensing Medical Examination (Examen Nacional de Medicina) [37], and in the USMLE soft skill assessments [28], to name a few. Nonetheless, ChatGPT-4 is a paid model and thus not accessible to everyone, which is not the case for the most recent free-to-use ChatGPT-4o mini.

Another important aspect is consistency. The results of this study revealed that ChatGPT-3.5 Turbo was less stable when asked to confirm its original answers. These results are consistent with those of Brin et al [28], who found that ChatGPT-3.5 altered its answers 82.5% of the time in the USMLE assessments [28]. Unfortunately, in this study, it was not shown that by changing the original answers, ChatGPT-3.5 Turbo improves its accuracy. This contrasts with studies on human students, which have shown that changing their answers usually improves their test scores [38]. One can wonder, since the "awareness of what one knows and does not know depends in part on how much one knows" [39], does ChatGPT-3.5 Turbo

XSL•FO
RenderX

change its answers because it does not know, or does it simply change answers to satisfy the user when prompted?

When evaluating the AI models against human respondents, it was found that in part 2 of the PNA examination (74 questions resolved by ChatGPT-4o mini plus 1 question by GPT-4o), the LLM outperformed the average accuracy of human participants. In contrast, in part 1 of the PNA examination (74 questions resolved by ChatGPT-3.5 Turbo plus 1 question by GPT-4o), LLM showed lower accuracy than human respondents. This indicates that while earlier versions, like ChatGPT-3.5 Turbo, may have required a high degree of human oversight, more recent and advanced versions, like ChatGPT-4o mini, have the potential to match or exceed human performance in medical domains. Although no previous studies have analyzed the performance of ChatGPT-4o mini, and no direct comparisons can be made, some studies have already noted that LLMs outperformed human candidates in several medical examinationinations (eg, the German Medical State Examinations of 2022 [40], part 1 of the Fellowship of the Royal College of Ophthalmologists MCQ examination [41], and the University of Toronto Family Medicine Residency Progress Test [42]).

Limitations

This study has several limitations regarding the performance evaluation of ChatGPT-3.5 Turbo and ChatGPT-4o mini. The analysis was based solely on ChatGPT's indication of the correct answer, which, while aligning with expectations for human candidates, does not consider other aspects of examination performance. Additionally, the grading did not account for the complexity or length of the questions, providing an incomplete assessment of the models' performance. Further studies should incorporate a more comprehensive evaluation framework that considers the reasoning process and evaluates performance across a broader range of question types and difficulties.

Future Perspectives

This study highlights the importance of continuous improvement in ChatGPT models to further enhance their reliability and accuracy. The superior performance of ChatGPT-4o mini compared to its predecessor offers promising applications in medical education. Its higher accuracy and consistency suggest that it could serve as an effective tool for training medical students. However, a broader assessment of ChatGPT-4o mini across various tests and real-world scenarios is required, as good performance on a specific test may not indicate abilities for general and reliable medical education usage. Additionally, there are known drawbacks and ethical considerations when using AI applications, including the potential for fabricated, incorrect, or biased information [43]. Other issues include limited training periods and the possibility of providing different answers to the same question depending on how the question is phrased [43]. A recent systematic scoping review by Xu et al [44] advises medical students to use ChatGPT cautiously, cross-checking information with reliable sources and disclosing AI-generated content in their work. Teachers should guide students on the effective and ethical use of ChatGPT, assess its reliability, and explore mixed assessment methods to evaluate student abilities while considering its impact on traditional assignments [44].

Conclusion

On the 2023 Portuguese National Examination for Access to Specialized Training, ChatGPT-4o mini achieved an accuracy rate of 65% (48/74), surpassing ChatGPT-3.5 Turbo. This demonstrates a superior capability in handling medical questions. ChatGPT-4o mini outperformed medical candidates, while ChatGPT-3.5 Turbo had a more moderate performance. This study highlights the advancements and potential of ChatGPT models in medical education, emphasizing the importance of careful implementation with teacher oversight and further research.

Acknowledgments

This study was supported by national funds through FCT (Fundação para a Ciência e a Tecnologia, I.P.) within CINTESIS R&D Unit (reference UIDB/4255/2020) and within the scope of the project RISE, Associated Laboratory (reference LA/P/0053/2020). During the preparation of this manuscript, the author used ChatGPT-4o mini in order to improve the language of the manuscript and correct grammatical errors. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Conflicts of Interest

None declared.

References

1. Feuerriegel S, Hartmann J, Janiesch C, Zschech P. Generative AI. Bus Inf Syst Eng 2024 Feb;66(1):111-126. [doi: 10.1007/s12599-023-00834-7]
2. Ramdurai B, Adhithya P. The impact, advancements and applications of generative AI. Int J Comput Sci Eng 2023;10(6):1-8. [doi: 10.14445/23488387/IJCSE-V10I6P101]
3. Cao Y, Li S, Liu Y, et al. A comprehensive survey of AI-generated content (AIGC): a history of generative AI from GAN to ChatGPT. arXiv. Preprint posted online on Mar 7, 2023. [doi: 10.48550/arXiv.2303.04226]
4. Introducing ChatGPT. OpenAI. 2022 Nov 30. URL: https://openai.com/index/chatgpt [accessed 2024-07-30]

XSL•FO
RenderX

5. Conte N. Ranked: the most popular AI tools. Visual Capitalist. 2024 Jan 24. URL: https://www.visualcapitalist.com/ranked-the-most-popular-ai-tools [accessed 2025-02-19]
6. Wang X, Sanders HM, Liu Y, et al. ChatGPT: promise and challenges for deployment in low- and middle-income countries. Lancet Reg Health West Pac 2023 Dec;41:100905. [doi: 10.1016/j.lanwpc.2023.100905] [Medline: 37731897]
7. Aydin Ö, Karaarslan E. Is ChatGPT leading generative AI? What is beyond expectations? Acad Platform J Eng Smart Sys 2023;11(3):118-134. [doi: 10.21541/apjess.1293702]
8. BuHamdan S, Alwisy A, Bouferguene A. Generative systems in the architecture, engineering and construction industry: a systematic review and analysis. Int J Archit Comput 2021 Sep;19(3):226-249. [doi: 10.1177/1478077120934126]
9. Ali O, Abdelbaki W, Shrestha A, Elbasi E, Alryalat MAA, Dwivedi YK. A systematic literature review of artificial intelligence in the healthcare sector: benefits, challenges, methodologies, and functionalities. J Innov Knowl 2023 Jan;8(1):100333. [doi: 10.1016/j.jik.2023.100333]
10. Younis HA, Eisa TAE, Nasser M, et al. A systematic review and meta-analysis of artificial intelligence tools in medicine and healthcare: applications, considerations, limitations, motivation and challenges. Diagnostics (Basel) 2024 Jan 4;14(1):109. [doi: 10.3390/diagnostics14010109] [Medline: 38201418]
11. Ruksakulpiwat S, Thorngthip S, Niyomyart A, et al. A systematic review of the application of artificial intelligence in nursing care: where are we, and what's next? J Multidiscip Healthc 2024;17:1603-1616. [doi: 10.2147/JMDH.S459946] [Medline: 38628616]
12. Bitkina OV, Park J, Kim HK. Application of artificial intelligence in medical technologies: a systematic review of main trends. Digit Health 2023;9. [doi: 10.1177/20552076231189331] [Medline: 37485326]
13. Sedaghat S. Early applications of ChatGPT in medical practice, education and research. Clin Med (Lond) 2023 May;23(3):278-279. [doi: 10.7861/clinmed.2023-0078] [Medline: 37085182]
14. Gandomani HS. ChatGPT in medical education: how we can use in medical education: challenges and opportunities. J Multidiscip Care 2023;12(1):1-2. [doi: 10.34172/jmdc.1232]
15. Sani I. Enhancing medical education with ChatGPT: a promising tool for the future. Can J Med 2024 Apr 1;6(1):1-4. [doi: 10.33844/cjm.2024.6032]
16. Hosseini M, Gao CA, Liebovitz DM, et al. An exploratory survey about using ChatGPT in education, healthcare, and research. PLoS One 2023;18(10):e0292216. [doi: 10.1371/journal.pone.0292216] [Medline: 37796786]
17. Guo AA, Li J. Harnessing the power of ChatGPT in medical education. Med Teach 2023 Sep;45(9):1063. [doi: 10.1080/0142159X.2023.2198094] [Medline: 37036161]
18. Jeblick K, Schachtner B, Dexl J, et al. ChatGPT makes medicine easy to swallow: an exploratory case study on simplified radiology reports. Eur Radiol 2024 May;34(5):2817-2825. [doi: 10.1007/s00330-023-10213-1] [Medline: 37794249]
19. Cheung BHH, Lau GKK, Wong GTC, et al. ChatGPT versus human in generating medical graduate exam multiple choice questions—a multinational prospective study (Hong Kong S.A.R., Singapore, Ireland, and the United Kingdom). PLoS One 2023;18(8):e0290691. [doi: 10.1371/journal.pone.0290691] [Medline: 37643186]
20. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023 Feb;2(2):e0000198. [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
21. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 8;9:e45312. [doi: 10.2196/45312] [Medline: 36753318]
22. Levin G, Horesh N, Brezinov Y, Meyer R. Performance of ChatGPT in medical examinations: a systematic review and a meta-analysis. BJOG 2024 Feb;131(3):378-380. [doi: 10.1111/1471-0528.17641] [Medline: 37604703]
23. Sumbal A, Sumbal R, Amir A. Can ChatGPT-3.5 pass a medical exam? A systematic review of ChatGPT's performance in academic testing. J Med Educ Curric Dev 2024;11. [doi: 10.1177/23821205241238641] [Medline: 38487300]
24. Cherif H, Moussa C, Missaoui AM, Salouage I, Mokaddem S, Dhahri B. Appraisal of ChatGPT's aptitude for medical education: comparative analysis with third-year medical students in a pulmonology examination. JMIR Med Educ 2024 Jul 23;10:e52818. [doi: 10.2196/52818] [Medline: 39042876]
25. Wang X, Gong Z, Wang G, et al. ChatGPT performs on the Chinese National Medical Licensing Examination. J Med Syst 2023 Aug 15;47(1):86. [doi: 10.1007/s10916-023-01961-0] [Medline: 37581690]
26. Alfertshofer M, Hoch CC, Funk PF, et al. Sailing the seven seas: a multinational comparison of ChatGPT's performance on medical licensing examinations. Ann Biomed Eng 2024 Jun;52(6):1542-1545. [doi: 10.1007/s10439-023-03338-3] [Medline: 37553555]
27. Prova Nacional de Acesso à Formação Especializada 2023 Perguntas Frequentes. ACSS. 2023. URL: https://www.acss.min-saude.pt/wp-content/uploads/2018/09/FAQ_PNA2023.pdf [accessed 2024-08-01]
28. Brin D, Sorin V, Vaid A, et al. Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. Sci Rep 2023 Oct 1;13(1):16492. [doi: 10.1038/s41598-023-43436-9] [Medline: 37779171]
29. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 1947 Jun;12(2):153-157. [doi: 10.1007/BF02295996] [Medline: 20254758]

XSL•FO
RenderX

30. Edwards AL. Note on the correction for continuity in testing the significance of the difference between correlated proportions. Psychometrika 1948 Sep;13(3):185-187. [doi: 10.1007/BF02289261] [Medline: 18885738]
31. Oztermeli AD, Oztermeli A. ChatGPT performance in the medical specialty exam: an observational study. Medicine (Baltimore) 2023 Aug 11;102(32):e34673. [doi: 10.1097/MD.0000000000034673] [Medline: 37565917]
32. Ebrahimian M, Behnam B, Ghayebi N, Sobhrakhshankhah E. ChatGPT in Iranian medical licensing examination: evaluating the diagnostic accuracy and decision-making capabilities of an AI-based model. BMJ Health Care Inform 2023 Dec 11;30(1):e100815. [doi: 10.1136/bmjhci-2023-100815] [Medline: 38081765]
33. Wójcik S, Rulkiewicz A, Pruszczyk P, Lisik W, Poboży M, Domienik-Karłowicz J. Reshaping medical education: performance of ChatGPT on a PES medical examination. Cardiol J 2024;31(3):442-450. [doi: 10.5603/cj.97517] [Medline: 37830257]
34. Rosoł M, Gąsior JS, Łaba J, Korzeniewski K, Młyńczak M. Evaluation of the performance of GPT-3.5 and GPT-4 on the Polish Medical Final Examination. Sci Rep 2023 Nov 22;13(1):20512. [doi: 10.1038/s41598-023-46995-z] [Medline: 37993519]
35. Guillen-Grima F, Guillen-Aguinaga S, Guillen-Aguinaga L, et al. Evaluating the efficacy of ChatGPT in navigating the Spanish Medical Residency Entrance Examination (MIR): promising horizons for AI in clinical medicine. Clin Pract 2023 Nov 20;13(6):1460-1487. [doi: 10.3390/clinpract13060130] [Medline: 37987431]
36. Kaneda Y, Takahashi R, Kaneda U, et al. Assessing the performance of GPT-3.5 and GPT-4 on the 2023 Japanese nursing examination. Cureus 2023 Aug;15(8):e42924. [doi: 10.7759/cureus.42924] [Medline: 37667724]
37. Flores-Cohaila JA, García-Vicente A, Vizcarra-Jiménez SF, et al. Performance of ChatGPT on the Peruvian National Licensing Medical Examination: cross-sectional study. JMIR Med Educ 2023 Sep 28;9:e48039. [doi: 10.2196/48039] [Medline: 37768724]
38. Kruger J, Wirtz D, Miller DT. Counterfactual thinking and the first instinct fallacy. J Pers Soc Psychol 2005 May;88(5):725-735. [doi: 10.1037/0022-3514.88.5.725] [Medline: 15898871]
39. Coutinho MVC, Thomas J, Fredricks-Lowman I, Alkaabi S, Couchman JJ. Unskilled and unaware: second-order judgments increase with miscalibration for low performers. Front Psychol 2024;15:1252520. [doi: 10.3389/fpsyg.2024.1252520] [Medline: 38952836]
40. Roos J, Kasapovic A, Jansen T, Kaczmarczyk R. Artificial intelligence in medical education: comparative analysis of ChatGPT, Bing, and medical students in Germany. JMIR Med Educ 2023 Sep 4;9:e46482. [doi: 10.2196/46482] [Medline: 37665620]
41. Fowler T, Pullen S, Birkett L. Performance of ChatGPT and Bard on the official part 1 FRCOphth practice questions. Br J Ophthalmol 2024 Oct;108(10):1379-1383. [doi: 10.1136/bjo-2023-324091]
42. Huang RS, Lu KJQ, Meaney C, Kemppainen J, Punnett A, Leung FH. Assessment of resident and AI chatbot performance on the University of Toronto Family Medicine Residency Progress Test: comparative study. JMIR Med Educ 2023 Sep 19;9:e50514. [doi: 10.2196/50514] [Medline: 37725411]
43. Wong RSY, Ming LC, Raja Ali RA. The intersection of ChatGPT, clinical medicine, and medical education. JMIR Med Educ 2023 Nov 21;9:e47274. [doi: 10.2196/47274] [Medline: 37988149]
44. Xu X, Chen Y, Miao J. Opportunities, challenges, and future directions of large language models, including ChatGPT in medical education: a systematic scoping review. J Educ Eval Health Prof 2024;21(6):6. [doi: 10.3352/jeehp.2024.21.6] [Medline: 38486402]

Abbreviations

AI: artificial intelligence
LLM: large language model
MCQ: multiple-choice question
PNA: Prova Nacional de Acesso à Formação Especializada
USMLE: United States Medical Licensing Examination

Edited by B Lesselroth; submitted 05.08.24; peer-reviewed by B Thies, LA Digiampietri, R Pellegrino; revised version received 30.11.24; accepted 12.12.24; published 05.03.25.

Please cite as:
Prazeres F
ChatGPT's Performance on Portuguese Medical Examination Questions: Comparative Analysis of ChatGPT-3.5 Turbo and ChatGPT-4o Mini
JMIR Med Educ 2025;11:e65108
URL: https://mededu.jmir.org/2025/1/e65108
doi:10.2196/65108

XSL•FO
RenderX

© Filipe Prazeres. Originally published in JMIR Medical Education (https://mededu.jmir.org), 5.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Performance of ChatGPT-4 on Taiwanese Traditional Chinese Medicine Licensing Examinations: Cross-Sectional Study

Liang-Wei Tseng[1], MD; Yi-Chin Lu[2], MD; Liang-Chi Tseng[3], MSc; Yu-Chun Chen[4,5,6], MD, MSc; Hsing-Yu Chen[2,7], MD, PhD

[1]Division of Chinese Acupuncture and Traumatology, Center of Traditional Chinese Medicine, Chang Gung Memorial Hospital, Taoyuan, Taiwan

[2]Division of Chinese Internal Medicine, Center for Traditional Chinese Medicine, Chang Gung Memorial Hospital, No. 123, Dinghu Rd, Gueishan Dist, Taoyuan, Taiwan

[3]Google International LLC Taiwan Branch, Taipei, Taiwan

[4]School of Medicine, Faculty of Medicine, National Yang-Ming Chiao Tung University, Taipei, Taiwan

[5]Taipei Veterans General Hospital, Yuli Branch, Taipei, Taiwan

[6]Institute of Hospital and Health Care Administration, National Yang-Ming Chiao Tung University, Taipei, Taiwan

[7]School of Traditional Chinese Medicine, College of Medicine, Chang Gung University, Taoyuan, Taiwan

Corresponding Author:
Hsing-Yu Chen, MD, PhD
Division of Chinese Internal Medicine, Center for Traditional Chinese Medicine, Chang Gung Memorial Hospital, No. 123, Dinghu Rd, Gueishan Dist, Taoyuan, Taiwan

Abstract

Background: The integration of artificial intelligence (AI), notably ChatGPT, into medical education, has shown promising results in various medical fields. Nevertheless, its efficacy in traditional Chinese medicine (TCM) examinations remains understudied.

Objective: This study aims to (1) assess the performance of ChatGPT on the TCM licensing examination in Taiwan and (2) evaluate the model's explainability in answering TCM-related questions to determine its suitability as a TCM learning tool.

Methods: We used the GPT-4 model to respond to 480 questions from the 2022 TCM licensing examination. This study compared the performance of the model against that of licensed TCM doctors using 2 approaches, namely direct answer selection and provision of explanations before answer selection. The accuracy and consistency of AI-generated responses were analyzed. Moreover, a breakdown of question characteristics was performed based on the cognitive level, depth of knowledge, types of questions, vignette style, and polarity of questions.

Results: ChatGPT achieved an overall accuracy of 43.9%, which was lower than that of 2 human participants (70% and 78.4%). The analysis did not reveal a significant correlation between the accuracy of the model and the characteristics of the questions. An in-depth examination indicated that errors predominantly resulted from a misunderstanding of TCM concepts (55.3%), emphasizing the limitations of the model with regard to its TCM knowledge base and reasoning capability.

Conclusions: Although ChatGPT shows promise as an educational tool, its current performance on TCM licensing examinations is lacking. This highlights the need for enhancing AI models with specialized TCM training and suggests a cautious approach to utilizing AI for TCM education. Future research should focus on model improvement and the development of tailored educational applications to support TCM learning.

(JMIR Med Educ 2025;11:e58897) doi:10.2196/58897

KEYWORDS

artificial intelligence; AI language understanding tools; ChatGPT; natural language processing; machine learning; Chinese medicine license exam; Chinese medical licensing examination; medical education; traditional Chinese medicine; large language model

Introduction

Traditional Chinese medicine (TCM), recognizeartid as one of the most renowned traditional medical systems, boasts a history spanning thousands of years. In the modern era, TCM has evolved to form an integral part of the formal health care system in East Asian countries, particularly in China and Taiwan [1,2]. TCM encompasses a wealth of theoretical knowledge and features unique diagnostic and treatment methods, such as acupuncture and herbal therapy. As a highly practical discipline, TCM learning traditionally relies on the accumulation of experience and the mentorship inherent in the master-apprentice

XSL•FO
RenderX

system; hence, this education model may not be sufficiently reliable or comprehensive. However, with the emerging need for integrative medicine over time, TCM has been integrated into the modern medical education system. This integration has led to prominent changes in educational approaches. The incorporation of TCM into academic institutions resulted in the establishment of formal examination systems. For instance, in Taiwan, TCM practitioners must pass a biannual licensing examination, termed the National Senior Professional and Technical Examinations for Chinese Medicine Practitioners (hereinafter called the "TCM licensing examinations"), to practice as a licensed TCM doctor, similar to their Western medicine counterparts [3].

The advancements in technology and the development of artificial intelligence (AI) have begun to impact and challenge the medical field, with TCM being no exception [4,5]. In the past year, significant progress has been made in AI language models, particularly those based on the generative pretrained transformer (GPT) architecture. ChatGPT, a conversational variant of the GPT model, has demonstrated its potential across various domains [6]. Recognized for its foundational medical knowledge and conversational capabilities, ChatGPT is considered a valuable tool in medical education, aiding in the understanding and application of medical knowledge [7], thereby facilitating student learning [8]. However, its responses are not consistently reliable. Unlike humans who answer questions based on an understanding of the content, it generates replies by drawing from a vast database. Therefore, although it can produce human-like conversations and respond to inquiries, it cannot guarantee the accuracy of its responses [9,10].

Discussions have emerged regarding the sufficiency of AI for clinical decision-making and basic medical consultation [7,11]. In addition, to be a potential mentor for medical students, one benchmark is the ability of AI to pass national licensing examinations (the minimum standard for practicing physicians). Thus, the application of ChatGPT in medical examinations has opened a new research direction. Studies have shown that GPT models, especially GPT-4, can achieve commendable scores on a variety of standardized tests for multiple professions, such as physicians [12-14], pharmacists [15], and nurses [16]. This success in examination settings has sparked interest in the potential of ChatGPT as a self-learning tool, suggesting its use for examination preparation and knowledge enhancement [17].

As previously mentioned, while TCM is a traditional medical system distinct from modern medicine, it has been integrated into modern medical education systems and subjected to formal examinations. The question arises: does ChatGPT possess the requisite knowledge level to assist TCM students in their learning? Only 1 study examined GPT's ability to answer TCM questions, but it focused on questions sourced from online TCM texts rather than formally recognized examination questions and utilized older GPT models (GPT-3 Turbo) [18]. In contrast, a more rigorous study on traditional Korean medicine found that, due to the unique nature of traditional medicine, GPT models require specially optimized prompts, such as language-related adjustments, to pass examinations [19]. However, considering the classical Chinese language barrier and different medical theories in TCM, whether GPT models would face challenges in TCM licensing examinations remains unexplored.

The aim of this study is to evaluate whether ChatGPT can accurately understand and respond to TCM questions by assessing its performance in simulated examination environments. By analyzing the accuracy of AI-generated answers, we sought to identify factors affecting their correctness. This study also aims to understand the consistency between AI-generated answers and their accompanying explanations, offering insights into the depth of understanding of this model. By analyzing the performance of ChatGPT in simulated TCM licensing examinations and comparing it with human performance, this study hopes to provide new insights and recommendations for innovation and development in TCM education.

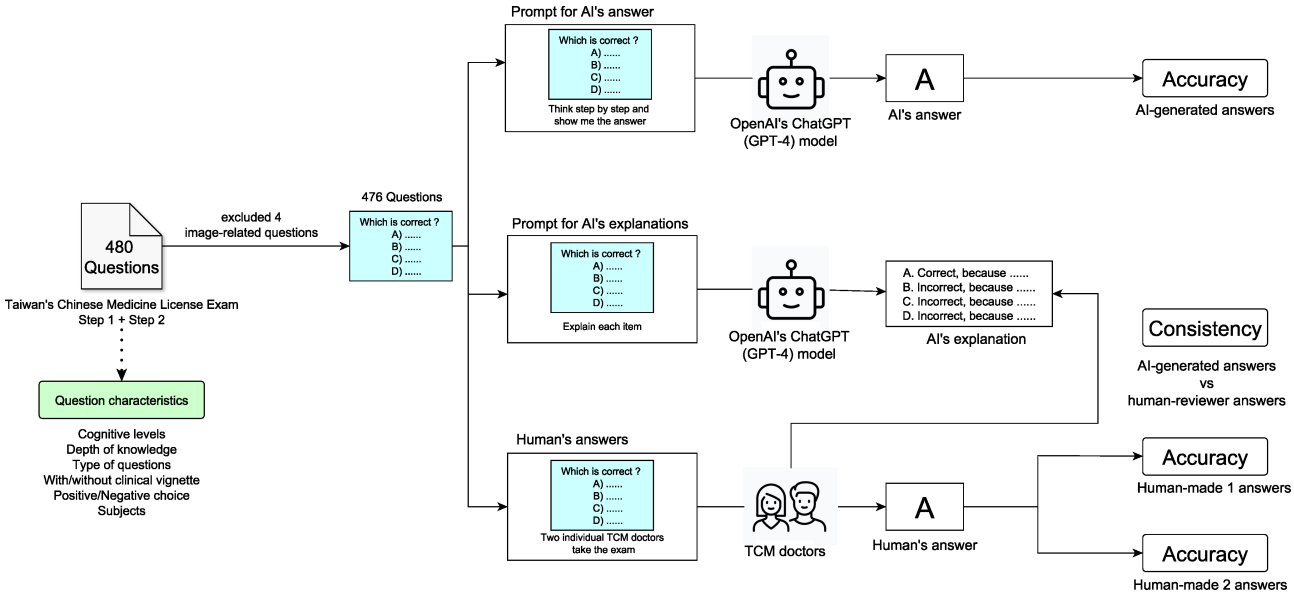
Methods

Study Design

Figure 1 shows the data processing flowchart of this study. The feasibility of using ChatGPT (GPT-4 model, with a knowledge cutoff date of September 2021), developed by OpenAI, with 2 different prompts on responding to the first National Senior Professional and Technical Examinations for Chinese Medicine Practitioners was assessed by comparing the responses of the model to those of licensed TCM resident doctors. A total of 480 questions from the 2022 examination were inputted into ChatGPT, and 2 different approaches were used to obtain responses from ChatGPT. The first step involved prompting AI to select the correct answer directly from the question options. The second step required ChatGPT to explain why each option was correct or incorrect before selecting the correct answer. For the second step, individual answers and explanations from ChatGPT were manually assessed for accuracy and consistency. Subsequently, accuracy was measured by comparing the AI-selected answers with the correct answers. Additionally, the performance of AI was benchmarked against that of human experts. Two individual TCM resident doctors took the same examination without preparation, and their answers were also evaluated for accuracy. Finally, consistency was evaluated by comparing explanations against a standard set of answers for logical coherence, and the reasons for inconsistency were also verified by the 2 TCM doctors.

XSL•FO
RenderX

Figure 1. Flowchart of this study. GPT: generative pretrained transformer; TCM: traditional Chinese medicine.



The TCM Licensing Examination in Taiwan

In Taiwan, TCM doctors are qualified through 2 stages of licensing examinations after graduation from their TCM course at the university. The contents and answers are freely downloadable after each examination from the following website [20]. The examinations contain 2 stages corresponding to 10 subjects. The first stage consists of basic theory, including 黃帝內經 (Huangdi Neijing), 難經 (Nanjing) (domain I), and basic pharmacology and formulation (domain II). The second stage consists of principles of diagnosis and treatment, including 傷寒論 (Shanghanlun) and 金匱要略 (Jinguiyaolue) (domain III), TCM internal medicine (domain IV), TCM gynecology and obstetrics (domain IV), TCM pediatrics (domain IV), TCM dermatology (domain V), TCM otorhinolaryngology (domain V, including questions regarding the specialty concerning ears, nose, and throat [ENT]) and ophthalmology (domain V), TCM traumatology (domain V), and acupuncture (domain VI). Each domain contains 80 multiple-choice questions with single answers. The full score of each domain is 100. The examination score is calculated by dividing the total score by the number of subjects. Only examinees obtaining average scores ≥60 pass the examination. TCM students are eligible to take the first-stage examination when they have earned the requisite fourth-year university credits. Before the second-stage examination, TCM students must first pass the first-stage examination and graduate from the 7-year university course.

Question Characteristics

A total of 5 factors were used to characterize the examination questions, including the cognitive level, depth of knowledge (DOK), type of questions, vignette style, and polarity of questions (Table S1 in Multimedia Appendix 1). LWT and YCL independently reviewed and classified all questions according to the definitions of these 5 factors. In case of disagreement, HYC was consulted, and the disagreement was resolved by reaching a consensus among all authors. Bloom's taxonomy was modified to classify the questions into lower-order thinking skills (LOTS) and higher-order thinking skills (HOTS). LOTS include remembering, understanding, and applying knowledge to questions, while HOTS include further analyzing, evaluating, and creating after learning [21,22]. For the DOK, 3 levels, ranging from low to high based on Webb's framework on science, were defined as recall, concept, and strategic thinking. Questions with higher levels of DOK indicate the recruitment of sophisticated thinking [23]. Furthermore, the licensing examinations in Taiwan are presented as single-choice questions, adhering to the 1 stem, 4 choices policy. However, 2 types of questions were used to add variety to examination questions, including single-answer multiple-choice (SAMC) and single-answer, multiple-response multiple-choice (SAMRMC) questions. SAMC questions had only 1 most appropriate answer, while SAMRMC questions require the tester to choose the most appropriate answer composed of multiple correct options provided in each question (Table S2 in Multimedia Appendix 2). Moreover, if the content of a question presents clinical scenarios, this question would be categorized as the clinical vignette type. This type of question typically aims to examine the ability of the tester to analyze the clinical conditions and corresponding actions. The polarity of a question depended on whether the question was positively or negatively framed. A "positive-choice question" solicits the correct or affirmative answer, whereas a "negative choice question" demands the identification of the incorrect or negative answer.

Prompt for AI-Generated Answers

To enhance the precision and brevity of responses obtained from ChatGPT (GPT-4 model), we strategically added "think step-by-step" to our queries. This approach aimed to guide the model toward a methodical and sequential problem-solving process. Subsequently, by integrating the command "but show me only the answer, do not explain it," we aimed to extract a more refined and consolidated answer, significantly boosting the response accuracy of the model. An example of a prompt with response is demonstrated in Table S3 in Multimedia Appendix 3. We created a collection of unique prompts derived from an equal number of questions in the question database, submitting them sequentially to the AI model. To solve the issue

XSL•FO
RenderX

of memory retention between submissions, we used a specialized application designed to initiate separate application programming interface requests for each prompt. This approach guaranteed that each application programming interface interaction would be initiated separately. This ensures that the processing of each prompt and the generation of its answer were conducted in isolation, thereby preserving the integrity of the responses without interference from a prior response [24,25].

Prompt for Explanations Provided By AI Through Step-By-Step and Human-Curated Answers

Furthermore, to understand the thinking process of GPT and evaluate the accuracy of its interpretation of our inquiries, we prompted ChatGPT to "explain each item" for each question. This prompt directed the AI to furnish exhaustive explanations for each item [26] (Table S4 in Multimedia Appendix 4). LWT and YCL reviewed all explanations to items and reached decisive responses based on AI-generated explanations. This process was termed "human-curated responses." To authentically represent the logic of AI, we refrained from making any human amendments, even if the explanations provided by AI were incorrect. The answer would be marked as "wrong" if the AI-generated explanations were incorrect.

Outcome Assessment

We evaluated the accuracy of answers generated by the GPT, those made by humans, and explanations provided by the GPT and curated by humans. This was achieved by calculating the ratio of accurate responses to the total number of questions and representing the results as a percentage. This measure of accuracy underwent comparative analysis across different attributes of the questions. The human-curated answers, which encapsulated the interpretation of questions by AI, were evaluated by LWT, YCL, and HYC, who reached a consensus to identify instances of misinterpretation of the question (GPT cannot understand the question and does not provide an answer), misunderstanding of concepts (GPT can understand the question, but lacks knowledge of the topic), and incorrect application of principles (the responses GPT provides are correct in general but fail to answer the question).

Statistical Analysis

Proportions and percentages were used to present categorical data. A logistic regression approach was adopted to assess the effect of various attributes of questions on the correctness of responses generated by GPT-4. The cognitive complexity of the questions, their structural format, the inclusion of clinical vignettes, the overall polarity of questions, and the subjects were used as covariates in the logistic regression with univariable and multivariable models. The influence of each variable on the probability of the AI producing accurate answers was quantified using the adjusted odds ratio, accompanied by 95% CIs. Additionally, the κ statistic was used to evaluate the agreement between responses generated by GPT and curated by humans. This represented the different viewpoints concerning the same explanation between GPT and humans. $P<.05$ was used as the threshold for statistical significance. All statistical evaluation was performed utilizing Stata 17 (StataCorp LLC).

Ethical Considerations

This study did not require ethical approval, as it analyzed data obtained from a publicly available database. The test questions and answers used were originally created and copyrighted by the Taiwan Ministry of Examination and made accessible for academic research purposes. The Ministry retains full copyright over the examination content and confirmed that this research adhered to copyright regulations without any infringement.

Results

Question Characteristics

The examination encompassed a total of 480 questions spanning 10 specialties. Four image-related questions were excluded. Our findings indicated that most questions were HOTS, SAMC, negative-choice, and without a clinical vignette. According to Bloom's taxonomy of cognitive learning, the majority of questions across all subjects required HOTS (263/476, 55.3%; LOTS: 213/476, 44.7%). In particular, principles of diagnosis and treatment, TCM internal medicine, TCM dermatology, and TCM traumatology predominantly featured HOTS (58/80, 72.5%; 37/48, 77.1%; 13/19, 68.4%; and 17/20, 85%, respectively), while TCM pediatrics mainly involved LOTS (11/16, 68.8%). Within the LOTS category, "remembering" was the most common type (121/213, 56.8%), while "analyzing" dominated the HOTS category (255/263, 97%). In terms of Webb's DOK analysis of question types, the basic application of skill/concept represented the largest proportion (248/476, 52.1%), surpassing recall (85/476, 17.9%) and strategic thinking (143/476, 30%). A large portion of the questions were formatted as SAMC (439/476, 92.2%). Negative-choice questions comprised 62.2% (296/476) of the total, while 23.9% (180/476) of the questions included a clinical vignette (Table 1, Figures 2 and 3).

XSL•FO
RenderX

Table . Characteristics of TCM[a] licensing examinations in Taiwan, 2022.

Cognitive level	Total (n=476)	Basic theory (n=80)	Basic pharmacology and formulation (n=80)	Principle of diagnosis and treatment (n=80)	TCM internal medicine (n=48)	TCM GYN/OBS[b] (n=16)	TCM pediatrics (n=16)	TCM dermatology (n=19)	TCM ENT, ophthalmology (n=37)	TCM traumatology (n=20)	TCM acupuncture (n=80)
LOTS [c]											
Remembering	121 (25.4)	28 (35)	19 (23.8)	7 (8.8)	1 (2.1)	6 (37.5)	8 (50)	5 (26.3)	13 (35.1)	3 (15)	31 (38.8)
Understanding	41 (8.6)	10 (12.5)	10 (12.5)	7 (8.8)	2 (4.2)	0 (0)	0 (0)	0 (0)	5 (13.5)	0 (0)	7 (8.8)
Applying	51 (10.7)	7 (8.8)	9 (11.3)	8 (10)	8 (16.7)	3 (18.8)	3 (18.8)	1 (5.3)	4 (10.8)	0 (0)	8 (10)
HOTS [d]											
Analyzing	255 (53.6)	34 (42.5)	40 (50)	54 (67.5)	37 (77.1)	7 (43.8)	5 (31.3)	13 (68.4)	15 (40.5)	17 (85)	33 (41.3)
Evaluating	8 (1.7)	1 (1.3)	2 (2.5)	4 (5.0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (1.3)
Depth of knowledge											
Recall	85 (17.9)	20 (25)	18 (22.5)	6 (7.5)	0 (0)	4 (25)	5 (31.3)	3 (15.8)	3 (8.1)	4 (20)	22 (27.5)
Basic application of skill/concept	248 (52.1)	34 (42.5)	44 (55)	44 (55)	28 (58.3)	7 (43.8)	6 (37.5)	8 (42.1)	25 (67.6)	8 (40)	44 (55)
Strategic thinking	143 (30)	26 (32.5)	18 (22.5)	30 (37.5)	20 (41.7)	5 (31.3)	5 (31.3)	8 (42.1)	9 (24.3)	8 (40)	14 (17.5)
Type of question options and choices											
SAMC[e]	439 (92.2)	78 (97.5)	76 (95)	75 (93.8)	48 (100)	11 (68.8)	13 (81.3)	19 (100)	30 (81.1)	20 (100)	69 (86.3)
SAMRMC[f]	37 (7.8)	2 (2.5)	4 (5)	5 (6.3)	0 (0)	5 (31.3)	3 (18.8)	0 (0)	7 (18.9)	0 (0)	11 (13.8)
Clinical vignette											
Without clinical vignette	362 (76.1)	63 (78.8)	63 (78.8)	61 (76.3)	22 (45.8)	7 (43.8)	14 (87.5)	13 (68.4)	29 (78.4)	16 (80)	74 (92.5)
With clinical vignette	114 (23.9)	17 (21.3)	17 (21.3)	19 (23.8)	26 (54.2)	9 (56.3)	2 (12.5)	6 (31.6)	8 (21.6)	4 (20)	6 (7.5)
Polarity of question options											
Positive	180 (37.8)	22 (27.5)	27 (33.8)	36 (45)	21 (43.8)	3 (18.8)	5 (31.3)	9 (47.4)	8 (21.6)	13 (65)	36 (45)
Negative	296 (62.2)	58 (72.5)	53 (66.3)	44 (55)	27 (56.3)	13 (81.3)	11 (68.8)	10 (52.6)	29 (78.4)	7 (35)	44 (55)

[a]TCM: traditional Chinese medicine.

[b]GYN/OBS: gynecology/obstetrics.

[c]LOTS: lower-order thinking skills.

[d]HOTS: higher-order thinking skills.

[e]SAMC: single-answer multiple-choice.

[f]SAMRMC: single-answer, multiple-response multiple-choice.

XSL•FO
RenderX

Figure 2. Distribution of subjects in TCM licensing examinations. The detailed numbers and proportion of each subject's question types can be seen in Table 1. ENT: ears, nose, and throat; GYN/OBS: gynecology/obstetrics; HOTS: higher-order thinking skills; LOTS: lower-order thinking skills; SAMC: single-answer multiple-choice; SAMRMC: single-answer, multiple-response multiple-choice; TCM: traditional Chinese medicine.

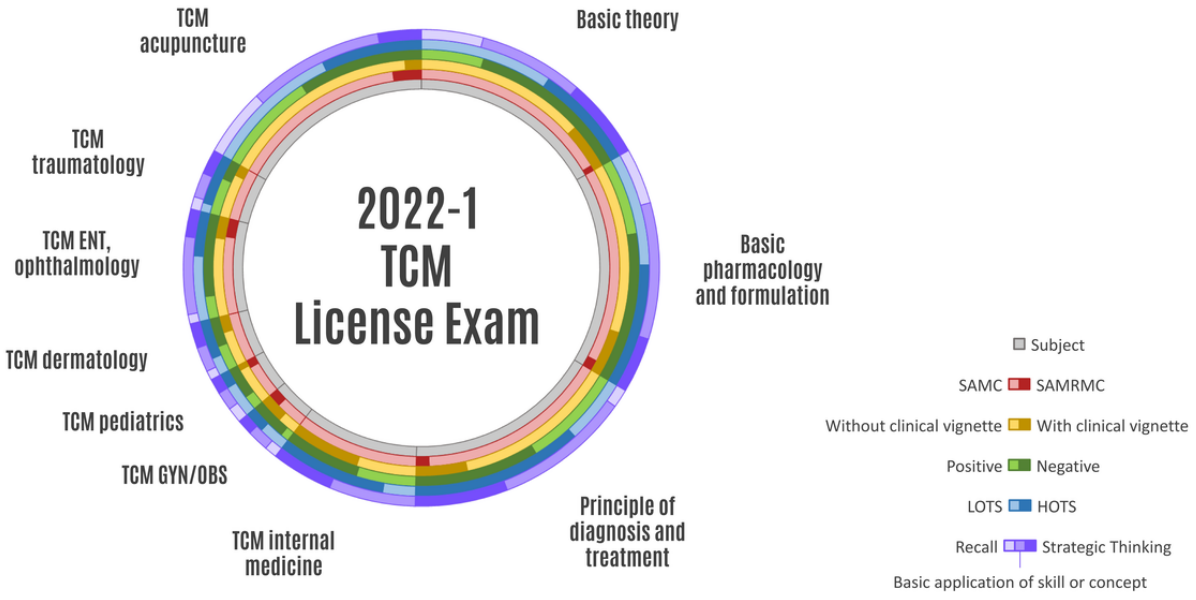
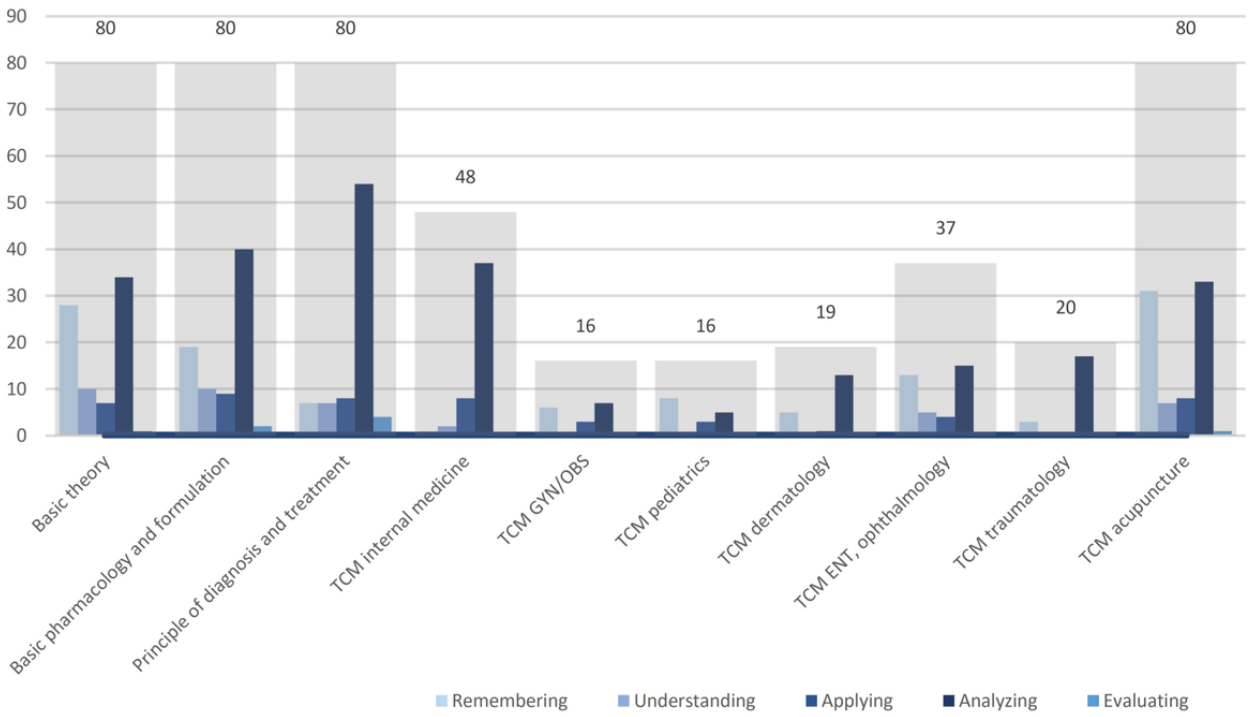


Figure 3. Analysis of question types according to Bloom's cognitive level in TCM licensing examinations. ENT: ears, nose, and throat; TCM: traditional Chinese medicine.



GPT-4 Model Performance and Accuracy Across Different Question Characteristics

We observed that the performance of the GPT-4 model was inferior to that of humans and did not demonstrate significant variation across different categories of examination questions. The GPT-4 model demonstrated an overall accuracy of only 43.9% (209/476). In comparison, 2 human evaluators achieved accuracy rates of 70% (333/476) and 78.4% (373/476), respectively (Table 2). The performance of ChatGPT across various variables is shown in Table 3. The accuracy of AI-generated answers did not show a significant correlation

XSL•FO
RenderX

with the characteristics of the questions, regardless of the classification method used (Figure 4). The GPT-4 model demonstrated a performance close to that of humans in TCM dermatology and TCM traumatology. The accuracy of AI-generated answers varied among the test subjects, ranging from 31.3% in TCM pediatrics to 73.7% in TCM dermatology. Notably, only TCM internal medicine (adjusted odds ratio [aOR] 3.07, 95% CI 1.41 - 6.68; *P*=.005), TCM dermatology (aOR 5.11, 95% CI 1.65 - 15.85; *P*=.005), and TCM acupuncture (aOR 2.14, 95% CI 1.12 - 4.11; *P*=.02) showed statistically significant better performance (Figure 4). On the other hand, GPT had a higher, but not statistically significant, accuracy rate for questions categorized as LOTS (96/213, 45.1%), SAMC (197/439, 44.9%), strategic thinking (66/143, 46.2%), with clinical vignette (52/114, 45.6%), and positive-choice (85/180, 47.2%).

Table . Accuracy rates of testers and ChatGPT-4 for TCM[a] licensing examinations.

	Number of questions	Number of correct responses	Accuracy, %
Human-made 1	476	333	70
Human-made 2	476	373	78.4
ChatGPT-4[b]	476	209	43.9
Human-curated answer 1	476	192	40.3
Human-curated answer 2	476	186	39.1

[a]TCM: traditional Chinese medicine.

[b]ChatGPT did not show answers to 7 questions although an explanation was provided.

XSL•FO
RenderX

Table . Accuracy rates of testers and ChatGPT-4 across different types and subjects of questions.

	Accuracy, %				
	Human-made 1	Human-made 2	ChatGPT-4	Human-curated 1	Human-curated 2
Bloom's cognitive level					
LOTS[a]	150 (70.4)	164 (77)	96 (45.1)	78 (36.6)	75 (35.2)
HOTS[b]	183 (69.6)	209 (79.5)	113 (43)	114 (43.3)	111 (42.2)
Depth of knowledge					
Recall	57 (67.1)	65 (76.5)	34 (40)	27 (31.8)	22 (25.9)
Basic application of skill/concept	172 (69.4)	193 (77.8)	109 (44)	103 (41.5)	102 (41.1)
Strategic thinking	104 (72.7)	115 (80.4)	66 (46.2)	62 (43.4)	62 (43.4)
Type of questions					
SAMC[c]	312 (71.1)	346 (78.8)	197 (44.9)	180 (41)	176 (40.1)
SAMRMC[d]	21 (56.8)	27 (73)	12 (32.4)	12 (32.4)	10 (27)
Vignette style question					
Without clinical vignette	248 (68.5)	283 (78.2)	157 (43.4)	143 (39.5)	137 (37.8)
With clinical vignette	85 (74.6)	90 (78.9)	52 (45.6)	49 (43)	49 (43)
Polarity of question					
Positive	129 (71.7)	142 (78.9)	85 (47.2)	78 (43.3)	76 (42.2)
Negative	204 (68.9)	231 (78)	124 (41.9)	114 (38.5)	110 (37.2)
Subjects					
Basic theory	51 (63.7)	63 (78.8)	29 (36.3)	29 (36.3)	28 (35)
Basic pharmacology and formulation	63 (78.8)	66 (82.5)	30 (37.5)	32 (40)	28 (35)
Principle of diagnosis and treatment	57 (71.3)	58 (72.5)	29 (36.3)	29 (36.3)	29 (36.3)
TCM[e] internal medicine	41 (85.4)	44 (91.7)	30 (62.5)	24 (50)	24 (50)
TCM gynecology and obstetrics	10 (62.5)	12 (75)	8 (50)	4 (25)	4 (25)
TCM pediatrics	11 (68.8)	13 (81.3)	5 (31.3)	7 (43.8)	7 (43.8)
TCM dermatology	14 (73.7)	17 (89.5)	14 (73.7)	12 (63.2)	12 (63.2)
TCM ENT[f], ophthalmology	21 (56.8)	26 (70.3)	12 (32.4)	13 (35.1)	13 (35.1)
TCM traumatology	9 (45)	14 (70)	9 (45)	8 (40)	8 (40)
TCM acupuncture	56 (70)	60 (75)	43 (53.8)	34 (42.5)	33 (41.3)

[a]LOTS: lower-order thinking skills.

[b]HOTS: higher-order thinking skills.

[c]SAMC: single-answer multiple-choice.

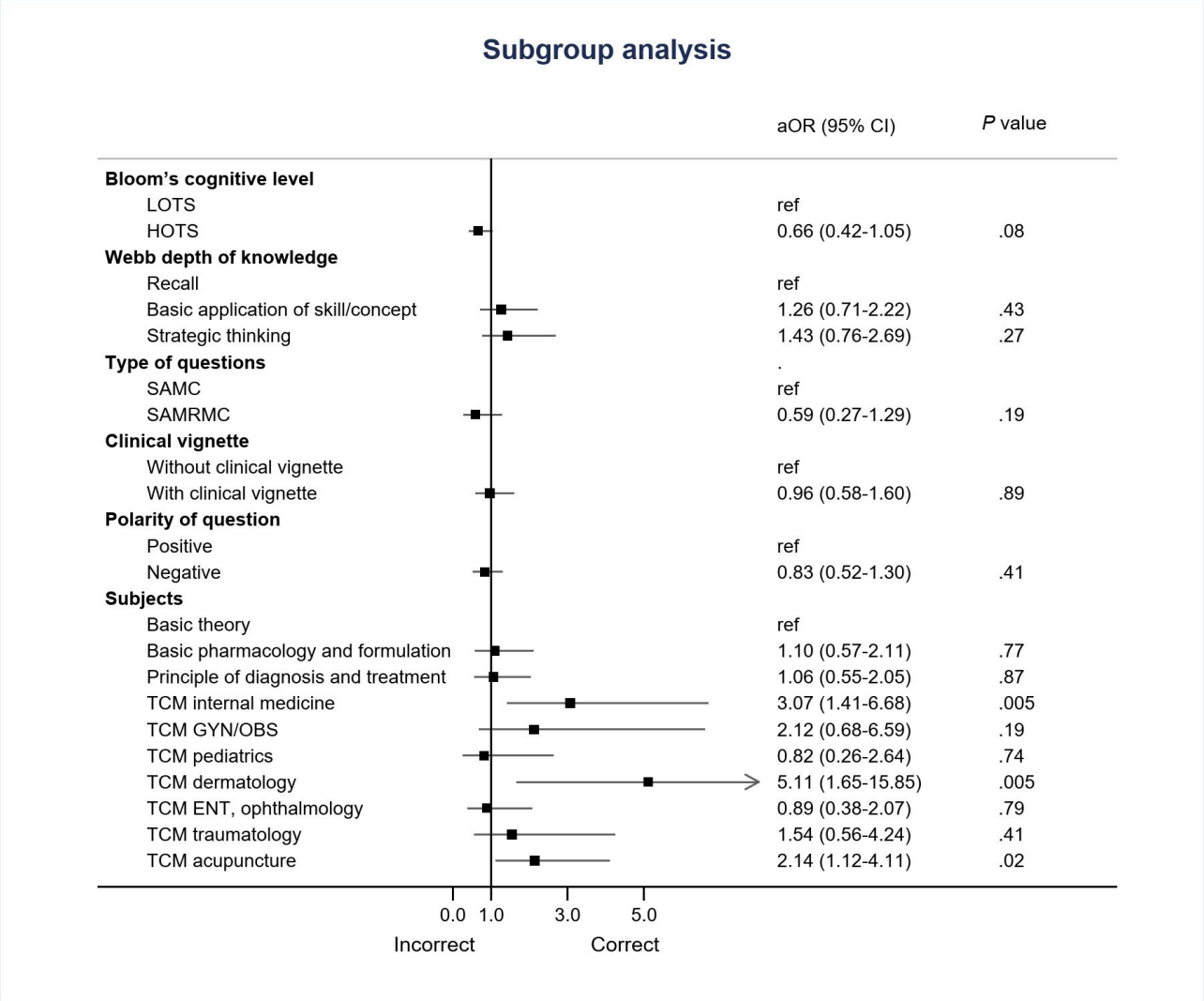
[d]SAMRMC: single-answer, multiple-response multiple-choice.

[e]TCM: traditional Chinese medicine.

[f]ENT: ears, nose, and throat.

XSL•FO
RenderX

Figure 4. Factors associated with correct answers provided by ChatGPT-4. aOR: adjusted odds ratio; ENT: ears, nose, and throat; GYN/OBS: gynecology/obstetrics; HOTS: higher-order thinking skills; LOTS: lower-order thinking skills; SAMC: single-answer multiple-choice; SAMRMC: single-answer, multiple-response multiple-choice; TCM: traditional Chinese medicine.



Consistency Between AI-Generated Answers and Human-Curated Answers and Analysis of Incorrect Responses Provided by the GPT-4 Model

The consistency between AI-generated and human-curated results was low (κ=0.504; Figure 5). After human review, the accuracy of the human-curated answers showed an overall trend of slight decrease, except for some minor increases in basic pharmacology and formulation, TCM pediatrics, and TCM otorhinolaryngology and ophthalmology. The accuracies for the remaining specialties were slightly lower, ranging from 43.9% to 40.3% (Table 2, Figures 5 and 6). For human reviewer 1, discrepancies were observed between AI-generated responses and those reviewed by humans, with 23.96% (115 of 480 questions) of the answers provided by AI conflicting with its own explanations. For 33% of correctly answered questions (69 of 209 questions), the AI provided an incorrect explanation, indicating a scenario of "correct answer, incorrect explanation." Conversely, for 17% of incorrectly answered questions (46 of 267 questions), the AI provided a correct explanation, suggesting a case of "incorrect answer, correct explanation." This reduced the overall accuracy of the AI model to 43.9%.

XSL•FO
RenderX

Figure 5. Accuracy rates of humans and ChatGPT-4 for TCM licensing examinations. The passing standard is an average score of 60. With 476 questions, the threshold is at least 286 correct answers (red dashed line). AI: artificial intelligence; TCM: traditional Chinese medicine.

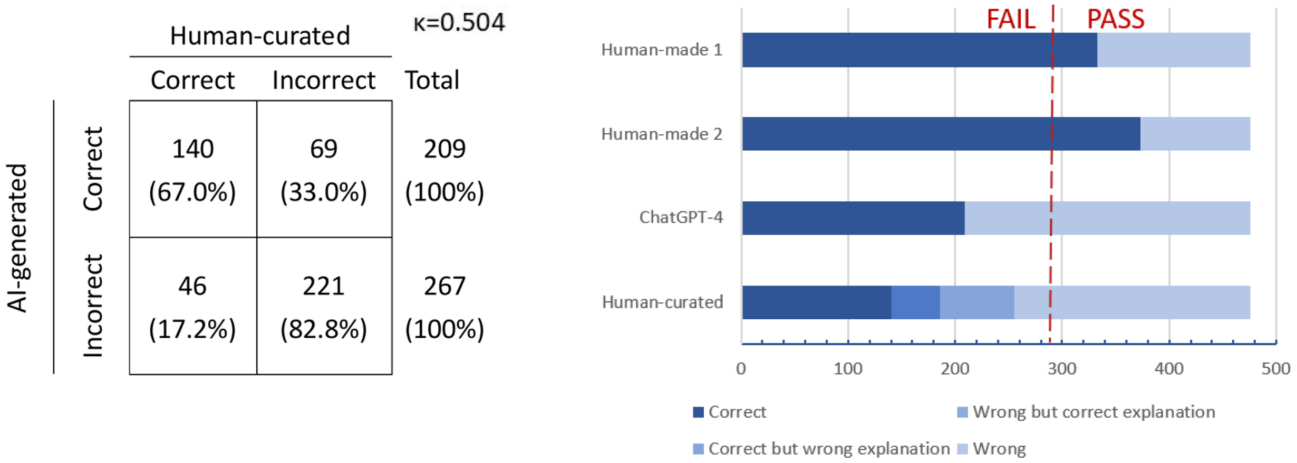
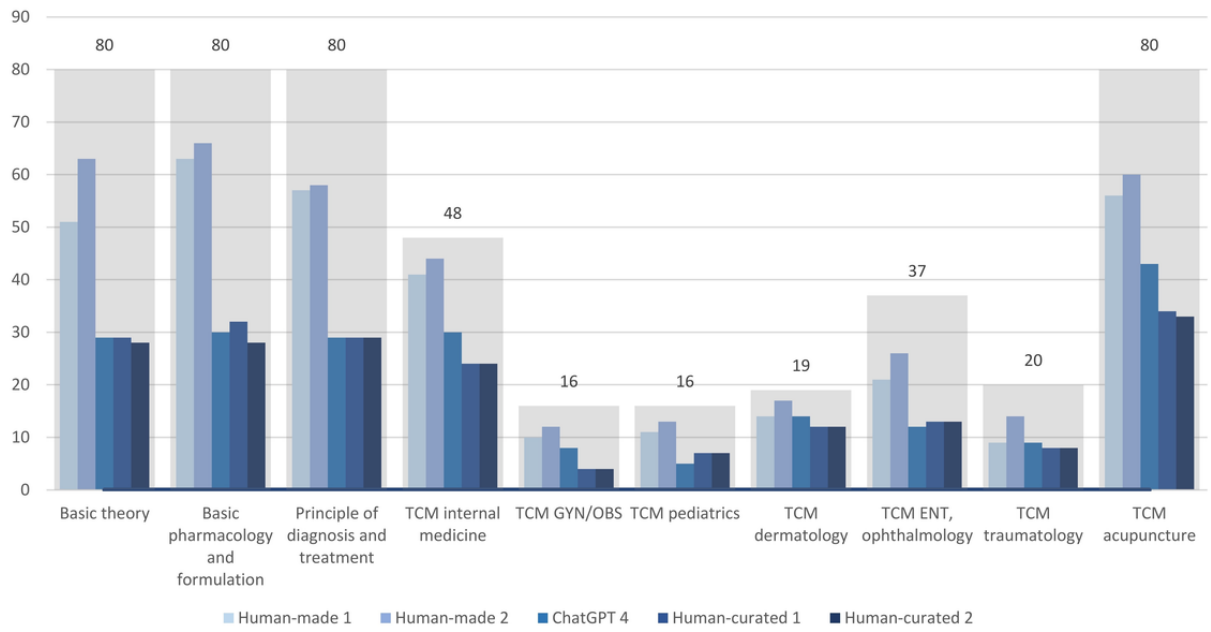


Figure 6. Performance of humans and ChatGPT-4 across various subjects. ENT: ears, nose, and throat; GYN/OBS: gynecology/obstetrics; TCM: traditional Chinese medicine.



We further analyzed the reasons responsible for the incorrect answers provided by the GPT. For this purpose, we categorized the potential reasons for these errors into 3 types: misinterpretation of the question (failing to understand the question), misunderstanding of concepts (lacking knowledge of the topic), and incorrect application of principles (the content is correct, but it does not answer the question). The results revealed that most of the errors (263/476, 55.3%) were attributed to the misunderstanding of concepts (Table 4, Figure 7). However, a closer examination of the different characteristics of the questions indicated that misunderstanding of concepts was more common in LOTS, recall, and SAMRMC compared to their counterparts. The second most common cause of error was incorrect application of principles (20/476, 4.2%), followed by misinterpretation of questions (7/476, 1.5%).

XSL•FO
RenderX

Table . Reasons responsible for incorrect artificial intelligence–generated responses (human-curated).

	Correct (n=186)	Misinterpretation of the question (n=7)	Misunderstanding of concepts (n=263)	Incorrect application of principles (n=20)	*P* value
Bloom's cognitive level					.25
LOTS[a]	75 (40.3)	5 (71.4)	124 (47.1)	9 (45)	
HOTS[b]	111 (59.7)	2 (28.6)	139 (52.9)	11 (55)	
Depth of knowledge					.06
Recall	22 (11.8)	0 (0)	60 (22.8)	3 (15)	
Basic application of skill/concept	102 (54.8)	5 (71.4)	132 (50.2)	9 (45)	
Strategic thinking	62 (33.3)	2 (28.6)	71 (27)	8 (40)	
Type of questions					.16
SAMC[c]	176 (94.6)	6 (85.7)	237 (90.1)	20 (100)	
SAMRMC[d]	10 (5.4)	1 (14.3)	26 (9.9)	0 (0)	
Vignette style question					.39
Without clinical vignette	137 (73.7)	6 (85.7)	206 (78.3)	13 (65)	
With clinical vignette	49 (26.3)	1 (14.3)	57 (21.7)	7 (35)	
Polarity of question					.28
Positive	76 (40.9)	1 (14.3)	98 (37.3)	5 (25)	
Negative	110 (59.1)	6 (85.7)	165 (62.7)	15 (75)	
Subjects					<.001
Basic theory	28 (15.1)	0 (0)	35 (13.3)	17 (85)	
Basic pharmacology and formulation	28 (15.1)	1 (14.3)	49 (18.6)	2 (10)	
Principle of diagnosis and treatment	29 (15.6)	3 (42.9)	48 (18.3)	0 (0)	
TCM[e] internal medicine	24 (12.9)	0 (0)	24 (9.1)	0 (0)	
TCM gynecology and obstetrics	4 (2.2)	0 (0)	12 (4.6)	0 (0)	
TCM pediatrics	7 (3.8)	1 (14.3)	8 (3.0)	0 (0)	
TCM dermatology	12 (6.5)	0 (0)	7 (2.7)	0 (0)	
TCM ENT[f], ophthalmology	13 (7)	1 (14.3)	22 (8.4)	1 (5)	
TCM traumatology	8 (4.3)	0 (0)	12 (4.6)	0 (0)	
TCM acupuncture	33 (17.7)	1 (14.3)	46 (17.5)	0 (0)	

[a]LOTS: lower-order thinking skills.

[b]HOTS: higher-order thinking skills.

[c]SAMC: single-answer multiple-choice.

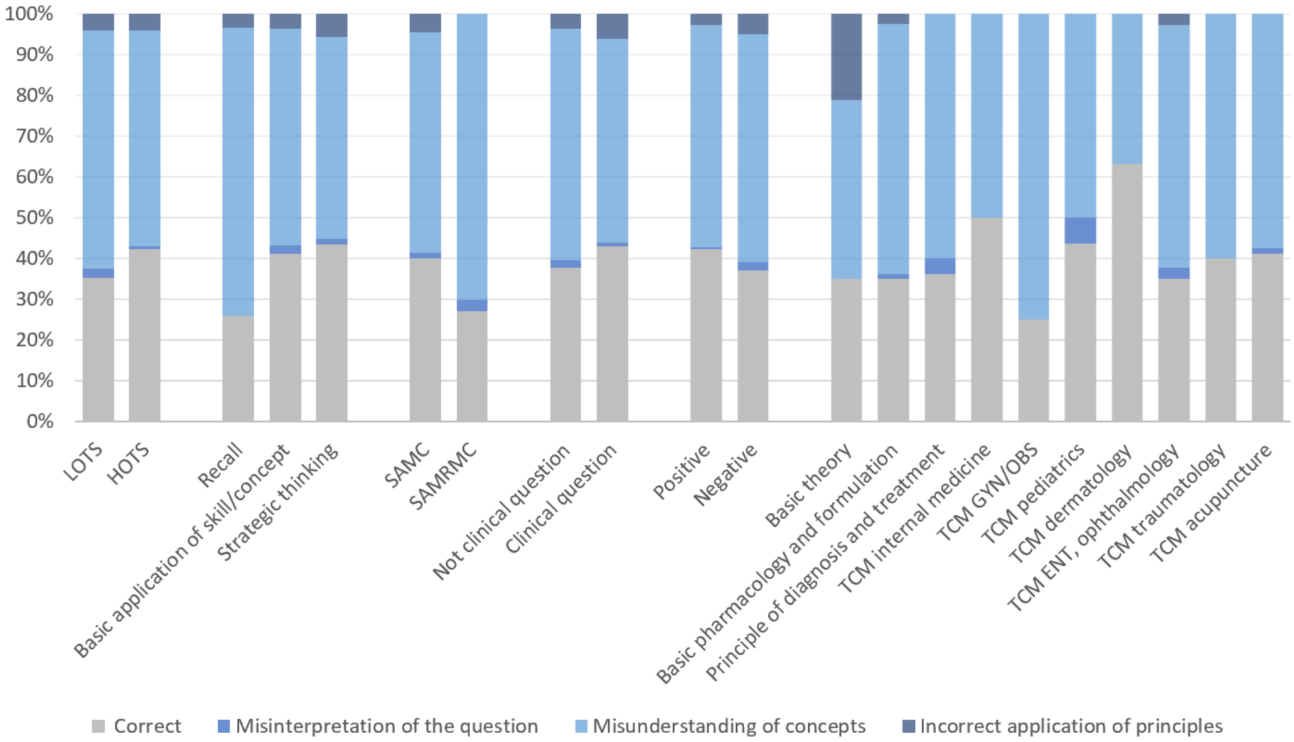
[d]SAMRMC: single-answer, multiple-response multiple-choice.

[e]TCM: traditional Chinese medicine.

[f]ENT: ears, nose, and throat.

XSL•FO
RenderX

Figure 7. Distribution of reasons for incorrect answers provided by ChatGPT-4. ENT: ears, nose, and throat; GYN/OBS: gynecology/obstetrics; HOTS: higher-order thinking skills; LOTS: lower-order thinking skills; SAMC: single-answer multiple-choice; SAMRMC: single-answer, multiple-response multiple-choice; TCM: traditional Chinese medicine.



Discussion

Performance of ChatGPT in Medical Examinations

This is the first study to test the capabilities of ChatGPT in TCM examinations. ChatGPT has undergone rigorous testing for its proficiency in medical examinations. Nonetheless, its effectiveness in TCM licensing examinations remains unexplored. Hence, this study fills a research void by examining the capability of an advanced language model like ChatGPT in the context of TCM. Generally, most studies indicate ChatGPT can meet the medical examination pass standards. For example, ChatGPT 3.5 scored around the pass mark on the United States Medical Licensing Examination [14] and exhibited strong performance in specialties such as radiation oncology and neurosurgery [27,28]. GPT-4 surpassed 70% in its score for UK medical licensing examinations [12], and its competency extends to examinations in different languages. For example, GPT 3.5 typically scored around the passing mark on the Japanese nursing examinations [16] and Korean medical student parasitology examinations [29]. Although GPT-3.5 Turbo is not yet capable, GPT-4 passed the medical licensing examinations of China [30,31] and achieved 88.6% accuracy in the equivalent examinations of Saudi Arabia [32]. Interestingly, it even outperformed human residents in the residency training examinations of Japan [33].

Published research has identified 2 trends in this setting. First, GPT-4 surpasses GPT-3.5 in identical medical examinations, as demonstrated in medical student finals in Poland [34] and the medical licensing examinations of Peru [35]. A systematic review and meta-analysis of ChatGPT use in medical licensing examinations worldwide observed similar results [36]. Second, ChatGPT models showed higher accuracy when answering questions translated into English compared with the original language [34,37]. In Taiwan, traditional Chinese is the language used for medical licensing examinations. Despite this disadvantage, ChatGPT performed near the pass threshold for the nursing [38] and pharmacy licensing examinations in Taiwan [15]; translating pharmacy examination questions into English indeed improved scores across all subjects [15]. Thus, it was hypothesized that GPT-4 would perform similarly in TCM licensing examinations. However, the results were surprising. The study used the first 2022 TCM licensing examinations in Taiwan as a case study to assess the performance of the model. GPT-4 failed the exam with an overall accuracy of 43.9%; following human revision of AI-provided explanations, the accuracy further decreased to 40.3% (human 1) and 39.1% (human 2). These results underscore the need for further research and development on the application of AI models to TCM examination preparation and highlight the existing knowledge gap. The reasons behind these outcomes merit further investigation.

Challenges Encountered by ChatGPT When Answering Medical Questions

Previous literature has discussed the shortcomings and challenges of ChatGPT in answering examination questions, including a decreased proficiency in languages other than English [34,37], AI “hallucinations” originating from erroneous data [10,38], and proficiency limited to certain types of questions [13,39]. The tendency for ChatGPT to be less proficient in answering questions posed in languages other than English stems from the fact that ChatGPT is an LLM trained primarily on English language data, which includes a wide variety of sources such as books, websites, and news articles [6]. The questions for TCM licensing examinations are not presented in

XSL•FO
RenderX

English. Although ChatGPT can fluently interact in traditional Chinese, its responses to medical examination questions, which require specific expertise and have standard answers, may reveal its inadequacies. AI "hallucinations" indicate a tendency to produce "hallucinations" or factually incorrect content due to incorrect data. This poses the risk of generating misleading or fabricated information, which complicates the use of AI as a reliable self-learning tool [7,10]. We also encountered seemingly plausible but incorrect content in AI-generated responses in our research. We even found that verifying the authenticity of these answers is more time-consuming and requires deeper professional knowledge than the questions themselves. Our study also showed that ChatGPT had higher, albeit not statistically significant, accuracy rates for questions posed such as SAMC (n=197, 44.9%) and presented with clinical vignettes (n=52, 45.6%). This trend aligns with findings of previous studies, such as a lower proficiency in multiple-choice questions [13] and a poorer aptitude for conceptual questions compared with clinical scenarios [39]. Despite these limitations, which we have also encountered, other research has shown that ChatGPT can pass examinations. Therefore, the use of ChatGPT in the context of TCM may pose its own unique set of challenges and necessitates further investigation.

Challenges Encountered by ChatGPT When Answering TCM Examination Questions

We identified 3 main reasons for incorrect answers according to AI-generated responses, namely misinterpretation of the question, misunderstanding of concepts, and incorrect application of principles. Misunderstanding of concepts was the most prevalent, especially in questions with lower cognitive demand such as recall and LOTS, as well as in questions where a single item encompasses multiple questions (eg, SAMRMC), indicating either a lack of knowledge or incorrect knowledge. We believe that this primarily stems from 2 factors. First, the database for TCM is currently incomplete. Second, compared with Western medicine, TCM is often considered alternative medicine. If an LLM such as ChatGPT answers questions based solely on the Western medical knowledge system, then TCM content may be ignored. Additionally, TCM focuses on personalized treatment without a golden standard, leading to the absence of definitive answers for the same disease.

The incomplete TCM database is due to challenges such as insufficient data, lack of standardization, and unrepresentative data sources. Although the specific TCM data that ChatGPT uses for training are unclear, it is evident that the current online data for TCM are significantly less comprehensive than those for Western medicine. For instance, a bibliometric analysis over the past 20 years did not show a significant presence of TCM-related keywords in the context of pediatric allergic rhinitis [40]. However, the usage rate of TCM for allergic diseases in Taiwan is approximately 30% - 50% [41]. Therefore, a model constructed based on such a database is likely to exhibit discrepancies with reality. Furthermore, online data often contain inaccuracies or incomplete information. Previous research has shown that uncleaned training texts can affect performance and could underpin the subpar performance of the trained model [42].

It is important to note that, due to challenges in translation and cultural appropriation, certain medical terms have different connotations in the TCM and Western medical systems. However, ChatGPT tends to interpret these terms with a preference for their meanings within Western medicine. For instance, in some AI-generated responses, the TCM term for "肝" was mistakenly translated and described as the physical organ "liver" in Western medicine. Similarly, the term for "瘧" in TCM was translated and described as "malaria" in some AI-generated responses. The understanding of "肝" in TCM is not entirely the same as in modern medicine, and "瘧" in TCM refers to a broad category of symptoms similar to malaria but not restricted to infections caused by *Plasmodium*.

The crux of TCM is personalized treatment, which is antithetical to gold-standard treatments. Hence, multiple therapeutic approaches may exist for the same disease. If the examination questions do not specify a particular scope or clear criteria, there may be no standard answer or multiple possible solutions. This study revealed that the decrease in the overall accuracy rate after human review was primarily driven by a reduction in accuracy for LOTS questions, whereas the accuracy rate for HOTS remained stable or even increased. Regarding DOK, the decrease in accuracy following human review was primarily in recall, with less of a decrease noted in more advanced DOK (eg, basic application of skill/concept, strategic thinking). This suggests that GPT-4 is more adept at providing detailed explanations for complex logical reasoning questions, as opposed to simple memorization, which might be influenced by incorrect information. In addition, if users intend to use GPT to answer TCM questions, they should be particularly cautious of potential hallucinations in lower cognitive demand questions.

Our study revealed that the GPT-4 model is currently unable to pass the TCM licensing examinations. This research underscores the limitations of the performance of AI in TCM licensing examinations, as well as illuminates broader challenges within the realm of integrating TCM knowledge into AI development.

Limitations

Although this study provides valuable insights into the use of the GPT-4 model for TCM licensing examination preparation, several limitations have been identified. The focus solely on the GPT-4 model of ChatGPT might neglect the complexities and potential capabilities of other recently developed AI-driven language models, such as Claude 3 by Anthropic, Bard (Gemini Pro) by Google, or LLaMa2 by Facebook. Notably, we did not use expert-level AI, such as Med-PaLM by Google [43]. Moreover, we did not use other traditional Chinese-language LLMs, such as Taiwan-LLM [44,45]. Nevertheless, GPT models are the most widely used and studied models, and it is necessary to use the same tool to facilitate comparisons with other research studies [36].

Considering the cultural context specific to the TCM licensing examination of Taiwan, the generalizability of our findings to different regions or educational systems may be limited. Notably, model performance may change over time, indicating that our results may not be replicated in the future. This study also did not account for potential inconsistencies in responses provided by ChatGPT to identical queries during different

XSL•FO
RenderX

sessions. However, this issue could be minimized by explicitly setting the parameters of ChatGPT.

Additionally, the difficulty of each exam can vary, which might affect ChatGPT's performance. However, the difficulty is generally controlled and, as a national exam, the pass rates have been stable over the years [46]. Previous exam questions could potentially be part of the GPT model's training data (with a knowledge cutoff date of September 2021), introducing bias. Therefore, we only used the first exam of 2022 to mitigate this issue.

Implications for Practice and Future Research

This study investigated the use of the GPT-4 model for TCM licensing examination preparation. The findings revealed that AI-driven tools are not yet valuable assets for TCM educators and students. The observed limitations (ie, often providing responses based on incorrect facts) highlight the need for further development before this model can be effectively used as a self-learning tool. As the AI field continues to advance with the introduction of new models, educators must stay informed and utilize the most effective tools while being cognizant of their limitations. This study sets the stage for 2 potential research directions. In terms of TCM, considering the suboptimal examination results, we speculate that the primary drawback lies in the quality of the front-end data. Future improvements may include incorporating ancient TCM texts and customizing training for LLMs.

We must deliberately incorporate relevant resources into our training database materials, such as textbooks on TCM in Chinese and ancient TCM texts. Currently, the majority of descriptions and knowledge regarding TCM are in Chinese. When these data are published in journals or translated into English, they often adopt the framework and language of modern medicine as a medium for knowledge transmission. This approach tends to underemphasize the original content of TCM, which is mostly documented in Chinese literature. Therefore, the inclusion of TCM materials in LLM training and the standardization of TCM should be targeted for improvement.

Tailoring training data for LLMs presents another promising avenue for improvement. TCM comprises different schools, suggesting that narrowing the knowledge domain could be more advantageous. Hence, to excel in TCM, developing specialized ChatGPT models or custom LLMs might be a beneficial strategy. Considering the current limitations in enhancing the database, integrating specific prompts offers an alternative solution. For example, the chain-of-thoughts method, used in LLMs for complex problem-solving, articulates intermediate steps in reasoning. This approach is particularly effective for models with extensive parameters, enhancing their ability to manage multistep tasks [26]. It has been confirmed that this method can also improve the performance of ChatGPT in medical examinations [47]. Hence, the adoption of chain-of-thoughts may be a viable strategy to address the complexity of TCM examinations. Additionally, previous research indicated that restricting ChatGPT to a single response in a Basic Life Support examination may introduce bias. When ChatGPT generates 3 responses per question, it successfully passes the examination. Moreover, rephrasing incorrectly answered questions as open-ended questions significantly boosts the accuracy of ChatGPT. This implies that open-ended questioning or multiple inquiries might be more effective than single-choice formats [48].

Conclusion

Our study represents the first comprehensive assessment of the performance of ChatGPT in TCM licensing examinations. Despite advances in AI and its success in various medical licensing tests, ChatGPT demonstrated a limited ability to accurately respond to TCM examination questions, achieving an overall accuracy rate significantly lower than that of its human counterparts. This shortfall underscores the challenges posed by the unique concepts and terminologies of TCM, highlighting a significant knowledge gap in the understanding of TCM principles by AI. Our findings call for further advancements in AI training, specifically tailored toward the intricate domain of TCM, to enhance its utility in this specialized field of medicine.

Acknowledgments

This study was partially supported by Chang Gung Medical Foundation (grant CGRPG1Q0011), the Ministry of Health and Welfare (grants MOHW112-CMAP-M-113-000006-D, MOHW113-CMAP-M-113-000002-D, and MOHW113-CMAP-M-113-000003-B), and the National Science and Technology Council in Taiwan (grant MOST111-2320-B-182-035-MY3).

Authors' Contributions

LWT contributed to manuscript writing. HYC and YCC were responsible for the statistical analysis, project administration, funding acquisition, manuscript revision, and study design. Results were interpreted by LCT and YCL. HYC and YCC contributed equally as co-corresponding authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of the 5 factors, data definitions, and source citations.

XSL•FO
RenderX

[DOCX File, 18 KB - mededu_v11i1e58897_app1.docx]

Multimedia Appendix 2
Examples of single-answer multiple-choice and single-answer, multiple-response multiple-choice questions.
[DOCX File, 16 KB - mededu_v11i1e58897_app2.docx]

Multimedia Appendix 3
Examples of the prompt used to generate responses from questions.
[DOCX File, 57 KB - mededu_v11i1e58897_app3.docx]

Multimedia Appendix 4
Examples of the prompts used to generate responses from questions with explanations for each item.
[DOCX File, 179 KB - mededu_v11i1e58897_app4.docx]

References

1. Chi C. Integrating traditional medicine into modern health care systems: examining the role of Chinese medicine in Taiwan. Soc Sci Med 1994 Aug;39(3):307-321. [doi: 10.1016/0277-9536(94)90127-9] [Medline: 7939847]
2. Chi C, Lee JL, Lai JS, Chen CY, Chang SK, Chen SC. The practice of Chinese medicine in Taiwan. Soc Sci Med 1996 Nov;43(9):1329-1348. [doi: 10.1016/0277-9536(95)00429-7] [Medline: 8913003]
3. Park YL, Huang CW, Sasaki Y, Ko Y, Park S, Ko SG. Comparative study on the education system of traditional medicine in China, Japan, Korea, and Taiwan. Expl NY 2016;12(5):375-383. [doi: 10.1016/j.explore.2016.06.004] [Medline: 27546589]
4. Wang Y, Shi X, Li L, Efferth T, Shang D. The impact of artificial intelligence on traditional Chinese medicine. Am J Chin Med 2021;49(6):1297-1314. [doi: 10.1142/S0192415X21500622] [Medline: 34247564]
5. Li W, Ge X, Liu S, Xu L, Zhai X, Yu L. Opportunities and challenges of traditional Chinese medicine doctors in the era of artificial intelligence. Front Med (Lausanne) 2023;10:1336175. [doi: 10.3389/fmed.2023.1336175] [Medline: 38274445]
6. Brown TB, Mann B, Ryder N, et al. Language models are few-shot learners. 2020 Presented at: Proceedings of the 34th International Conference on Neural Information Processing Systems; Dec 6-12, 2020; Vancouver, BC, Canada.
7. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. N Med 2023 Aug;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
8. Han JW, Park J, Lee H. Analysis of the effect of an artificial intelligence chatbot educational program on non-face-to-face classes: a quasi-experimental study. BMC Med Educ 2022 Dec 1;22(1):830. [doi: 10.1186/s12909-022-03898-3] [Medline: 36457086]
9. Branum C, Schiavenato M. Can ChatGPT accurately answer a PICOT question? Assessing AI response to a clinical question. Nurse Educ 2023;48(5):231-233. [doi: 10.1097/NNE.0000000000001436] [Medline: 37130197]
10. Borji A. A categorical archive of ChatGPT failures. arXiv. Preprint posted online on Feb 6, 2023. [doi: 10.21203/rs.3.rs-2895792/v1]
11. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell 2023;6:1169595. [doi: 10.3389/frai.2023.1169595] [Medline: 37215063]
12. Lai UH, Wu KS, Hsu TY, Kan JKC. Evaluating the performance of ChatGPT-4 on the United Kingdom Medical Licensing Assessment. Front Med (Lausanne) 2023;10:1240915. [doi: 10.3389/fmed.2023.1240915] [Medline: 37795422]
13. Alfertshofer M, Hoch CC, Funk PF, et al. Sailing the seven seas: a multinational comparison of ChatGPT's performance on medical licensing examinations. Ann Biomed Eng 2024 Jun;52(6):1542-1545. [doi: 10.1007/s10439-023-03338-3] [Medline: 37553555]
14. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Dig Health 2023 Feb;2(2):e0000198. [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
15. Wang YM, Shen HW, Chen TJ. Performance of ChatGPT on the pharmacist licensing examination in Taiwan. J Chin Med Assoc 2023 Jul 1;86(7):653-658. [doi: 10.1097/JCMA.0000000000000942] [Medline: 37227901]
16. Taira K, Itaya T, Hanada A. Performance of the large language model ChatGPT on the National Nurse Examinations in Japan: evaluation study. JMIR Nurs 2023 Jun 27;6:e47305. [doi: 10.2196/47305] [Medline: 37368470]
17. Wang H, Wu W, Dou Z, He L, Yang L. Performance and exploration of ChatGPT in medical examination, records and education in Chinese: pave the way for medical AI. Int J Med Inform 2023 Sep;177:105173. [doi: 10.1016/j.ijmedinf.2023.105173] [Medline: 37549499]
18. Yizhen L, Shaohan H, Jiaxing Q, Lei Q, Dongran H, Zhongzhi L. Exploring the comprehension of ChatGPT in Traditional Chinese Medicine knowledge. arXiv. Preprint posted online on Mar 14, 2024. [doi: 10.48550/arXiv.2403.09164]
19. Jang D, Yun TR, Lee CY, Kwon YK, Kim CE. GPT-4 can pass the Korean National Licensing Examination for Korean Medicine Doctors. PLOS Dig Health 2023 Dec;2(12):e0000416. [doi: 10.1371/journal.pdig.0000416] [Medline: 38100393]

XSL•FO
RenderX

20. Taiwan Ministry of Examination - Examination Question Inquiry Platform [Website in Mandarin]. URL: https://wwwq.moex.gov.tw/exam/wFrmExamQandASearch.aspx [accessed 2025-01-04]
21. Zaidi NB, Hwang C, Scott S, Stallard S, Purkiss J, Hortsch M. Climbing Bloom's taxonomy pyramid: lessons from a graduate histology course. Anat Sci Ed 2017 Sep;10(5):456-464. [doi: 10.1002/ase.1685]
22. Krathwohl DR. A revision of Bloom's taxonomy: an overview. Theor Pract 2002 Nov 1;41(4):212-218. [doi: 10.1207/s15430421tip4104_2]
23. Webb NL. Depth-of-knowledge levels for four content areas. LA 2002;28(March):1-9.
24. Giannos P, Delardas O. Performance of ChatGPT on UK standardized admission tests: insights from the BMAT, TMUA, LNAT, and TSA examinations. JMIR Med Educ 2023 Apr 26;9:e47737. [doi: 10.2196/47737] [Medline: 37099373]
25. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 8;9:e45312. [doi: 10.2196/45312] [Medline: 36753318]
26. Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models. 2022 Presented at: 36th Conference on Neural Information Processing Systems; Nov 28 to Dec 9, 2022; New Orleans, LA, USA p. 24824-24837.
27. Huang Y, Gomaa A, Semrau S, et al. Benchmarking ChatGPT-4 on a radiation oncology in-training exam and Red Journal Gray Zone cases: potentials and challenges for ai-assisted medical education and decision making in radiation oncology. Front Oncol 2023;13:1265024. [doi: 10.3389/fonc.2023.1265024] [Medline: 37790756]
28. Ali R, Tang OY, Connolly ID, et al. Performance of ChatGPT and GPT-4 on neurosurgery written board examinations. Neurosurgery 2023 Dec 1;93(6):1353-1365. [doi: 10.1227/neu.0000000000002632] [Medline: 37581444]
29. Huh S. Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: a descriptive study. J Educ Eval Health Prof 2023;20:1. [doi: 10.3352/jeehp.2023.20.1] [Medline: 36627845]
30. Fang C, Wu Y, Fu W, et al. How does ChatGPT-4 preform on non-English national medical licensing examination? An evaluation in Chinese language. PLOS Dig Health 2023 Dec;2(12):e0000397. [doi: 10.1371/journal.pdig.0000397] [Medline: 38039286]
31. Zong H, Li J, Wu E, Wu R, Lu J, Shen B. Performance of ChatGPT on Chinese national medical licensing examinations: a five-year examination evaluation study for physicians, pharmacists and nurses. BMC Med Educ 2024 Feb 14;24(1):143. [doi: 10.1186/s12909-024-05125-7] [Medline: 38355517]
32. Aljindan FK, Al Qurashi AA, Albalawi IAS, et al. ChatGPT conquers the Saudi Medical Licensing Exam: exploring the accuracy of artificial intelligence in medical knowledge assessment and implications for modern medical education. Cureus 2023 Sep;15(9):e45043. [doi: 10.7759/cureus.45043] [Medline: 37829968]
33. Watari T, Takagi S, Sakaguchi K, et al. Performance comparison of ChatGPT-4 and Japanese medical residents in the General Medicine In-Training Examination: comparison study. JMIR Med Educ 2023 Dec 6;9:e52202. [doi: 10.2196/52202] [Medline: 38055323]
34. Rosoł M, Gąsior JS, Łaba J, Korzeniewski K, Młyńczak M. Evaluation of the performance of GPT-3.5 and GPT-4 on the Polish Medical Final Examination. Sci Rep 2023 Nov 22;13(1):20512. [doi: 10.1038/s41598-023-46995-z] [Medline: 37993519]
35. Flores-Cohaila JA, García-Vicente A, Vizcarra-Jiménez SF, et al. Performance of ChatGPT on the Peruvian National Licensing Medical Examination: cross-sectional study. JMIR Med Educ 2023 Sep 28;9:e48039. [doi: 10.2196/48039] [Medline: 37768724]
36. Liu M, Okuhara T, Chang X, et al. Performance of ChatGPT across different versions in medical licensing examinations worldwide: systematic review and meta-analysis. J Med Internet Res 2024 Jul 25;26:e60807. [doi: 10.2196/60807] [Medline: 39052324]
37. Wang X, Gong Z, Wang G, et al. ChatGPT performs on the Chinese National Medical Licensing Examination. J Med Syst 2023 Aug 15;47(1):86. [doi: 10.1007/s10916-023-01961-0] [Medline: 37581690]
38. Huang H. Performance of ChatGPT on Registered Nurse License Exam in Taiwan: a descriptive study. Healthcare (Basel) 2023 Oct 30;11(21):2855. [doi: 10.3390/healthcare11212855] [Medline: 37958000]
39. Scaioli G, Lo Moro G, Conrado F, Rosset L, Bert F, Siliquini R. Exploring the potential of ChatGPT for clinical reasoning and decision-making: a cross-sectional study on the Italian Medical Residency Exam. Ann Ist Super Sanita 2023;59(4):267-270. [doi: 10.4415/ANN_23_04_05] [Medline: 38088393]
40. Liu F, Chen N, Wang R, Zhang L, Li Y. Visual analysis of allergic rhinitis in children based on web of science and CiteSpace software. Front Pediatr 2022;10:911293. [doi: 10.3389/fped.2022.911293] [Medline: 36245734]
41. Lin PY, Chu CH, Chang FY, Huang YW, Tsai HJ, Yao TC. Trends and prescription patterns of traditional Chinese medicine use among subjects with allergic diseases: a nationwide population-based study. World Allergy Organ J 2019;12(2):100001. [doi: 10.1016/j.waojou.2018.11.001] [Medline: 30937136]
42. Rejeleene R, Xu X, Talburt J. Towards trustable language models: investigating information quality of large language models. arXiv. Preprint posted online on 2024.

XSL•FO
RenderX

43. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. Nat New Biol 2023 Aug;620(7972):172-180. [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
44. Chen PH, Cheng S, Chen WL, Lin YT, Chen YN. Measuring Taiwanese Mandarin language understanding. arXiv. Preprint posted online on Mar 29, 2024arXiv:2403.20180.
45. Lin YT, Chen YN. Taiwan llm: bridging the linguistic divide with a culturally aligned language model. arXiv. Preprint posted online on Nov 29, 2023arXiv:2311.17487.
46. Taiwan Ministry of Examination - Examination Statistics [Website in Mandarin]. URL: https://wwwc.moex.gov.tw/main/examreport/wfrmexamstatistics.aspx?menu_id=158 [accessed 2025-01-04]
47. Ting YT, Hsieh TC, Wang YF, et al. Performance of ChatGPT incorporated chain-of-thought method in bilingual nuclear medicine physician board examinations. Dig Health 2024;10:20552076231224074. [doi: 10.1177/20552076231224074] [Medline: 38188855]
48. Zhu L, Mou W, Yang T, Chen R. ChatGPT can pass the AHA exams: open-ended questions outperform multiple-choice format. Resuscitation 2023 Jul;188:109783. [doi: 10.1016/j.resuscitation.2023.109783] [Medline: 37349064]

Abbreviations

AI: artificial intelligence
aOR: adjusted odds ratio
DOK: depth of knowledge
ENT: ears, nose, and throat
GPT: generative pretrained transformer
GYN/OBS: gynecology/obstetrics
HOTS: higher-order thinking skills
LLM: large language model
LOTS: lower-order thinking skills
SAMC: single-answer multiple-choice
SAMRMC: single-answer, multiple-response multiple-choice
TCM: traditional Chinese medicine

Edited by B Lesselroth; submitted 27.03.24; peer-reviewed by B Shen, T Kikuchi, Z Hou; revised version received 27.07.24; accepted 09.11.24; published 19.03.25.

Please cite as:
Tseng LW, Lu YC, Tseng LC, Chen YC, Chen HY
Performance of ChatGPT-4 on Taiwanese Traditional Chinese Medicine Licensing Examinations: Cross-Sectional Study
JMIR Med Educ 2025;11:e58897
URL: https://mededu.jmir.org/2025/1/e58897
doi:10.2196/58897

© Liang-Wei Tseng, Yi-Chin Lu, Liang-Chi Tseng, Yu-Chun Chen, Hsing-Yu Chen. Originally published in JMIR Medical Education (https://mededu.jmir.org), 19.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Generative Artificial Intelligence in Medical Education—Policies and Training at US Osteopathic Medical Schools: Descriptive Cross-Sectional Survey

Tsunagu Ichikawa[1*], BS; Elizabeth Olsen[2*], BS, MS; Arathi Vinod[3*], BA; Noah Glenn[4*]; Karim Hanna[5*], MD; Gregg C Lund[*], MS, DO; Stacey Pierce-Talsma[1*], MS, DO

[1]College of Osteopathic Medicine, University of New England, 11 Hills Beach Road, Biddeford, ME, United States
[2]College of Osteopathic Medicine, Rocky Vista University, Parker, CO, United States
[3]College of Osteopathic Medicine, Touro University California, Vallejo, CA, United States
[4]McCombs School of Business, University of Texas at Austin, Austin, TX, United States
[5]Morsani College of Medicine, University of South Florida, Tampa, FL, United States
[*]all authors contributed equally

Corresponding Author:
Stacey Pierce-Talsma, MS, DO
College of Osteopathic Medicine, University of New England, 11 Hills Beach Road, Biddeford, ME, United States

Abstract

Background: Interest has recently increased in generative artificial intelligence (GenAI), a subset of artificial intelligence that can create new content. Although the publicly available GenAI tools are not specifically trained in the medical domain, they have demonstrated proficiency in a wide range of medical assessments. The future integration of GenAI in medicine remains unknown. However, the rapid availability of GenAI with a chat interface and the potential risks and benefits are the focus of great interest. As with any significant medical advancement or change, medical schools must adapt their curricula to equip students with the skills necessary to become successful physicians. Furthermore, medical schools must ensure that faculty members have the skills to harness these new opportunities to increase their effectiveness as educators. How medical schools currently fulfill their responsibilities is unclear. Colleges of Osteopathic Medicine (COMs) in the United States currently train a significant proportion of the total number of medical students. These COMs are in academic settings ranging from large public research universities to small private institutions. Therefore, studying COMs will offer a representative sample of the current GenAI integration in medical education.

Objective: This study aims to describe the policies and training regarding the specific aspect of GenAI in US COMs, targeting students, faculty, and administrators.

Methods: Web-based surveys were sent to deans and Student Government Association (SGA) presidents of the main campuses of fully accredited US COMs. The dean survey included questions regarding current and planned policies and training related to GenAI for students, faculty, and administrators. The SGA president survey included only those questions related to current student policies and training.

Results: Responses were received from 81% (26/32) of COMs surveyed. This included 47% (15/32) of the deans and 50% (16/32) of the SGA presidents (with 5 COMs represented by both the deans and the SGA presidents). Most COMs did not have a policy on the student use of GenAI, as reported by the dean (14/15, 93%) and the SGA president (14/16, 88%). Of the COMs with no policy, 79% (11/14) had no formal plans for policy development. Only 1 COM had training for students, which focused entirely on the ethics of using GenAI. Most COMs had no formal plans to provide mandatory (11/14, 79%) or elective (11/15, 73%) training. No COM had GenAI policies for faculty or administrators. Eighty percent had no formal plans for policy development. Furthermore, 33.3% (5/15) of COMs had faculty or administrator GenAI training. Except for examination question development, there was no training to increase faculty or administrator capabilities and efficiency or to decrease their workload.

Conclusions: The survey revealed that most COMs lack GenAI policies and training for students, faculty, and administrators. The few institutions with policies or training were extremely limited in scope. Most institutions without current training or policies had no formal plans for development. The lack of current policies and training initiatives suggests inadequate preparedness for integrating GenAI into the medical school environment, therefore, relegating the responsibility for ethical guidance and training to the individual COM member.

(JMIR Med Educ 2025;11:e58766) doi:10.2196/58766

XSL•FO
RenderX

KEYWORDS

artificial intelligence; medical education; faculty development; policy; AI; training; United States; school; university; college; institution; osteopathic; osteopathy; curriculum; student; faculty; administrator; survey; cross-sectional

Introduction

Artificial intelligence (AI) is a technology capable of performing tasks traditionally requiring human intelligence [1]. AI has a long-standing presence in medicine across clinical, educational, and administrative domains [2-4]. Generative artificial intelligence (GenAI) technologies are a subset of AI that can create new content.

In the clinical domain, GenAI has demonstrated proficiency in performing tasks ranging from passing the United States Medical Licensing Examination to providing empathetic patient communication [5,6]. At a more advanced level, these tools have answered real-world medical questions with more factual accuracy and more empathy than human physicians [7,8]. Such capabilities highlight GenAI's potential as a pivotal tool in both the learning environment of medical students and the broader context of patient care. However, the integration of GenAI into medical education raises important questions regarding the ethical, legal, and practical implications of its use.

Increased computing power, the development of a user-friendly conversational interface that lowers the technical barriers to use, and the availability to the public at little or no direct cost have made this technology nearly as available as web-based search engines or document spell-checking for medical educators and students. This has stimulated a great deal of interest by all constituencies in medicine and medical education. GenAI is only 1 component of the general field of AI. However, with the recent nearly ubiquitous availability to the general population in the United States, the yet clearly defined risks and benefits have significant implications for the short term in all aspects of medicine and the need for training and policies for medical trainees.

The rapid evolution of GenAI highlights the responsibility of medical schools to take a proactive approach to adapt their curricula and policies to harness the benefits of these technologies while mitigating potential risks. How medical schools currently fulfill their responsibilities is unclear. There are published reports highlighting individual AI-related training programs, as well as recommendations for AI curriculum, content, delivery, and challenges in medical schools [9-11]. While insightful, they do not describe the full educational landscape of US medical schools that grant either DO or MD degrees. This is particularly crucial in Colleges of Osteopathic Medicine (COMs) in the United States, which account for a significant and growing proportion of the country's medical student population. Understanding the current landscape of GenAI policies and training in COMs is essential for identifying gaps, setting benchmarks, and guiding future initiatives aimed at effectively integrating GenAI into medical education.

GenAI has rapidly become nearly ubiquitous in the United States and has the potential for significant benefits and risks. It is unclear whether COMs have included training or policy guidance in this domain. This study aimed to describe the status of policy and training, specifically in one aspect of AI, GenAI, for medical students, faculty, and administrators, as well as near-term plans for policy and training development at COMs. This analysis will provide an overview of the current state of GenAI integration in osteopathic medical education, which will demonstrate opportunities for future development.

Methods

Study Design and Population

This descriptive cross-sectional study targeted US COMs that held full accreditation by the Commission on Osteopathic College Accreditation as of the end of the 2022 - 2023 academic year. These COMs have at least 1 graduating class, ensuring that they possess a comprehensive experience with the full spectrum of undergraduate medical education. Approximately 28% of all US medical students are enrolled in COMs [12,13] in academic settings ranging from large public research universities to small private institutions. Therefore, we believe that studying COMs will offer a representative sample of the current GenAI integration in US medical education.

Ethical Considerations

Before initiating contact with potential participants, the institutional review board (number 0723-10) of the University of New England, Biddeford, Maine, granted this project an exemption status. Participation in the study was voluntary, and informed consent was provided in both the email invitation and beginning of the survey. Data collection procedures were designed for privacy and confidentiality with deidentification of respondents. There was no compensation for survey participation.

Survey Development and Data Collection

Due to the novel and rapidly developing field of GenAI, a survey was developed using an iterative process to obtain the availability, content, and development plans for training and policies for students, faculty, and administrators. The survey was designed to prioritize the general details of these domains. This strategy was to maximize the survey participation and to provide direction for potential future projects. The design was led by team members with experience in the user interface (GCL), survey development (GCL and SPT), COM medical curriculum development (GCL and SPT), and COM administrative management and operations (GCL and SPT). The survey was tested before implementation with a convenience sample of administrators and students to ensure that the questions were straightforward and the web-based survey system was usable. The order of survey items was the same for all participants in each group, with each question being presented on an individual screen. However, the surveys used an adaptive methodology to expose participants only to pertinent questions. For example, only those participants who answered that they currently provided training would be asked about the

XSL•FO
RenderX

content of the training. If a COM stated that they do not have a GenAI policy, they would be asked about future development.

Data were collected using a survey distributed via a web-based tool (Qualtrics XM). The recruitment for participation was sent by an email directly to the potential participant. The recruitment email described the project purpose and survey details, including that the survey was on the web, anonymous, and no incentives were provided for their participation. No personal data were collected, including the respondent's IP address. Two separate surveys were developed: one for the deans of the COMs and another for the presidents of the Student Government Association (SGA). The dean's survey included questions about current and planned GenAI policies and training for students, faculty, and administrators, as well as questions about the content of existing policies and training (Multimedia Appendix 1). Recognizing that students are unlikely to have knowledge of policy, curriculum planning, or those related to faculty or administrators, the SGA president's survey exclusively encompassed questions about current student policies and training (Multimedia Appendix 2). In both the dean and SGA president recruitment email, the recipient was informed that if there was a more appropriate survey responder, they may forward the email to that person, such as, the dean to an appropriate administrator, and the SGA president to a student.

Each dean and SGA president recruitment email included a unique survey URL to ensure that only 1 response represented each COM for each category. Qualtrics provides distribution data that are separate from the survey results. This allowed follow-up emails to nonresponders while maintaining the anonymity of the data. Data were collected from July 28, 2023, to September 14, 2023.

Data Analysis

Descriptive statistics were used to analyze the survey results. Response rates for both surveys were calculated as the number of completed surveys as a percentage of total COMs surveyed. The number of started but not completed surveys was calculated as a percentage of total COMs surveyed. For each COM not providing training or having a policy, the status of development was reported as the percentage of COMs surveyed without that characteristic. Due to the anonymity of the respondents and the institutional overlap of the dean and SGA presidents, no statistical comparison between the 2 groups was made.

Results

Response Rates

Of the 32 COMs surveyed, 47% (15/32) deans and 50% (16/32) SGA presidents completed the survey. Five surveys overlapped deans and SGA presidents. The dean or SGA president responded from 81% (26/32) of the COMs surveyed, providing a comprehensive understanding of the COMs. All surveys started were completed (100%).

GenAI Policies for Students

A vast majority of COMs reported a lack of established policies regarding the use of GenAI by students. Specifically, 93% (14/15) of deans and 88% (14/16) of SGA presidents indicated that their institutions had no student-focused GenAI policies. Among the few COMs with existing policies, the scope was primarily limited to GenAI use in graded assignments. Of the COMs with no policy, 79% (11/14) had no formal plans for policy development. The stages of planning for student policy are shown in Table 1.

Table . Status of student generative artificial intelligence policy and training development (Colleges of Osteopathic Medicine without policy or training).

	Student GenAI[a] policy	Student mandatory education	Student elective education
Total surveys, n	14	14	15
Status, n (%)			
Not working on a policy or education	3 (21.4)	3 (21.4)	8 (53.3)
Informal conversations	8 (57.1)	8 (57.1)	3 (20)
Workgroup in place	1 (7.1)	3 (21.4)	2 (13.3)
Being drafted and under review	2 (14.3)	0 (0)	1 (6.7)
Approved to take effect after July 1, 2023	0 (0)	0 (0)	1 (6.7)

[a]GenAI: generative artificial intelligence.

GenAI Training for Students

Only 1 COM was identified as having mandatory student training, which focused entirely on the ethics of using GenAI. None of the COMs offered any elective training. Most COMs had no formal plans to provide mandatory (11/14, 79%) or elective (11/15, 73%) training. The stages of planning for student training are shown in Table 1.

GenAI Policies for Faculty or Administrators

None of the COMs studied had a GenAI policy for faculty or administrators. Similar to the students, 80% (12/15) had no formal plans to develop one. The stages of planning for faculty or administrator policy are shown in Table 2.

XSL•FO
RenderX

Table . Status of faculty or administrator generative artificial intelligence policy and training development for Colleges of Osteopathic Medicine (COMs) with no policy or training.

		Faculty/administrator policy	Faculty/administrator training
Total surveys, n		15	10
Status, n (%)			
	We are not working on a policy or training	6 (40)	2 (20)
	Informal conversations	6 (40)	3 (30)
	Workgroup in place	2 (13.3)	2 (20)
	Being drafted and under review	1 (6.7)	3 (30)
	Approved and will take effect after July 1, 2023	0 (0)	0 (0)

GenAI Training for Faculty or Administrators

Only 33.3% (5/15) of COMs had initiated faculty or administrator-focused GenAI training. These predominantly covered basic use and ethical considerations. Except for examination question development, there was no specific focus on skills to enhance educational efficiency or reduce workload (Table 3). Fifty percent (5/10) of the COMs without faculty or administrator training had no formal plans to develop training (Table 2).

Table . Content of current faculty or administrator generative artificial intelligence training.

	Deans, n (%)
Total surveys	5 (100)
How to use the technology	4 (80)
Benefits/limitations of the technology	4 (80)
Ethics of using it	3 (60)
Legal perspective on using it	2 (40)
Development of examination questions	2 (40)

Discussion

Principal Findings

Our survey uncovers a pronounced gap in GenAI policies and training across US COMs, with the vast majority of institutions surveyed lacking formal policy guidelines (93% dean responses and 88% SGA president responses), and of the COMs with no current student policies, 79% (11/14) had no formal plans for future development. Furthermore, no COMs described any student GenAI elective training, with 73% (11/15) reporting no plans for mandatory educational programs. This underscores an urgent GenAI training imperative for medical schools to prepare future physicians for the imminent AI-enhanced health care landscape. Little has been done to support COM faculty to address these needs as no COMs surveyed had a formal policy regarding Gen AI for faculty or administration, 80% (12/15) did not have a plan to develop one, and only 33% (5/15) had focused training mainly in the realm of utilization and ethical considerations.

Comparison With Prior Work

In a recent national survey of US postsecondary schools, 8% had GenAI policies in place [14]. In that report, the focus of the policies was not described. If these were related to students, it is comparable with the data of this project, where 7% (1/15) of the deans or 12% (2/16) of the SGA presidents responded that they had student GenAI policies. In our sample of student GenAI policies, the focus was on using GenAI in graded assignments. While there were few COMs with student-focused policies, none of the COMs had faculty or administrator policies.

The survey results indicated that the status of COM AI policies is unlikely to change significantly in the near future, with few COMs having formal plans to evaluate and develop GenAI policies. The 21% (3/14) of COMs with formal plans for student policies and 20% (3/15) with plans for faculty or administrator policies demonstrate that they are far less engaged than the postsecondary programs, in which 57% are evaluating and developing policies [14].

As with policy, training for COM students, faculty, and administrators is minimal and does not focus on enabling students, faculty, or administrators to increase productivity, improve effectiveness, or decrease workload. Because the majority do not have formal plans to develop training, this situation is unlikely to change significantly in the near future.

Implications for Future Practice

The rapid advancement of AI technologies, including GenAI, necessitates a proactive stance from medical education institutions to integrate these tools effectively and ethically into teaching, learning, and clinical practice. COMs must move more quickly to develop AI policies and training. However, we do not propose indiscriminately replicating the nascent policies or

XSL•FO
RenderX

training approaches of other institutions, which may not be appropriate for their institution. Furthermore, we caution against a hasty and thoughtless development process merely for the sake of establishing provisional measures. Instead, we propose that medical educators and administrators use the growing body of resources to strategically and methodically create policies and training resources using interdisciplinary teams and continually improve them as future GenAI innovations progressively transform the paradigm of technology-assisted human labor.

One example of resources to be reviewed is the study by Chan [15] that presented an AI policy framework integrating their local data and the UNESCO (United Nations Educational, Scientific and Cultural Organization) AI policy guidance [16]. This policy framework is divided into 3 dimensions, governance, operational, and pedagogical, and can also be used as a competency framework, as shown in Table 4.

Table . Artificial intelligence (AI) education policy framework [15].

Domain	Explanation	Content	Leadership
Pedagogical	Teaching and learning aspects of AI integration.	• Rethinking assessments and examinations. Developing student holistic competencies/generic skills • Preparing students for the AI-driven workplace • Encouraging a balanced approach to AI adoption	Teachers
Operational	Practical implementation of AI in university settings	• Monitoring and evaluating AI implementation • Providing training and support for teachers, staff, and students in AI literacy	Teaching and learning and IT staff
Governance	Governance considerations surrounding AI usage in education	• Understanding, identifying, and preventing academic misconduct and ethical dilemmas • Addressing governance of AI: data privacy, transparency, accountability, and security • Attributing AI technologies • Ensuring equity in access to AI	Senior management

Further frameworks for describing AI literacy and learner competencies have emerged [9,10,17-20] and can form a starting point for COMs when developing a curriculum consistent with their institution's educational mission and existing pedagogical architecture. Building upon this framework, in addition to work done internally, the growing body of published content resources can be accessed and, where appropriate, integrated into their development process. Some resources may be adapted from general educational domains, including skills such as writing [21] or faculty development of course content [22]. Other resources are specific to clinical care [20,23], education [24], or ethical use [25,26]. By adopting and evolving these frameworks with growing evidence-based resources, medical schools can ensure that their curricula not only cover the operational aspects of GenAI but also address the ethical, social, and professional implications.

This general framework is appropriate for learners at any developmental stage. However, as in other areas of medical education, the learners' level of training [11,27] must be considered. For faculty or administrators, responsibilities in developing, integrating, and operationalizing the curriculum must also be considered [28].

In addition to the trainee level, medical school policy makers and educators must consider the systems in which future physicians will work. Physicians should be part of a team with diverse backgrounds and professional training to be most effective. With further AI development, these teams will include AI-powered computer assistants. The team must know how to interact effectively and appropriately with this new "team member," including how it affects the patients and families they care for. This awareness is similar to the early assessments of the effects of electronic health records during clinical encounters [29,30].

Implementing GenAI competencies or any new content is a challenge with an already crowded curriculum. We propose that GenAI be integrated into the current system, where other tools are used to minimize the negative effect. When trainees learn to search and evaluate background scientific publications, GenAI can be incorporated where appropriate as one of the tools they are trained with. Furthermore, when practicing for clinical encounters, whether an actual clinical encounter or their objective structured clinical exams, using GenAI as a tutor may potentially reinforce their preparation. There are many similar uses that will integrate GenAI as a tool and not necessitate a significant increase in curriculum time and may additionally make other aspects of their curriculum more effective. However, these efforts will need further evaluation.

XSL•FO
RenderX

By developing clear policies and offering robust training, medical schools can ensure that future physicians are adept at leveraging GenAI to improve health care outcomes while navigating the ethical and professional complexities it presents.

Limitations

This study's findings must be interpreted in light of several limitations. The availability of data limits this project. Ongoing assessment is needed that includes a larger group of medical schools, including those that grant either doctor of osteopathic medicine or doctor of medicine degrees. In addition, other aspects of the physician's life cycle (graduate medical education, clinical practice, and continuing education) must be studied.

The rapidly evolving nature of GenAI requires institutional policies and training initiatives that can quickly adapt, necessitating ongoing research to capture these developments accurately.

Conclusions and Future Directions

Most COMs do not provide AI policy guidance or training for medical students, faculty, or administrators. There also does not seem to be an appropriate prioritization by COMs to remedy this deficiency. While many philosophers, including the great baseball legend Yogi Berra, have opined that "It is difficult to make predictions, especially about the future" [31], this difficulty does not negate medical schools' responsibility while waiting for the future to become clear. They must assess future physicians' needs and implement appropriate training and guidance in their programs. If the COMs do not lead, their trainees will be unprepared for the future. This risks inappropriate use of AI and the medical equivalent to the lawyer who used GenAI to submit a brief in court that included fabricated references or "hallucinations" [32].

Future research should explore effective strategies for implementing GenAI education and policy development, including interdisciplinary approaches and stakeholder engagement.

Acknowledgments

The authors wish to thank Christopher Callaway, PhD, for his assistance with survey design and data collection.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COM (College of Osteopathic Medicine) dean survey.
[DOCX File, 23 KB - mededu_v11i1e58766_app1.docx]

Multimedia Appendix 2
SGA (Student Government Association) president survey.
[DOCX File, 19 KB - mededu_v11i1e58766_app2.docx]

References

1. McCarthy J, Hayes P. Some philosophical problems from the standpoint of artificial intelligence. In: Meltzer B, Michie D, editors. Machine Intelligence 4: Edinburgh University Press.
2. Vyborny CJ, Giger ML. Computer vision and artificial intelligence in mammography. AJR Am J Roentgenol 1994 Mar;162(3):699-708. [doi: 10.2214/ajr.162.3.8109525] [Medline: 8109525]
3. Keel S, Lee PY, Scheetz J, et al. Feasibility and patient acceptability of a novel artificial intelligence-based screening model for diabetic retinopathy at endocrinology outpatient services: a pilot study. Sci Rep 2018 Mar 12;8(1):4330. [doi: 10.1038/s41598-018-22612-2] [Medline: 29531299]
4. Takiddin A, Schneider J, Yang Y, Abd-Alrazaq A, Househ M. Artificial intelligence for skin cancer detection: scoping review. J Med Internet Res 2021 Nov 24;23(11):e22934. [doi: 10.2196/22934] [Medline: 34821566]
5. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023 Feb;2(2):e0000198. [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
6. Brin D, Sorin V, Vaid A, Soroush A, Glicksberg BS, Charney AW, et al. Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. Sci Rep 2023 Oct 1;13(1):16492. [doi: 10.1038/s41598-023-43436-9] [Medline: 37779171]
7. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023 Jun 1;183(6):589-596. [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
8. Goodman RS, Patrinely JR, Stone CA Jr, Zimmerman E, Donald RR, Chang SS, et al. Accuracy and reliability of chatbot responses to physician questions. JAMA Netw Open 2023 Oct 2;6(10):e2336483. [doi: 10.1001/jamanetworkopen.2023.36483]

XSL•FO
RenderX

9. Lee J, Wu AS, Li D, Kulasegaram KM. Artificial intelligence in undergraduate medical education: a scoping review. Acad Med 2021 Nov 1;96(11S):S62-S70. [doi: 10.1097/ACM.0000000000004291] [Medline: 34348374]
10. Charow R, Jeyakumar T, Younus S, et al. Artificial intelligence education programs for health care professionals: scoping review. JMIR Med Educ 2021 Dec 13;7(4):e31043. [doi: 10.2196/31043] [Medline: 34898458]
11. Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P. Introducing artificial intelligence training in medical education. JMIR Med Educ 2019 Dec 3;5(2):e16048. [doi: 10.2196/16048] [Medline: 31793895]
12. American Association of Colleges of Osteopathic Medicine (AACOM). Osteopathic medical college total enrollment by race/ethnicity 2000-2023. 2023 Oct 30. URL: https://www.aacom.org/docs/default-source/research-reports/2000-23-tebycom-re.xlsx?sfvrsn=7ee8361a_24 [accessed 2025-01-30]
13. Association of American Medical Colleges (AAMC). Applicants, matriculants, enrollment, and graduates of US MD-granting medical schools, 2013-2014 through 2022-202. 2013. URL: https://www.aamc.org/media/37816/download?attachment [accessed 2025-01-30]
14. Sebesta J, Davis VL. Supporting instruction and learning through artificial Intelligence: a survey of institutional practices & policies.: WICHE Cooperative for Educational Technologies; 2023. URL: https://wcet.wiche.edu/wp-content/uploads/sites/11/2023/07/AI-Survey-In-Depth-Analysis-Report-Summer-2023.pdf [accessed 2023-09-23]
15. Chan CKY. A comprehensive AI policy education framework for university teaching and learning. Int J Educ Technol High Educ 2023;20(1):1-25. [doi: 10.1186/s41239-023-00408-3]
16. Miao F, Holmes W. Artificial intelligence and education guidance for policymakers United Nations Educational, Scientific and Cultural Organization (UNESCO): Paris, France. 2021. URL: https://unesdoc.unesco.org/ark:/48223/pf0000376709 [accessed 2023-10-17]
17. Kong SC, Man-Yin Cheung W, Zhang G. Evaluation of an artificial intelligence literacy course for university students with diverse study backgrounds. Comput Educ Artif Intell 2021;2:100026. [doi: 10.1016/j.caeai.2021.100026]
18. Ng DTK, Leung JKL, Chu SKW, Qiao MS. Conceptualizing AI literacy: an exploratory review. Comput Educ Artif Intell 2021;2:100041. [doi: 10.1016/j.caeai.2021.100041]
19. McCoy LG, Nagaraj S, Morgado F, Harish V, Das S, Celi LA. What do medical students actually need to know about artificial intelligence? NPJ Digit Med 2020;3:86. [doi: 10.1038/s41746-020-0294-7] [Medline: 32577533]
20. Russell RG, Lovett Novak L, Patel M, et al. Competencies for the use of artificial intelligence-based tools by health care professionals. Acad Med 2023 Mar 1;98(3):348-356. [doi: 10.1097/ACM.0000000000004963] [Medline: 36731054]
21. Byrd A, Flores L, Green D, et al. MLA-CCCC joint task force on writing and AI working paper: overview of the issues, statement of principles, and recommendations. 2023 Jul. URL: https://hcommons.org/app/uploads/sites/1003160/2023/07/MLA-CCCC-Joint-Task-Force-on-Writing-and-AI-Working-Paper-1.pdf [accessed 2023-09-07]
22. Dickey E, Bejarano A. A model for integrating generative AI into course content development. . 2023 23. [doi: 10.48550/arXiv.2308.12276]
23. Ethical application of artificial intelligence in family medicine. American Academy of Family Physicians (AAFP). URL: https://www.aafp.org/about/policies/all/ethical-ai.html [accessed 2023-09-06]
24. Global Forum on Innovation in Health Professional Education, Board on Global Health, Health and Medicine Division, National Academies of Sciences, Engineering, and Medicine. Artificial intelligence in health professions education. In: Proceedings of a Workshop: National Academies Press; 2023:38166138 URL: https://www.nap.edu/catalog/27174 [doi: 10.17226/27174]
25. Nguyen A, Ngo HN, Hong Y, Dang B, Nguyen BPT. Ethical principles for artificial intelligence in education. Educ Inf Technol (Dordr) 2023;28(4):4221-4241. [doi: 10.1007/s10639-022-11316-w] [Medline: 36254344]
26. Weidener L, Fischer M. Teaching AI ethics in medical education: a scoping review of current literature and practices. Perspect Med Educ 2023;12(1):399-410. [doi: 10.5334/pme.954] [Medline: 37868075]
27. Sapci AH, Sapci HA. Artificial intelligence education and tools for medical and health informatics students: systematic review. JMIR Med Educ 2020 Jun 30;6(1):e19285. [doi: 10.2196/19285] [Medline: 32602844]
28. Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. Acad Med 2024 Jan 1;99(1):22-27. [doi: 10.1097/ACM.0000000000005439] [Medline: 37651677]
29. Margalit RS, Roter D, Dunevant MA, Larson S, Reis S. Electronic medical record use and physician-patient communication: an observational study of Israeli primary care encounters. Patient Educ Couns 2006 Apr;61(1):134-141. [doi: 10.1016/j.pec.2005.03.004] [Medline: 16533682]
30. Ventres W, Kooienga S, Vuckovic N, Marlin R, Nygren P, Stewart V. Physicians, patients, and the electronic health record: an ethnographic analysis. Ann Fam Med 2006;4(2):124-131. [doi: 10.1370/afm.425] [Medline: 16569715]
31. Dickstein DP. Editorial: it's difficult to make predictions, especially about the future: risk calculators come of age in child psychiatry. J Am Acad Child Adolesc Psychiatry 2021 Aug;60(8):950-951. [doi: 10.1016/j.jaac.2020.12.029] [Medline: 33383160]
32. New York Times. The ChatGPT lawyer explains himself. 2023. URL: https://www.nytimes.com/2023/06/08/nyregion/lawyer-chatgpt-sanctions.html [accessed 2023-09-06]

XSL•FO
RenderX

Abbreviations

AI: artificial intelligence
COMs: Colleges of Osteopathic Medicine
GenAI: generative artificial intelligence
SGA: Student Government Association
UNESCO: United Nations Educational, Scientific and Cultural Organization

Edited by B Lesselroth; submitted 24.03.24; peer-reviewed by C Zanetti, M Lin; revised version received 09.10.24; accepted 02.01.25; published 11.02.25.

Please cite as:
Ichikawa T, Olsen E, Vinod A, Glenn N, Hanna K, Lund GC, Pierce-Talsma S
Generative Artificial Intelligence in Medical Education—Policies and Training at US Osteopathic Medical Schools: Descriptive Cross-Sectional Survey
JMIR Med Educ 2025;11:e58766
URL: https://mededu.jmir.org/2025/1/e58766
doi:10.2196/58766

© Tsunagu Ichikawa, Elizabeth Olsen, Arathi Vinod, Noah Glenn, Karim Hanna, Gregg C Lund, Stacey Pierce-Talsma. Originally published in JMIR Medical Education (https://mededu.jmir.org), 11.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Assessing Familiarity, Usage Patterns, and Attitudes of Medical Students Toward ChatGPT and Other Chat-Based AI Apps in Medical Education: Cross-Sectional Questionnaire Study

Safia Elwaleed Elhassan, MBBS; Muhammad Raihan Sajid, MBBS, MMed; Amina Mariam Syed, MBBS; Sidrah Afreen Fathima, MBBS; Bushra Shehroz Khan, MBBS; Hala Tamim, PhD

College of Medicine, Alfaisal University, Takhasussi street, Riyadh, Saudi Arabia

Corresponding Author:
Muhammad Raihan Sajid, MBBS, MMed
College of Medicine, Alfaisal University, Takhasussi street, Riyadh, Saudi Arabia

Abstract

Background: There has been a rise in the popularity of ChatGPT and other chat-based artificial intelligence (AI) apps in medical education. Despite data being available from other parts of the world, there is a significant lack of information on this topic in medical education and research, particularly in Saudi Arabia.

Objective: The primary objective of the study was to examine the familiarity, usage patterns, and attitudes of Alfaisal University medical students toward ChatGPT and other chat-based AI apps in medical education.

Methods: This was a cross-sectional study conducted from October 8, 2023, through November 22, 2023. A questionnaire was distributed through social media channels to medical students at Alfaisal University who were 18 years or older. Current Alfaisal University medical students in years 1 through 6, of both genders, were exclusively targeted by the questionnaire. The study was approved by Alfaisal University Institutional Review Board. A χ^2 test was conducted to assess the relationships between gender, year of study, familiarity, and reasons for usage.

Results: A total of 293 responses were received, of which 95 (32.4%) were from men and 198 (67.6%) were from women. There were 236 (80.5%) responses from preclinical students and 57 (19.5%) from clinical students, respectively. Overall, males (n=93, 97.9%) showed more familiarity with ChatGPT compared to females (n=180, 90.09%; P=.03). Additionally, males also used Google Bard and Microsoft Bing ChatGPT more than females (P<.001). Clinical-year students used ChatGPT significantly more for general writing purposes compared to preclinical students (P=.005). Additionally, 136 (46.4%) students believed that using ChatGPT and other chat-based AI apps for coursework was ethical, 86 (29.4%) were neutral, and 71 (24.2%) considered it unethical (all Ps>.05).

Conclusions: Familiarity with and usage of ChatGPT and other chat-based AI apps were common among the students of Alfaisal University. The usage patterns of these apps differ between males and females and between preclinical and clinical-year students.

(JMIR Med Educ 2025;11:e63065) doi:10.2196/63065

KEYWORDS

ChatGPT; artificial intelligence; large language model; medical students; ethics; chat-based; AI apps; medical education; social media; attitude; AI

Introduction

ChatGPT is a sophisticated large language model of artificial intelligence (AI) that was created by OpenAI and released to the public in November 2022 [1]. It generates human-like responses to natural language inputs. The users can hold a conversation with the model where they input a prompt and receive a response [2]. It has many applications including email writing, solving math problems, grammar checking, generating answers to complex questions, and more [3]. Other similar chat-based AI apps include Google Bard, Microsoft Bing ChatGPT, Socrative by Google, Hugging Chat, Snapchat AI, Perplexity AI, and YouChat, among others. All these apps are similar to ChatGPT in terms of generating natural responses to prompts [4].

There has been a rise in new literature pertaining to the use of ChatGPT and other AI tools among medical students. Many published articles show that medical students have a positive attitude toward using ChatGPT in education [5-8]. Many students are eager to use AI tools as they believe it can revolutionize medicine and dentistry [9,10]. Additionally, ChatGPT and other chat-based AIs are continuing to evolve to expand their scope of usage, for example, making virtual histology slides for interactive learning [11-13]. Moreover,

XSL•FO
RenderX

ChatGPT can be used in medical education and medical specialties [14]. Its implications in the cardiovascular, cerebrovascular, and radiology fields are being extensively studied, as it can interpret medical imaging and potentially provide a diagnosis [15-17].

Regarding medical research, ChatGPT and similar AI apps can expedite the writing processes by enabling authors to allocate their time and resources more efficiently, by reducing the time spent on the laborious process of searching for relevant literature [18].

A few studies have been conducted to determine students' willingness to integrate AI tools such as ChatGPT into education. One study demonstrated that both undergraduate and postgraduate students in Hong Kong had a positive attitude toward integrating AI tools into higher education due to its ability to provide immediate solutions, help generate ideas, and handle tedious tasks, allowing students to focus on more important work [6]. Similarly, another study performed on students and faculty at Texas University showed a favorable perception of ChatGPT usage. The responses highlighted the benefits of having access to an AI instructor, which can assist in simplifying concepts by providing examples, offering study advice, and working with students on individual projects [5]. However, the studies were relatively recent and recommend further research, targeting different majors to understand the specialized use of AI in different fields.

Within the Middle East, limited recent studies have assessed medical students' knowledge and attitudes toward AI. A recent study assessed the awareness, perceptions, and opinions of pharmacy undergraduate students toward AI at King Saud University in Riyadh. The findings indicated a generally positive attitude, with demographic factors such as gender and year of study influencing their perceptions [19]. Another qualitative study investigated the knowledge, benefits, concerns, and limitations associated with the use of ChatGPT among medical college faculty and students in Saudi Arabia; the results highlighted both positive aspects such as enhanced communication and learning, and concerns regarding reliability and privacy [20]. Another study conducted at the University of Jordan involving 623 randomly selected medical students demonstrated a strong positive inclination toward using ChatGPT for learning. The findings recommended integrating ChatGPT into the university curricula, emphasizing benefits for students and the potential for misuse [21].

Due to the rise in popularity of ChatGPT and other chat-based AI in medical education, further research must be conducted to understand students' familiarity, usage habits, and attitudes toward these technologies. Despite data from other parts of the world and colleges, there is a significant lack of information on this topic in medical education and research, especially in Saudi Arabia. Therefore, this study was designed to study the familiarity, usage, and attitudes of medical students at Alfaisal University toward ChatGPT and other chat-based AI apps for medical education and research. Furthermore, it explores the perceived limitations, advantages, and ethical concerns that arise from their use. This paper addresses the research question "What are the familiarity, usage patterns, and attitudes of Alfaisal University medical students toward ChatGPT and other chat-based AI apps in medical education?" Based on existing literature, we hypothesize that Alfaisal University medical students are familiar with chat-based AI apps and hold positive attitudes toward their use.

Methods

Study Design and Enrollment

This study was a closed cross-sectional survey that was conducted among medical students at Alfaisal University. Alfaisal University is a private university in Riyadh, Saudi Arabia, that has around 1500 enrolled medical students.

Only current Alfaisal University medical students in years 1 through 6, of both genders, aged 18 years and above were targeted by the questionnaire; students who did not meet the eligibility criteria were not included in the study. The target sample size was calculated to be in the range of 290 - 310 students to achieve a 95% confidence level with a 5% CI, using a sample size calculator.

The online questionnaire was made using Google Forms, a web-based tool used to distribute surveys. The survey was open for responses over 6 weeks, from October 8, 2023, to November 22, 2023. The current survey was modified based on earlier published research [5,6]; the published surveys were chosen in accordance with the IDEE (Identify, Discern, Ethics, Engage) framework, which evaluates how students utilize chat-based AI to achieve specific educational goals, assesses the perceived level of AI integration, examines the effectiveness of AI tools, and explores the ethical considerations involved. The survey was revised to align with our requirements and complement the goals of the study, as previous articles targeted different populations. The answer choices were adapted to reflect the context specific to medical students. The survey was sent to students via email and through Whatsapp groups and other social media outlets including Instagram and Twitter. The survey was designed in accordance with the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [22,23].

The questionnaire (Multimedia Appendix 1) consisted of 21 questions distributed over 5 pages to assess familiarity, usage, and attitude of medical students toward ChatGPT and other AI apps. The survey consisted of four sections. The first section addressed demographic aspects, including gender (males and females) and academic year (preclinical: years 1 - 3; clinical: years 4 - 6). The second section had questions regarding the knowledge and use of ChatGPT and other chat-based AI apps, including familiarity, frequency of use, and purposes of usage. Participants were asked to rank their familiarity and frequency through Likert-scale questions. For the purpose of usage, the questions were divided according to uses in medical education and medical research. Participants were asked to select all the relevant choices. In the third section, participants were asked to rate their attitudes toward using ChatGPT or other chat-based AI apps in medical training using Likert-scale questions. They rated their beliefs about the enhancement of medical education through such tools, and their intentions to incorporate them into their future learning practices. The final section investigated

XSL•FO
RenderX

ethical considerations of students, including concerns about academic dishonesty.

Descriptive statistics were performed to describe the level of familiarity, reasons for usage, and attitudes of students toward ChatGPT and other chat-based AI apps. A χ^2 test was conducted to assess the relationships between gender and year of studies in terms of familiarity and reasons for usage of ChatGPT and other AI apps. Data analysis was carried out using SPSS (version 29.0; IBM Corp). Statistical significance was set at $P<.05$.

Ethical Considerations

This study received ethical approval from the Institutional Review Board at Alfaisal University (approval number: IRB-20247). The participants were informed of the purpose of the study; the survey was 4 - 5 minutes long and the principal investigator's email was provided for inquiries. The students provided written informed consent to participate in the research. Participation was voluntary and the students were not given any compensation. To maintain confidentiality, no personally identifiable information such as names or college identity numbers were gathered. The responses were only available to the primary investigators and coinvestigators, and data were anonymized.

Results

In total, 293 responses fit the inclusion criteria, 95 (32.4 %) of which were from men and 198 (67.6%) from women. There were 236 (80.5%) responses from preclinical and 57 (19.5%) from clinical students, respectively. Participant familiarity with various AI apps is summarized in Table 1. Most students were familiar with ChatGPT and other chat-based AI apps. However, men used ChatGPT, Google Bard, and Microsoft Bing ChatGPT significantly more than women (all *P*s<.05). Additionally, Socrative by Google was used more by students in the preclinical years when compared to students in clinical years (*P*=0.11) (Table 1).

Table . Familiarity of students toward various artificial intelligence (AI) apps.

AI apps	Total number of responses (N=293), n (%)	Gender			Academic year		
		Men (N=95), n (%)	Women (N=198), n (%)	*P* value	Preclinical (N=236), n (%)	Clinical (N=57), n (%)	*P* value
ChatGPT	273 (93)	93 (97.9)	180 (90.9)	.03	218 (92.4)	55 (96.5)	.27
Google Bard	63 (21.5)	32 (33.7)	31 (15.7)	<.001	50 (21.2)	13 (22.8)	.79
Microsoft Bing ChatGPT	89 (30.4)	41 (43.2)	48 (24.2)	<.001	73 (30.9)	16 (28.1)	.67
Socrative by Google	41 (14)	13 (13.7)	28 (14.1)	.92	39 (16.5)	2 (3.5)	.01
Snapchat AI	181 (61.8)	59 (62.8)	122 (61.6)	.85	145 (61.7)	36 (63.2)	.84
Perplexity AI	7 (2.4)	3 (3.2%)	4 (2)	.55	5 (2.1)	2 (3.5)	.54
YouChat	9 (3.1)	1 (1.1)	8 (4)	.17	9 (3.8)	0 (0.0)	.13
Poe-Telegram-Chatsonic-Replika-Huggingchat	14 (4.8)	7 (7.4)	7 (3.5)	.15	10 (4.2)	4 (7)	.38

Reasons for using various AI apps are summarized in Table 2. Men used ChatGPT for technical questions and solving practice questions significantly more than women (both *P*s<.05). Additionally, clinical students used ChatGPT significantly more for general writing compared to preclinical students (*P*=.005) (Table 2).

XSL•FO
RenderX

Table . Reasons for using various AI apps.

	Total (N=293), n (%)	Gender			Academic years		
		Men (N=95), n (%)	Women (N=198), n (%)	*P* value	Preclinical (N=236), n (%)	Clinical (N=57), n (%)	*P* value
Usage of Chat-GPT/other chat-based apps for medical education							
Asking technical questions	104 (35.5)	44 (46.3)	60 (30.3)	.007	82 (34.7)	22 (38.6)	.59
Asking general knowledge questions/advice on medical issues	113 (38.6)	39 (41.1)	74 (37.4)	.55	86 (36.4)	27 (47.4)	.13
Solving practice questions	84 (28.7)	34 (35.8)	50 (25.3)	.06	73 (30.9)	11 (19.3)	.08
Generating flashcards	34 (11.6)	13 (13.7)	21 (10.6)	.44	29 (12.3)	5 (8.8)	.46
Asking quick questions when stuck on a problem	115 (39.2)	39 (41.1)	76 (38.4)	.66	95 (40.3)	20 (35.1)	.47
Explaining concepts	101 (34.5)	39 (41.1)	62 (31.3)	.10	78 (33.1)	23 (40.4)	.30
Summarizing text	110 (37.5)	32 (33.7)	78 (39.8)	.31	87 (37)	23 (41.1)	.57
Usage of Chat-GPT/other chat-based apps for medical research							
Helping with assignments, making notes, drafting emails	9 (3.1)	3 (3.2)	6 (3)	.95	8 (3.4)	1 (1.8)	.52
General writing	4 (1.4)	2 (2.1)	2 (1)	.45	1 (0.4)	3 (5.3)	.005
Summarizing texts	97 (33.1)	35 (37.2)	62 (31.3)	.32	79 (33.6)	18 (31.6)	.77
Proofreading	59 (20.1)	23 (24.2)	36 (18.2)	.23	48 (20.3)	11 (19.3)	.86
Grammar checking	82 (28)	29 (30.5)	53 (26.8)	.50	67 (28.4)	15 (26.3)	.75
Paraphrasing	121 (41.3)	41 (43.2)	80 (40.4)	.65	91 (38.6)	30 (52.6)	.053
Writing sections of research	46 (15.7)	16 (17)	30 (15.2)	.68	31 (13.2)	15 (26.3)	.02
Generating citations	39 (13.3)	10 (10.5)	29 (14.6)	.33	35 (14.8)	4 (7)	.12
Searching for relevant articles	63 (21.5)	20 (21.1)	43 (21.7)	.90	45 (19.1)	18 (31.6)	.04
Analyzing literature	39 (13.3)	17 (17.9)	22 (11.1)	.11	33 (14)	6 (10.5)	.49

Attitudes and ethical knowledge toward AI apps are reported in Table 3. Notably, the findings showed that 136 (46.4%) of the participants believed using ChatGPT and other chat-based AI apps for coursework was ethical, 86 (29.4%) were neutral, and 71 (24.2%) considered it unethical (all *P*s>.05) (Table 3).

XSL•FO
RenderX

Table . Attitude and ethical knowledge toward AI apps.

Aspect	Agree/ethical, n (%)	Neutral, n (%)	Disagree/nonethical, n (%)
ChatGPT/other chat-based AI apps can enhance my medical education	171 (58.4)	96 (32.8)	26 (8.9)
In the future, I plan to incorporate ChatGPT/other chat-based AI apps into my learning procedures	149 (50.9)	96 (32.8)	48 (16.4)
ChatGPT/other chat-based AI apps can help me save time in medical research	188 (64.2)	79 (27)	26 (8.9)
ChatGPT/other chat-based AI apps can provide me with unique perspectives that I may not have thought of myself	193 (65.9)	80 (27.3)	20 (6.8)
ChatGPT/other chat-based AI apps can provide me with personalized and immediate feedback for my assignments	180 (61.4)	81 (27.6)	32 (10.9)
I can become overreliant on ChatGPT/other chat-based AI apps	104 (35.5)	78 (26.6)	111 (37.9)
ChatGPT/other chat-based AI apps will enable academic dishonest behaviors	226 (77.1)	54 (18.4)	13 (4.4)
I understand ChatGPT/other chat-based AI apps can generate output that is factually inaccurate	206 (70.3)	64 (21.8)	23 (7.8)
To what extent do you think using ChatGPT/other chat-based AI apps is ethical for coursework?	136 (46.4)	86 (29.4)	71 (24.2)

Discussion

Principal Findings

This study investigated the familiarity, usage patterns, and attitudes toward chat-based AI apps among medical students at Alfaisal University, Riyadh, Saudi Arabia. The findings reveal interesting insights into how this technology is integrated into medical education and research.

When evaluating familiarity, it was found that a significant majority of students (>90%) were familiar with ChatGPT, the most popular application. Additionally, male students exhibited a statistically greater familiarity with, and use of certain apps compared to female students. Furthermore, preclinical students were more familiar with Socrative by Google than other AI apps.

For usage, the primary reasons for using chat-based AI were related to medical education, including asking questions, solving practice problems, generating flash cards, and summarizing texts. Nearly 40% of the students reported using AI to ask quick questions when stuck on a problem and explain concepts. While less prevalent, AI was also used for tasks such as summarizing research texts, proofreading, and paraphrasing.

When questioned about attitudes, most students agreed that chat-based AI could enhance learning, save time, and provide unique perspectives. A vast majority of medical students were willing to incorporate ChatGPT and similar AI apps in their learning strategies and believed that it enabled them to save time. It also provided them with unique perspectives and personalized and immediate feedback on their assignments. Despite the positive outlook, a significant portion of students (37.9%) expressed concerns about overreliance on AI; they also had varying opinions regarding the ethical use of AI for coursework. Despite the positive views on chat-based AI for learning, a significant concern emerged among students. Nearly 77% students feared that these AI apps could contribute to academic dishonesty.

Implications of Findings

The findings have significant implications for medical education. The high awareness of chat-based AI, particularly among male students suggests that integrating technology into early medical education could enhance learning outcomes. The varied app usage between preclinical and clinical students highlights the importance of tailored educational tools at different training stages. Furthermore, students' comfort in using AI for daily problem-solving underscores its potential to streamline research workflows and enhance study efficiency, emphasizing the importance of incorporating AI literacy and ethical considerations into curricula.

The findings of this study reinforce the idea that the conventional memory-based medical curriculum, which is primarily memory based, must be followed by advancements in AI. This model has been effective for centuries but demonstrates limitations in the context of the AI age, where

XSL•FO
RenderX

technology is evolving to assist with information retrieval, data processing, and clinical decision-making. While memory and foundational knowledge remain important, there is an increasing need for critical thinking, problem-solving, and technological literacy.

Competence in the efficient integration and using knowledge from an expanding range of sources, including the ethical use of AI must be taught to aspiring doctors[16,24]. These findings, unique to this study, reinforce the importance of using AI in medical education [9,16,21,25]. However, students also acknowledged the potential for misuse, highlighting the importance of clear guidelines and fostering a culture of academic integrity alongside the integration of AI in medical education.

Comparison of Literature

Regarding awareness, our findings are similar to a previously published study from Saudi Arabia that assessed the awareness, perceptions, and opinions toward AI among pharmacy undergraduates. Several students had a positive awareness toward AI and its implications in health care [19].

A cross-sectional study on medical and dental students' perceptions of AI noted a lack of basic AI education in medical and dental schools. Furthermore, raised concerns about AI-competent doctors may replace doctors those less knowledgeable in using AI. This suggests that educational resources are crucial during earlier stages of medical training to keep up with advancements in AI [9]. Additionally, another study showed that pharmacy students deemed it essential to incorporate AI into college curriculum to effectively educate students on apps in the health care field [19].

A study in Canada reported similar results in terms of attitudes toward AI in research. Conducted on Canadian entry-to–health care students, it found that students who were interested in research generally had a more favorable outlook toward AI [26]. This suggests a potential role of AI in enhancing research efficiency.

However, concerns about the overreliance on AI were similarly found in other studies. For instance, a cross-sectional study conducted among pharmacy students in Saudi Arabia found that 46% of students believed that the use of AI reduced the humanistic aspect of health care, while 7.6% believed that AI devalued the medical profession [12]. Similarly, another study conducted on Canadian health care students expressed those concerns that AI could eventually take over their jobs [26].

There were also varying opinions about the ethical use of AI for coursework. A study at the University of Jordan encouraged educators to integrate ChatGPT into medical curricula and teaching practices, while also addressing student concerns and the potential for misuse [21]. Similarly, a cross-sectional study by Weidner and Fischer [27] in German-speaking European countries highlighted the necessity of incorporating teaching AI ethics into the undergraduate medical curricula. This highlights the need for discussion around responsible AI integration in medical education [9,16,25,28].

A previously published study emphasized the potential for misuse, raising concerns that students might rely on ChatGPT to outsource their assessment tasks [29]. Additionally, in a qualitative study conducted at the Faculty of Medicine, Jazan University in Saudi Arabia, respondents expressed ethical concerns related to threats to academic integrity, plagiarism, privacy, and confidentiality issues [7]. Our findings are similar to these studies, highlighting the importance of clear guidelines and fostering a culture of academic integrity [8,30].

Limitations of the Study

This study focuses on self-reported data, which may not always reflect actual practices and can cause information bias. There may be a chance of selection bias due to convenient sampling. Additionally, the results may not be generalizable to other countries, as cultural differences could lead to varying attitudes and responses in different contexts.

Conclusion

Overall, this study provides valuable insights into the growing integration of chat-based AI apps within medical education. As technology evolves, it will be crucial to address ethical concerns and ensure responsible use while maximizing the potential benefits for student learning and research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey.
[DOCX File, 22 KB - mededu_v11i1e63065_app1.docx]

References

1. Marr B. A short history of ChatGPT: how we got to where we are today. Forbes. URL: https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/ [accessed 2024-01-30]
2. How does ChatGPT work? Zapier. URL: https://zapier.com/blog/how-does-chatgpt-work/ [accessed 2024-01-30]
3. 50 ChatGPT use cases with real-life examples in 2024. AI Multiple Research. URL: https://research.aimultiple.com/chatgpt-use-cases/ [accessed 2024-04-29]
4. The best AI productivity tools in 2024. Zapier. URL: https://zapier.com/blog/best-ai-productivity-tools/ [accessed 2024-04-29]

XSL•FO
RenderX

5. Amani S, White L, Balart T, et al. Generative AI perceptions: a survey to measure the perceptions of faculty, staff, and students on generative AI tools in academia. arXiv. Preprint posted online on Apr 21, 2023. [doi: 10.48550/arXiv.2304.14415]
6. Chan CKY, Hu W. Students' voices on generative AI: perceptions, benefits, and challenges in higher education. Int J Educ Technol High Educ 2023;20(1). [doi: 10.1186/s41239-023-00411-8]
7. Salih SM. Perceptions of faculty and students about use of artificial intelligence in medical education: a qualitative study. Cureus. URL: https://www.cureus.com/articles/237686-perceptions-of-faculty-and-students-about-use-of-artificial-intelligence-in-medical-education-a-qualitative-study [accessed 2024-10-21]
8. Magalhães Araujo S, Cruz-Correia R. Incorporating ChatGPT in medical informatics education: mixed methods study on student perceptions and experiential integration proposals. JMIR Med Educ 2024 Mar 20;10:e51151 [FREE Full text] [doi: 10.2196/51151] [Medline: 38506920]
9. Bisdas S, Topriceanu CC, Zakrzewska Z, et al. Artificial intelligence in medicine: a multinational multi-center survey on the medical and dental students' perception. Front Public Health 2021;9:795284. [doi: 10.3389/fpubh.2021.795284] [Medline: 35004598]
10. Sit C, Srinivasan R, Amlani A, et al. Attitudes and perceptions of UK medical students towards artificial intelligence and radiology: a multicentre survey. Insights Imaging 2020 Feb 5;11(1):14. [doi: 10.1186/s13244-019-0830-7] [Medline: 32025951]
11. Lee H. The rise of ChatGPT: exploring its potential in medical education. Anat Sci Educ 2024;17(5):926-931. [doi: 10.1002/ase.2270] [Medline: 36916887]
12. Alkhaaldi SMI, Kassab CH, Dimassi Z, et al. Medical student experiences and perceptions of ChatGPT and artificial intelligence: cross-sectional study. JMIR Med Educ 2023 Dec 22;9:e51302. [doi: 10.2196/51302] [Medline: 38133911]
13. Veras M, Dyer JO, Rooney M, Barros Silva PG, Rutherford D, Kairy D. Usability and efficacy of artificial intelligence chatbots (ChatGPT) for health sciences students: protocol for a crossover randomized controlled trial. JMIR Res Protoc 2023 Nov 24;12:e51873. [doi: 10.2196/51873] [Medline: 37999958]
14. Mohammad B, Supti T, Alzubaidi M, et al. The pros and cons of using ChatGPT in medical education: a scoping review. Stud Health Technol Inform 2023 Jun 29;305:644-647. [doi: 10.3233/SHTI230580] [Medline: 37387114]
15. Srivastav S, Chandrakar R, Gupta S, et al. ChatGPT in radiology: the advantages and limitations of artificial intelligence for medical imaging diagnosis. Cureus 2023 Jul;15(7):e41435. [doi: 10.7759/cureus.41435] [Medline: 37546142]
16. Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P. Introducing artificial intelligence training in medical education. JMIR Med Educ 2019 Dec 3;5(2):e16048. [doi: 10.2196/16048] [Medline: 31793895]
17. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023 Mar 19;11(6):887. [doi: 10.3390/healthcare11060887] [Medline: 36981544]
18. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell 2023;6:1169595 [FREE Full text] [doi: 10.3389/frai.2023.1169595] [Medline: 37215063]
19. Syed W, Basil A Al-Rawi M. Assessment of awareness, perceptions, and opinions towards artificial intelligence among healthcare students in Riyadh, Saudi Arabia. Medicina (Kaunas) 2023 Apr 24;59(5):828. [doi: 10.3390/medicina59050828] [Medline: 37241062]
20. Abouammoh N, Alhasan K, Raina R, et al. Exploring perceptions and experiences of ChatGPT in medical education: a qualitative study among medical college faculty and students in Saudi Arabia. medRxiv. Preprint posted online on Jul 16, 2023. [doi: 10.1101/2023.07.13.23292624]
21. Ajlouni AO, Wahba FAA, Almahaireh AS. Students' attitudes towards using ChatGPT as a learning tool: the case of the University of Jordan. Int J Interact Mob Technol 2023;17(18):99-117. [doi: 10.3991/ijim.v17i18.41753]
22. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34. [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
23. Su J, Yang W. Unlocking the power of ChatGPT: a framework for applying generative AI in education. ECNU Review of Education 2023;6(3):355-366. [doi: 10.1177/20965311231168423]
24. Wartman SA, Combs CD. Medical education must move from the information age to the age of artificial intelligence. Acad Med 2018 Aug;93(8):1107-1109. [doi: 10.1097/ACM.0000000000002044] [Medline: 29095704]
25. Gong B, Nugent JP, Guest W, et al. Influence of artificial intelligence on Canadian medical students' preference for radiology specialty: a national survey study. Acad Radiol 2019 Apr;26(4):566-577. [doi: 10.1016/j.acra.2018.10.007] [Medline: 30424998]
26. Teng M, Singla R, Yau O, et al. Health care students' perspectives on artificial intelligence: countrywide survey in Canada. JMIR Med Educ 2022 Jan 31;8(1):e33390. [doi: 10.2196/33390] [Medline: 35099397]
27. Weidener L, Fischer M. Artificial intelligence in medicine: cross-sectional study among medical students on application, education, and ethical aspects. JMIR Med Educ 2024 Jan 5;10:e51247. [doi: 10.2196/51247] [Medline: 38180787]
28. Kapsali MZ, Livanis E, Tsalikidis C, Oikonomou P, Voultsos P, Tsaroucha A. Ethical concerns about ChatGPT in healthcare: a useful tool or the tombstone of original and reflective thinking? Cureus 2024 Feb;16(2):e54759. [doi: 10.7759/cureus.54759] [Medline: 38523987]

XSL•FO
RenderX

29. Zhai X. ChatGPT user experience: implications for education. SSRN Journal 2022 Dec 27. [doi: 10.2139/ssrn.4312418]
30. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT - reshaping medical education and clinical management. Pak J Med Sci 2023;39(2):605-607. [doi: 10.12669/pjms.39.2.7653] [Medline: 36950398]

Abbreviations

AI: artificial intelligence
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
IDEE : Identify, Discern, Ethics, Engage

Edited by B Lesselroth; submitted 09.06.24; peer-reviewed by E Bai, I Zaletel; revised version received 21.10.24; accepted 02.01.25; published 30.01.25.

Please cite as:
Elhassan SE, Sajid MR, Syed AM, Fathima SA, Khan BS, Tamim H
Assessing Familiarity, Usage Patterns, and Attitudes of Medical Students Toward ChatGPT and Other Chat-Based AI Apps in Medical Education: Cross-Sectional Questionnaire Study
JMIR Med Educ 2025;11:e63065
URL: https://mededu.jmir.org/2025/1/e63065
doi:10.2196/63065

© Safia Elwaleed Elhassan, Muhammad Raihan Sajid, Amina Mariam Syed, Sidrah Afreen Fathima, Bushra Shehroz Khan, Hala Tamim. Originally published in JMIR Medical Education (https://mededu.jmir.org), 30.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Performance Evaluation and Implications of Large Language Models in Radiology Board Exams: Prospective Comparative Analysis

Boxiong Wei, MD

Department of Ultrasound, Peking University First Hospital, 8 Xishiku Rd, Xicheng District, Beijing, China

Corresponding Author:
Boxiong Wei, MD
Department of Ultrasound, Peking University First Hospital, 8 Xishiku Rd, Xicheng District, Beijing, China

Abstract

Background: Artificial intelligence advancements have enabled large language models to significantly impact radiology education and diagnostic accuracy.

Objective: This study evaluates the performance of mainstream large language models, including GPT-4, Claude, Bard, Tongyi Qianwen, and Gemini Pro, in radiology board exams.

Methods: A comparative analysis of 150 multiple-choice questions from radiology board exams without images was conducted. Models were assessed on their accuracy for text-based questions and were categorized by cognitive levels and medical specialties using χ^2 tests and ANOVA.

Results: GPT-4 achieved the highest accuracy (83.3%, 125/150), significantly outperforming all other models. Specifically, Claude achieved an accuracy of 62% (93/150; *P*<.001), Bard 54.7% (82/150; *P*<.001), Tongyi Qianwen 70.7% (106/150; *P*=.009), and Gemini Pro 55.3% (83/150; *P*<.001). The odds ratios compared to GPT-4 were 0.33 (95% CI 0.18 - 0.60) for Claude, 0.24 (95% CI 0.13 - 0.44) for Bard, and 0.25 (95% CI 0.14 - 0.45) for Gemini Pro. Tongyi Qianwen performed relatively well with an accuracy of 70.7% (106/150; *P*=0.02) and had an odds ratio of 0.48 (95% CI 0.27 - 0.87) compared to GPT-4. Performance varied across question types and specialties, with GPT-4 excelling in both lower-order and higher-order questions, while Claude and Bard struggled with complex diagnostic questions.

Conclusions: GPT-4 and Tongyi Qianwen show promise in medical education and training. The study emphasizes the need for domain-specific training datasets to enhance large language models' effectiveness in specialized fields like radiology.

(JMIR Med Educ 2025;11:e64284) doi:10.2196/64284

KEYWORDS

large language models; LLM; artificial intelligence; AI; GPT-4; radiology exams; medical education; diagnostics; medical training; radiology; ultrasound

Introduction

Artificial intelligence (AI) in radiology has significantly improved diagnostic accuracy and educational methods for radiologists. By using advanced machine learning and deep learning techniques, AI applications have evolved from enhancing image interpretation to supporting complex diagnostic decisions [1]. These advancements not only increase the efficiency of diagnostic processes but also provide radiologists with interactive training simulations, crucial for their professional growth and certification readiness [2-9].

Recent advancements have also emerged with the development of large language models (LLMs) like GPT-4, Claude, Bard, Tongyi Qianwen and Gemini Pro. These models have added a new aspect to medical education by producing medically accurate content and supporting advanced diagnostic reasoning exercises [10,11]. These features are crucial for establishing safe learning spaces where future radiologists can practice detailed diagnostic reasoning and decision-making without real-world clinical risks [12,13]. Moreover, these LLMs are crucial in developing and clarifying complex medical scenarios and test questions, improving the educational experience and boosting the diagnostic abilities of students [14-16].

Despite these advancements, recent research has pinpointed limitations in the use of LLMs in medical exams, particularly in specialties like radiology that demand extensive clinical insight. Studies have shown that while LLMs such as GPT-4 can manage simple diagnostic questions effectively, they encounter difficulties with more complex cases that require a deeper clinical understanding and the integration of diverse medical information [17,18]. These findings highlight a significant gap in the existing literature; there is a lack of

XSL•FO
RenderX

comprehensive comparative studies that evaluate the performance of various LLMs across different diagnostic scenarios in radiology [19].

This study addresses this gap by comparing several mainstream LLMs in text-based radiology board exams, without imaging components, evaluating their overall performance. While a secondary objective is to analyze performance by question type and topic. This study hypothesizes that GPT-4 will outperform other models, particularly in handling complex diagnostic questions.

Methods

Study Design

This research was structured as a prospective, comparative analysis that aimed to test the effectiveness of various notable LLMs within a controlled environment resembling radiology board examinations without images. The radiology exams comprehensively evaluated a candidate's radiology knowledge, reasoning, and clinical skills. China does not currently have a unified national licensing exam specifically for radiologists. Given that the Canadian Royal College and American Board of Radiology exams are viewed as authoritative and widely recognized, test questions were selected according to the standards of these two exams for model testing [20]. Both of the exams assess candidates on a broad spectrum of radiology topics using multiple-choice questions.

Ethical Considerations

Despite the reliance on nonpersonal, pre-existing data and the lack of direct involvement of human or animal subjects, ethical approval and the need for informed consent were waived by the Institutional Review Board of Peking University First Hospital, Beijing, China. The radiologists who participated in question validation and categorization were compensated at a rate of 300 Chinese Yuan (US $40.91) per hour for their professional expertise. All data used in the study were anonymized exam questions, with no personal identifiable information involved. The research strictly adhered to ethical standards, with data integrity meticulously upheld throughout the study.

Models Selection

The models chosen for this investigation included GPT-4 (OpenAI), Claude 2.1 (Anthropic), Bard (Google, PaLM 2), Tongyi Qianwen (Alibaba, Qwen-72B), and Gemini Pro 1.0 (Google). All models were tested from late November to early December 2023. These models represent significant advancements in AI, particularly in natural language processing. They were selected based on their demonstrated success in academic and professional settings, indicating their potential effectiveness in educational applications.

Dataset Composition

The dataset for this study consisted of 150 multiple-choice questions drawn from historical radiology board exams similar to those given by the Canadian Royal College and the American Board of Radiology. These questions were sourced from the websites of Board Vitals [21] and CanadaQBank [22], which are widely recognized for providing questions that closely reflect the content and format of North American radiology board exams. Each question was individually reviewed and validated by two academic radiologists—one specializing in ultrasound with 20 years of experience and the other in abdominal radiology with 4 years of experience. Questions were only included if both reviewers concurred on their relevance and appropriateness for this study. Questions that involved images were excluded.

Question Categorization

All questions were classified according to their primary assessment objectives using Bloom's Taxonomy, including two main categories: lower-order thinking (remembering and understanding) and higher-order thinking (applying, analyzing, and evaluating) [23]. Higher-order thinking questions were further divided into specific groups such as description and analysis of image findings, application of concepts, clinical management, and calculation and classification. Additionally, questions were also classified based on the specific area of disease focus, including digestive, genitourinary, musculoskeletal, respiratory, cardiovascular (including angiography and intervention), nervous, breast and thyroid, pediatrics, and imaging basics and physics. Each question was reviewed and categorized independently by the two board-certified radiologists mentioned above. Any disagreements were then discussed collectively to arrive at a consensus.

Scoring Criteria

The Canadian Royal College examination uses a pass-fail system based on achieving at least 70% on all written components of the examination. The American Board of Radiology uses a criterion-referenced scoring system. This means that candidates are evaluated against a predefined standard, not in comparison to other test-takers. The passing standard is typically set by a group of experts, including residency program directors and experienced clinicians, who determine the difficulty level of each question to ensure it aligns with the required competency for independent practice. To pass, candidates must meet or exceed the passing standard for all categories scored together. For both exams, the questions undergo psychometric validation, and questions that are not effective in discriminating between candidates or are found too difficult may be removed. The threshold for passing in this study was set at 70% to align with the standards of the Royal College examinations in Canada. This study did not use the criterion-referenced scoring system used by the American Board of Radiology because its standards were difficult to ascertain. Each multiple-choice question was inputted into different LLMs, and the first response from each model was recorded as the subject of analysis.

Statistical Analysis

To evaluate the association between model type and accuracy for categorical variables, χ^2 tests were used. For categories with small sample sizes, the Fisher exact test was used to ensure the validity of the statistical results. Odds ratios and their corresponding 95% CIs were calculated using GPT-4 as the benchmark. ANOVA was used to compare the mean accuracy rates across different models. Following the results from the ANOVA, Tukey's honestly significant difference test was

XSL•FO
RenderX

applied to identify specific pairs of models that demonstrated significant differences in performance. Cohen *d* was calculated to quantify the magnitude of differences between the models, providing a clearer understanding of the practical significance of the findings. Split-half reliability testing was used to assess the consistency of each model's performance across different subsets of data, ensuring the reliability of the models over varied test conditions. Statistical significance was set at an α level of .05.

Results

Overall Model Performance

GPT-4 emerged as the leading model with an accuracy rate of 83.3% (125/150), significantly outperforming its peers. Tongyi Qianwen also displayed strong performance, recording a 70.7% (106/150) accuracy. Moderate effectiveness was observed in models like Claude and Gemini Pro, with accuracy rates of 62.0% (93/150) and 55.3% (83/150), respectively. Bard trailed with a 54.7% (82/150) accuracy rate, highlighting its challenges in handling complex medical data under exam conditions (Table 1).

Table . Performance of different large language models on radiology board–styled multiple-choice questions without images.

Parameter	Test score, n (%)				
	GPT4	Claude	Bard	Tongyi Qianwen	Gemini Pro
All questions (n=150)	125 (83.3)	93 (62.0)	82 (54.7)	106 (70.7)	83 (55.3)
Question type					
Lower order thinking (n=46)	38 (82.6)	34 (73.9)	27 (58.7)	34 (73.9)	29 (63)
Higher order thinking (n=104)	87 (83.7)	59 (56.7)	55 (52.9)	72 (69.2)	54 (51.9)
Higher order thinking question categories					
Description and analyze of image findings (n=35)	30 (85.7)	23 (65.7)	20 (57.1)	28 (80)	21 (60)
Application of concepts (n=38)	34 (89.5)	19 (50)	17 (44.7)	26 (68.4)	17 (44.7)
Clinical management (n=19)	14 (73.7)	12 (63.2)	12 (63.2)	13 (68.4)	11 (57.9)
Calculation and classification (n=12)	9 (75)	5 (41.7)	6 (50)	5 (41.7)	5 (41.7)
Question topic					
Digestive (n=15)	10 (66.7)	7 (46.7)	5 (33.3)	10 (66.7)	9 (60)
Genitourinary (n=21)	19 (90.5)	15 (71.4)	14 (66.7)	15 (71.4)	11 (52.4)
Musculoskeletal (n=11)	8 (72.7)	6 (54.5)	7 (63.6)	9 (81.8)	7 (63.6)
Respiratory (n=15)	12 (80)	9 (60)	8 (53.3)	8 (53.3)	8 (53.3)
Cardiovascular (n=22)	19 (86.4)	14 (63.6)	8 (36.4)	18 (81.8)	11 (50)
Nervous (n=11)	11 (100)	9 (81.8)	7 (63.6)	8 (72.7)	9 (81.8)
Breast and thyroid (n=14)	11 (78.6)	9 (64.3)	9 (64.3)	9 (64.3)	7 (50)
Pediatrics (n=19)	15 (78.9)	11 (57.9)	11 (57.9)	13 (68.4)	9 (47.4)
Imaging Basics and physics (n=22)	19 (86.4)	11 (50)	12 (54.5)	15 (68.2)	12 (54.5)

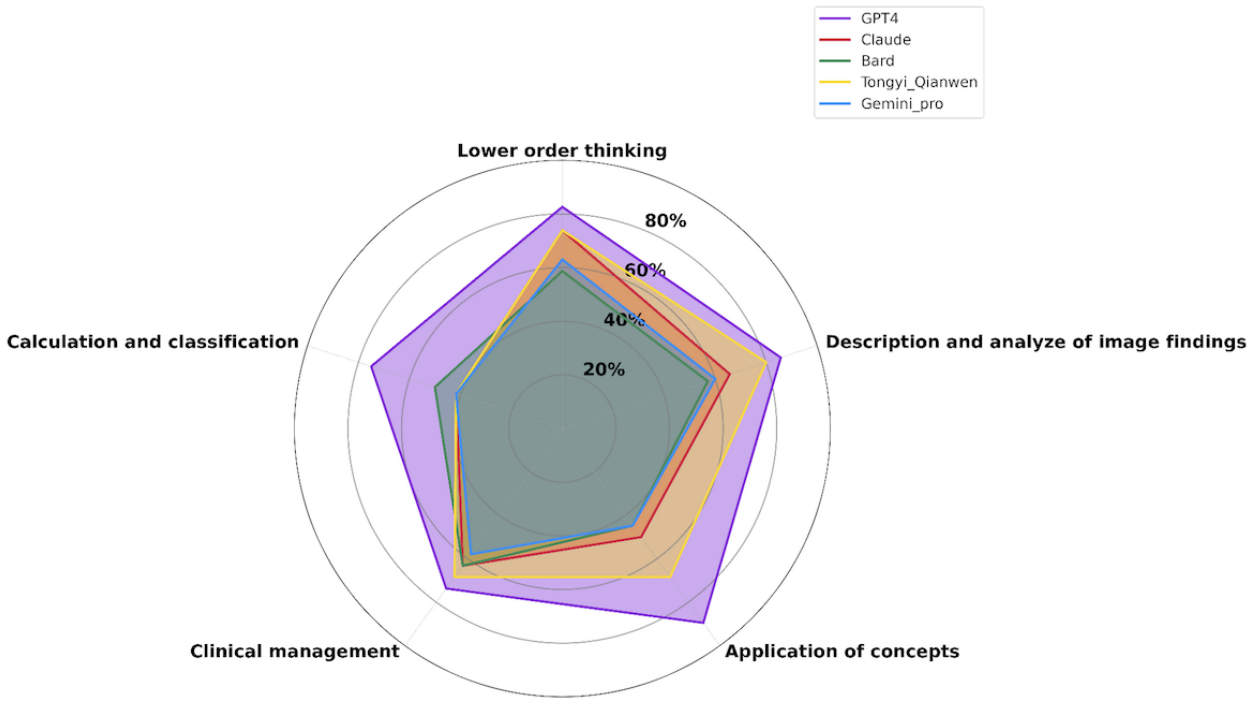
Detailed Performance Analysis by Question Type

The breakdown by question type revealed that GPT-4 consistently excelled in both lower-order and higher-order thinking questions, scoring 82.6% (38/46) and 83.7% (87/104), respectively. This indicated GPT-4's capability to manage both basic recall and more complex analytical tasks effectively. In contrast, models such as Claude and Bard demonstrated a drop in performance with higher-order thinking questions, achieving only 56.7% (59/104) and 52.9% (55/104) accuracy in this

XSL•FO
RenderX

category, respectively. This gradient in performance highlighted the difficulties faced by current LLMs in simulating the complex cognitive processes involved in clinical reasoning (Figure 1).

Figure 1. Model accuracy by question type, illustrating the differentiation in model performance between lower-order and higher-order thinking questions.



Performance Across Medical Specialties

Performance analysis segmented by medical specialty showed marked variances. GPT-4 demonstrated exceptional proficiency in neurology with a perfect score of 100% (11/11), and also performed well in genitourinary and cardiovascular categories, with accuracies of 90.5% (19/21) and 86.4% (19/22), respectively. However, challenges were apparent in areas like musculoskeletal and digestive categories, where high-performing models like GPT-4 experienced reduced accuracy rates of 72.7% (8/11) and 66.7% (10/15), respectively. These results indicated that some specialties may need more tailored domain-specific training for models to enhance their effectiveness (Table 1).

Detailed odds ratios and CIs for each model are presented in Multimedia Appendix 1. The odds ratio results show that GPT-4 had the highest performance. All the other models had significantly lower odds ratios compared to GPT-4. Tongyi Qianwen had the highest odds ratio among the other models. As shown in Multimedia Appendix 2, the pairwise comparisons showed that GPT-4 significantly outperformed all other models, with statistically significant differences observed in its comparison with Claude ($P<.001$), Bard ($P<.001$), Tongyi Qianwen ($P=.009$), and Gemini Pro ($P<.001$). Additionally, Tongyi Qianwen exhibited a significantly higher accuracy compared to Bard ($P=.004$) and Gemini Pro ($P=.006$). In contrast, no statistically significant differences were found between Claude and Bard ($P=.20$), Claude and Gemini Pro ($P=.24$), or Bard and Gemini Pro ($P=.90$). These results suggest that the performance of these models was relatively similar in this dataset.

Discussion

Principal Findings

The exceptional performance of GPT-4 in this study aligns with recent findings that highlight its advanced reasoning capabilities and improvements over previous versions, such as GPT-3.5, in various professional contexts, including various kinds of medical exams [24]. GPT-4's extensive training on diverse datasets and its refined architecture enable it to adeptly handle complex questions, which are typical in the specialized language and scenario-based queries found in medical board examinations [25]. Nevertheless, the performance differences observed among models like Bard and Claude can be attributed to the nature of their training and inherent limitations in processing complex cognitive tasks, which are crucial in radiology examinations. This is largely due to the absence of specialized medical training data during their development phases. These findings are in line with the research, which indicated that while GPT-4's textual reasoning is strong, its integration and analysis of image-based information remains inadequate [26].

Models such as GPT-4 and Tongyi Qianwen, which displayed superior performance, likely benefited from training datasets that included medical scenarios. The significance of domain-specific training is well-documented, emphasizing that for LLMs to excel in specialized fields like radiology, they require training with pertinent medical data. Both GPT-4 and Tongyi Qianwen exceeded the 70% passing threshold for the simulated radiology board exams. This marks a significant achievement and shows the potential of these models in

XSL•FO
RenderX

academic and professional environments. The threshold mirrors real medical licensing exam criteria, offering a realistic measure of AI's potential performance in actual educational assessments. The robust performance of Tongyi Qianwen, particularly in an English-based setup, is notable. Despite generally not being ranked as highly as Western models in AI benchmarks, its performance indicates significant progress in China's AI development [27]. This supports calls for more inclusive and diverse training datasets to reduce biases and improve the global applicability of AI technologies.

GPT-4 has demonstrated the capability to pass simulated UK Radiology Fellowship Examinations, especially in sections focused on physics and single best answers [28]. However, challenges remain when these models are tested with image-based questions, highlighting a persisting gap between current AI capabilities and the complex demands of radiological diagnostics [26]. While integrating LLMs into medical education and assessments promises transformative changes in how content is delivered and evaluated, there is a risk of excessive reliance on AI. This overdependence could potentially undermine the development of critical thinking and diagnostic skills vital for medical practice [25].

Limitations

This study's limitations include its sole focus on text-based questions and the exclusion of visual components, which are integral to radiology. Future research should incorporate multimodal assessments and also aim to integrate image recognition capabilities with textual analysis to improve the applicability of LLMs in radiology. These models will need to be fine-tuned with domain-specific datasets to enhance their practical utility in medical education and clinical diagnostics. Another notable limitation is the delay between the submission and publication of peer-reviewed articles, which can result in outdated assessments of rapidly evolving LLMs. The models evaluated in this paper were based on their versions from late November to early December 2023, and significant advancements have occurred since then, particularly with models like Claude, which has been regularly updated, with multiple new versions released by Anthropic. In future work, we intend to continue discussing the accuracy comparisons among new models as they are released. Additionally, if sufficient technical resources are available, we aim to create a platform to maintain an up-to-date database of LLM performance on this benchmark.

Conclusion

This article underscores the evolving capabilities and limitations of LLMs in medical education. While models like GPT-4 show promise, the path to their effective integration in clinical practice requires ongoing refinement and a deeper understanding of their operational dynamics in complex medical settings.

Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The odds ratios and CIs of each model using GPT-4 as the benchmark.
[DOCX File, 17 KB - mededu_v11i1e64284_app1.docx]

Multimedia Appendix 2
Hypothetical pairwise comparison table.
[DOCX File, 17 KB - mededu_v11i1e64284_app2.docx]

References

1. Aggarwal R, Sounderajah V, Martin G, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. NPJ Digit Med 2021 Apr 7;4(1):65. [doi: 10.1038/s41746-021-00438-z] [Medline: 33828217]
2. Cabitza F, Campagner A, Balsano C. Bridging the "last mile" gap between AI implementation and operation: "data awareness" that matters. Ann Transl Med 2020 Apr;8(7):501. [doi: 10.21037/atm.2020.03.63] [Medline: 32395545]
3. Thrall JH, Li X, Li Q, et al. Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. J Am Coll Radiol 2018 Mar;15(3 Pt B):504-508. [doi: 10.1016/j.jacr.2017.12.026] [Medline: 29402533]
4. Akinci D'Antonoli T, Stanzione A, Bluethgen C, et al. Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions. Diagn Interv Radiol 2024 Mar 6;30(2):80-90. [doi: 10.4274/dir.2023.232417] [Medline: 37789676]
5. Nassiri K, Akhloufi MA. Recent advances in large language models for healthcare. BioMed Inform 2024;4(2):1097-1143. [doi: 10.3390/biomedinformatics4020062]

XSL•FO
RenderX

6. Duong MT, Rauschecker AM, Rudie JD, et al. Artificial intelligence for precision education in radiology. Br J Radiol 2019 Nov;92(1103):20190389. [doi: 10.1259/bjr.20190389] [Medline: 31322909]
7. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health 2019 Oct;1(6):e271-e297. [doi: 10.1016/S2589-7500(19)30123-2] [Medline: 33323251]
8. Papadimitroulas P, Brocki L, Christopher Chung N, et al. Artificial intelligence: deep learning in oncological radiomics and challenges of interpretability and data harmonization. Phys Med 2021 Mar;83:108-121. [doi: 10.1016/j.ejmp.2021.03.009] [Medline: 33765601]
9. Seah JCY, Tang CHM, Buchlak QD, et al. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. Lancet Digit Health 2021 Aug;3(8):e496-e506. [doi: 10.1016/S2589-7500(21)00106-0] [Medline: 34219054]
10. Nichols JA, Herbert Chan HW, Baker MAB. Machine learning: applications of artificial intelligence to imaging and diagnosis. Biophys Rev 2019 Feb;11(1):111-118. [doi: 10.1007/s12551-018-0449-9] [Medline: 30182201]
11. Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ 2020 Mar 25;368:m689. [doi: 10.1136/bmj.m689] [Medline: 32213531]
12. Brown T, Mann B, Ryder N, et al. Language models are few-shot learners. 2020 Presented at: 34th Conference on Neural Information Processing Systems; Dec 6-12, 2020; Vancouver, Canada p. 1877-1901 URL: https://dl.acm.org/doi/abs/10.5555/3495724.3495883 [accessed 2025-01-06]
13. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Presented at: 31st International Conference on Neural Information Processing Systems; Dec 4-9, 2017; Long Beach, CA p. 6000-6010 URL: https://dl.acm.org/doi/10.5555/3295222.3295349 [accessed 2025-01-06]
14. Yu AC, Mohajer B, Eng J. External validation of deep learning algorithms for radiologic diagnosis: a systematic review. Radiol Artif Intell 2022 May;4(3):e210064. [doi: 10.1148/ryai.210064] [Medline: 35652114]
15. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. Radiology 2018 Mar;286(3):800-809. [doi: 10.1148/radiol.2017171920] [Medline: 29309734]
16. Fischer AM, Eid M, De Cecco CN, et al. Accuracy of an artificial intelligence deep learning algorithm implementing a recurrent neural network with long short-term memory for the automated detection of calcified plaques from coronary computed tomography angiography. J Thorac Imaging 2020 May;35 Suppl 1:S49-S57. [doi: 10.1097/RTI.0000000000000491] [Medline: 32168163]
17. McBee MP, Awan OA, Colucci AT, et al. Deep learning in radiology. Acad Radiol 2018 Nov;25(11):1472-1480. [doi: 10.1016/j.acra.2018.02.018] [Medline: 29606338]
18. Li L, Qin L, Xu Z, et al. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. Radiology 2020 Aug;296(2):E65-E71. [doi: 10.1148/radiol.2020200905] [Medline: 32191588]
19. Rauschecker AM, Rudie JD, Xie L, et al. Artificial intelligence system approaching neuroradiologist-level differential diagnosis accuracy at brain MRI. Radiology 2020 Jun;295(3):626-637. [doi: 10.1148/radiol.2020190283] [Medline: 32255417]
20. Wang YE, Liu M, Jin L, et al. Radiology education in China. J Am Coll Radiol 2013 Mar;10(3):213-219. [doi: 10.1016/j.jacr.2012.11.006] [Medline: 23571062]
21. The standard in healthcare board exam prep & CME. Board Vitals. URL: https://www.boardvitals.com/ [accessed 2025-01-06]
22. CanadaQBank. URL: https://www.canadaqbank.com/ [accessed 2025-01-06]
23. Krathwohl DR. A revision of Bloom's taxonomy: an overview. Theor Pract 2002 Nov 1;41(4):212-218. [doi: 10.1207/s15430421tip4104_2]
24. Bhayana R, Bleakney RR, Krishna S. GPT-4 in radiology: improvements in advanced reasoning. Radiology 2023 Jun;307(5):e230987. [doi: 10.1148/radiol.230987] [Medline: 37191491]
25. Lourenco AP, Slanetz PJ, Baird GL. Rise of ChatGPT: it may be time to reassess how we teach and test radiology residents. Radiology 2023 Jun;307(5):e231053. [doi: 10.1148/radiol.231053] [Medline: 37191490]
26. Kim H, Kim P, Joo I, Kim JH, Park CM, Yoon SH. ChatGPT vision for radiological interpretation: an investigation using medical school radiology examinations. Korean J Radiol 2024 Apr;25(4):403-406. [doi: 10.3348/kjr.2024.0017] [Medline: 38528699]
27. Jiang L, Wu Z, Xu X, et al. Opportunities and challenges of artificial intelligence in the medical field: current application, emerging problems, and problem-solving strategies. J Int Med Res 2021 Mar;49(3):3000605211000157. [doi: 10.1177/03000605211000157] [Medline: 33771068]
28. Ariyaratne S, Jenko N, Mark Davies A, Iyengar KP, Botchu R. Could ChatGPT pass the UK radiology fellowship examinations? Acad Radiol 2024 May;31(5):2178-2182. [doi: 10.1016/j.acra.2023.11.026] [Medline: 38160089]

XSL•FO
RenderX

Abbreviations

AI: artificial intelligence
LLM: large language model

Edited by B Lesselroth; submitted 14.07.24; peer-reviewed by B Thies, R Yin; revised version received 10.10.24; accepted 03.12.24; published 16.01.25.

Please cite as:
Wei B
Performance Evaluation and Implications of Large Language Models in Radiology Board Exams: Prospective Comparative Analysis
JMIR Med Educ 2025;11:e64284
URL: https://mededu.jmir.org/2025/1/e64284
doi:10.2196/64284

© Boxiong Wei. Originally published in JMIR Medical Education (https://mededu.jmir.org), 16.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Factors Associated With the Accuracy of Large Language Models in Basic Medical Science Examinations: Cross-Sectional Study

Naritsaret Kaewboonlert, MD; Jiraphon Poontananggul, MD; Natthipong Pongsuwan, MD; Gun Bhakdisongkhram, PhD, MD

Institute of Medicine, Suranaree University of Technology, 111 University Avenue, Nakhon Ratchasima, Thailand

Corresponding Author:
Naritsaret Kaewboonlert, MD
Institute of Medicine, Suranaree University of Technology, 111 University Avenue, Nakhon Ratchasima, Thailand

Abstract

Background: Artificial intelligence (AI) has become widely applied across many fields, including medical education. Content validation and its answers are based on training datasets and the optimization of each model. The accuracy of large language model (LLMs) in basic medical examinations and factors related to their accuracy have also been explored.

Objective: We evaluated factors associated with the accuracy of LLMs (GPT-3.5, GPT-4, Google Bard, and Microsoft Bing) in answering multiple-choice questions from basic medical science examinations.

Methods: We used questions that were closely aligned with the content and topic distribution of Thailand's Step 1 National Medical Licensing Examination. Variables such as the difficulty index, discrimination index, and question characteristics were collected. These questions were then simultaneously input into ChatGPT (with GPT-3.5 and GPT-4), Microsoft Bing, and Google Bard, and their responses were recorded. The accuracy of these LLMs and the associated factors were analyzed using multivariable logistic regression. This analysis aimed to assess the effect of various factors on model accuracy, with results reported as odds ratios (ORs).

Results: The study revealed that GPT-4 was the top-performing model, with an overall accuracy of 89.07% (95% CI 84.76% - 92.41%), significantly outperforming the others (P<.001). Microsoft Bing followed with an accuracy of 83.69% (95% CI 78.85% - 87.80%), GPT-3.5 at 67.02% (95% CI 61.20% - 72.48%), and Google Bard at 63.83% (95% CI 57.92% - 69.44%). The multivariable logistic regression analysis showed a correlation between question difficulty and model performance, with GPT-4 demonstrating the strongest association. Interestingly, no significant correlation was found between model accuracy and question length, negative wording, clinical scenarios, or the discrimination index for most models, except for Google Bard, which showed varying correlations.

Conclusions: The GPT-4 and Microsoft Bing models demonstrated equal and superior accuracy compared to GPT-3.5 and Google Bard in the domain of basic medical science. The accuracy of these models was significantly influenced by the item's difficulty index, indicating that the LLMs are more accurate when answering easier questions. This suggests that the more accurate models, such as GPT-4 and Bing, can be valuable tools for understanding and learning basic medical science concepts.

(JMIR Med Educ 2025;11:e58898) doi:10.2196/58898

KEYWORDS

accuracy; performance; artificial intelligence; AI; ChatGPT; large language model; LLM; difficulty index; basic medical science examination; cross-sectional study; medical education; datasets; assessment; medical science; tool; Google

Introduction

Advances in artificial intelligence (AI), machine learning, and large language models (LLMs) have made these tools widely used across a variety of industries. Education and other fields are increasingly using these technologies for decision-making and predictive analysis, using machine learning fed by large databases [1]. Their utility has expanded to a wide range of applications, including speech recognition, image categorization, and language translation [2].

The application of computer technologies to study and create models for decision-making, prediction, and simulation is known as machine learning. Model performance is based on training datasets. The incorporation of AI into traditional health care and medical education has had a substantial impact on medical practices [3]. It has accelerated diagnostic processes in radiography [4], pathology, endoscopy, and ultrasonography, has improved clinical decision-making, and has decreased the workloads of health care personnel. AI has had an impact on pharmaceutical development and management and medical education, resulting in a new paradigm [5].

XSL•FO
RenderX

A study on the accuracy of ChatGPT in answering questions that were contextually similar to those in the United States Medical Licensing Examination (USMLE) reported accuracy rates of 44% - 64% for step 1 and 42% - 57.8% for step 2, depending on the dataset [6]. This research indicated that the model's accuracy in answering questions matched the passing score for third-year medical students, suggesting that further development is required for ChatGPT to meet or exceed the USMLE passing criteria [7]. Additionally, the model has the potential to generate insightful content that could aid human learners in studying medical sciences [8].

Evaluations of ChatGPT's accuracy in answering university-level physiology examination questions have shown it can correctly answer more than 75% of them. Furthermore, it can provide explanations that align with expert assessments [9]. For specialized surgical studies, ChatGPT's GPT-4 model, an evolution of GPT-3.5, has been used to assess surgical question accuracy, revealing an overall accuracy of 76.4%, compared to 46.8% with GPT-3.5, a statistically significant difference ($P<.05$). GPT-4 showed an accuracy range of 63.6% - 88.3% across different topics, outperforming GPT-3.5 in every subtopic [10].

In terms of answering questions for family medicine experts in Taiwan, ChatGPT demonstrated an accuracy of 41.6% in a study that also found that the length of the questions did not affect the model's accuracy. However, the authors noted that the AI's accuracy might depend on the difficulty of the test, the local language, and medical practices, which differ by region and could reduce the model's accuracy [11].

This study investigated the accuracy of responses from widely used LLM AIs, including ChatGPT (with GPT-3.5 and GPT-4), Bing, and Google Bard. Also, we compared their accuracy and determined relationships with the difficulty index for multiple-choice questions closely related to the content of the Thailand Center for Medical Competency Assessment step 1, as well as other factors that may affect the AI's accuracy, such as the length of the question, the presence of negatively worded questions, and the variety of topics across various systems. This research was undertaken to explore these dimensions.

Methods

Study Design and Setting

This study was carried out at the Institute of Medicine, Suranaree University of Technology, Thailand. The curriculum has been accredited by the World Federation for Medical Education since 2021, and the program enrolls 92 medical students annually. Preclinical medical students receive instruction through a collaboration between the School of Preclinic, the Institute of Science, and the Institute of Medicine.

Ethical Considerations

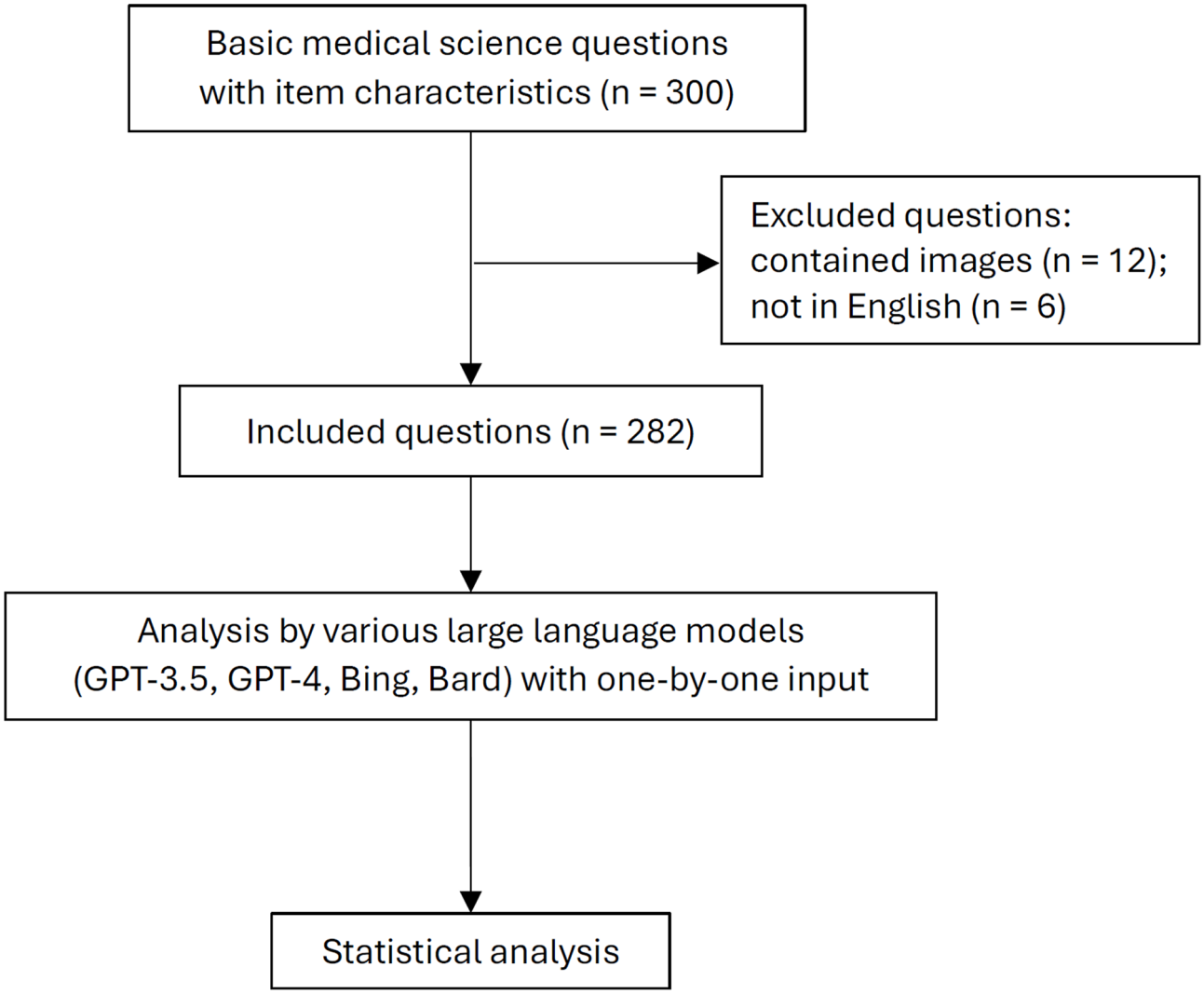
The Human Research Ethics Committee at Suranaree University of Technology approved an exemption (certificate of exemption 117/2566) for this study, which was conducted in accordance with international guidelines for human research.

Data Source

This study used a set of 300 multiple-choice questions that closely matched the content and topic distribution of Thailand's step 1 National Medical Licensing Examination. These questions were voluntarily administered to third-year medical students in February 2021 and 2022. This timing was chosen because the students had already completed courses relevant to the examination. The difficulty index and discrimination index of each question were assessed from the test. The same set of questions was used for both years without any modifications to the content of the exam. The study excluded questions that contained pictures or were not written in English. These exclusion criteria were applied to ensure consistency in the type of questions assessed and to maintain a focus on the textual comprehension and response accuracy of the LLMs (Figure 1).

XSL•FO
RenderX

Figure 1. Study flow chart.



Question Characteristics

Question length was defined as the number of words contained within a question. Negative word questions were identified as those containing the terms "not," "no," "exclude," or "neither." Case scenario questions were characterized by the inclusion of a clinical case scenario, providing a contextual background to the question being asked.

We also used item analyses [12-14], such as the difficulty index, discrimination index, and internal consistency reliability, as independent factors associated with the LLMs' accuracy.

Difficulty index (represented by the letter *p*) is the proportion of examinees who answered a specific question correctly. If a question is easy and every examinee answers it correctly, *p* will be 1. Conversely, if no examinees answer the question correctly, *p* will be 0. This index helps in evaluating the relative difficulty of each question in an examination [12].

Discrimination index (represented by the letter *r*) refers to a question's ability to differentiate between examinees who have high scores and those who do not. Questions with a high discrimination ability are characterized by high scorers typically answering them correctly, while low scorers tend to answer them incorrectly [13]. The most widely used metric for assessing a question's discrimination ability is the point-biserial correlation. The point-biserial correlation coefficient ranges from −1 to 1. A higher point-biserial correlation indicates a question with better discriminatory power.

Internal consistency reliability was measured with Cronbach α. It ranges from 0 to 1, with higher values indicating greater internal consistency. A Cronbach α value above 0.7 is generally considered acceptable, values above 0.8 are considered good, and values above 0.9 are considered excellent.

Prompt Input for LLMs

We used the prompt "Choose the best one answer." Each question was asked to each LLM after inputting the prompt during the same period, from January 18 to 24, 2024. We individually inputted the selected questions into various LLMs, including ChatGPT (with GPT-3.5 and GPT-4), Microsoft Bing, and Google Bard (one session contained one prompt and individual question). The responses from these models were then categorized as either correct or incorrect.

Statistical Analysis

In this study, discrete variables are represented as percentages, while continuous variables are represented as either the mean (SD) or median (IQR). The association between categorical

XSL•FO
RenderX

variables was analyzed using the χ^2 test or Fisher exact test. The relationships between variables and the ability of the LLMs to provide correct answers was examined using multivariable logistic regression, with results reported as odds ratios (ORs) and 95% CIs. Statistical significance was determined at a *P* value of <.05 for all tests. The analysis was facilitated by Stata (version 17; StataCorp), which was used for data analysis and chart creation.

Results

We evaluated the LLMs by using a set of 300 multiple-choice questions that were closely aligned with the content and topic distribution of Thailand's Step 1 National Medical Licensing Examination. According to the exclusion criteria, 12 picture-containing questions and 6 non-English questions were excluded; therefore, 282 eligible questions were included. All eligible questions were concurrently input into various LLMs (Figure 1). The responses were then recorded, categorizing the outcomes as either correct or incorrect.

The questions were categorized according to the block system (Table 1), with distributions as follows: 32.3% on general principles, 5.7% on the hematopoietic system, 8.2% on the nervous system, 3.9% on skin and connective tissues, 4.3% on the musculoskeletal system, 7.8% on the respiratory system, 8.9% on the cardiovascular system, 7.5% on the gastrointestinal system, 6.7% on the urinary system, 7.1% on the reproductive system, and 7.8% on the endocrine system. The average question length was 49.10 (SD 18.94) words, with 24 questions (8.2%) containing negative wording. More than half of the questions, specifically 53.2%, were based on clinical case scenarios (more descriptive statistics for the item analysis for each block are provided in Multimedia Appendix 1). The mean difficulty index was 0.35, indicating moderately difficult to difficult questions. The discrimination index was 0.16, suggesting a poor ability to distinguish between higher and lower performers. Otherwise, the internal consistency reliability, at 0.84, highlighted an acceptable level of consistency across the examination.

Table . Question characteristics (n=282).

Characteristics	Values
Number of questions by block, n (%)	
General principles[a]	91 (32.3)
Hematopoietic system	16 (5.7)
Nervous system	23 (8.2)
Skin and connective tissue	11 (3.9)
Musculoskeletal system	12 (4.3)
Respiratory system	22 (7.8)
Cardiovascular system	25 (8.9)
Gastrointestinal system	21 (7.5)
Urinary system	19 (6.7)
Reproductive system	20 (7.1)
Endocrine system	22 (7.8)
Question length (words), mean (SD)	49.10 (18.94)
Negative-word questions, n (%)	24 (8.5)
Case scenario questions, n (%)	150 (53.2)
Average difficulty index (*p*)	0.35
Average discrimination index (*r*)	0.16
Internal consistency reliability (α)	0.84

[a]"General principle" questions refer to fundamental principles in biochemistry, molecular biology, human development, genetics, normal immune responses, basic pathological processes, laboratory investigations, general pharmacology, epidemiology, and biostatistics.

The overall accuracy of the LLMs in the basic medical science examination was as follows (Table 2): GPT-4 achieved the highest accuracy at 89.07% (95% CI 84.76% - 92.41%), Microsoft Bing had an accuracy of 83.69% (95% CI 78.85% - 87.80%), GPT-3.5 recorded an accuracy of 67.02% (95% CI 61.20% - 72.48%), and Google Bard demonstrated an accuracy of 63.83% (95% CI 57.92% - 69.44%). The Fisher exact test showed that GPT-4 performed more accurately than Microsoft Bing, and that the difference was statistically significant (*P*<.001)

XSL•FO
RenderX

Table . Accuracy of large language models with 95% CIs, compared based on category (n=282).

	GPT-3.5	GPT-4	Microsoft Bing	Google Bard
Number of correct answers	189	251	236	180
Overall accuracy, % (95% CI)	67.02 (61.20 - 72.48)	89.07 (84.76 - 92.41)	83.69 (78.85 - 87.80)	63.83 (57.92 - 69.44)
General principles, % (95% CI)	84.62 (75.54 - 91.33)	90.11 (82.05 - 95.38)	84.62 (75.54 - 91.33)	72.53 (62.17 - 81.37)
Block system, % (95% CI)	61.78 (54.49 - 68.70)	88.48 (83.08 - 92.64)	83.25 (77.18 - 88.25)	59.69 (52.36 - 66.70)

The GPT-4 model demonstrated the highest accuracy among the LLMs in the general principles section for basic science, achieving 90.11% (95% CI 82.05% - 95.38%), as shown in Table 2. GPT-3.5 and Bing exhibited equal accuracy in this section, with the lowest accuracy being 72.53% (95% CI 62.17% - 81.37%) for Bard. Additionally, GPT-4 maintained its position as the top performer in the block system with an accuracy of 88.48% (95% CI 83.08% - 92.64%), whereas Bard again displayed the lowest performance in this segment (Figure 2). Overall, GPT-4 stood out for its superior performance in overall accuracy, general principles, and the block system.

Figure 2. Comparative accuracy with 95% CIs for artificial intelligence models across different question categories.

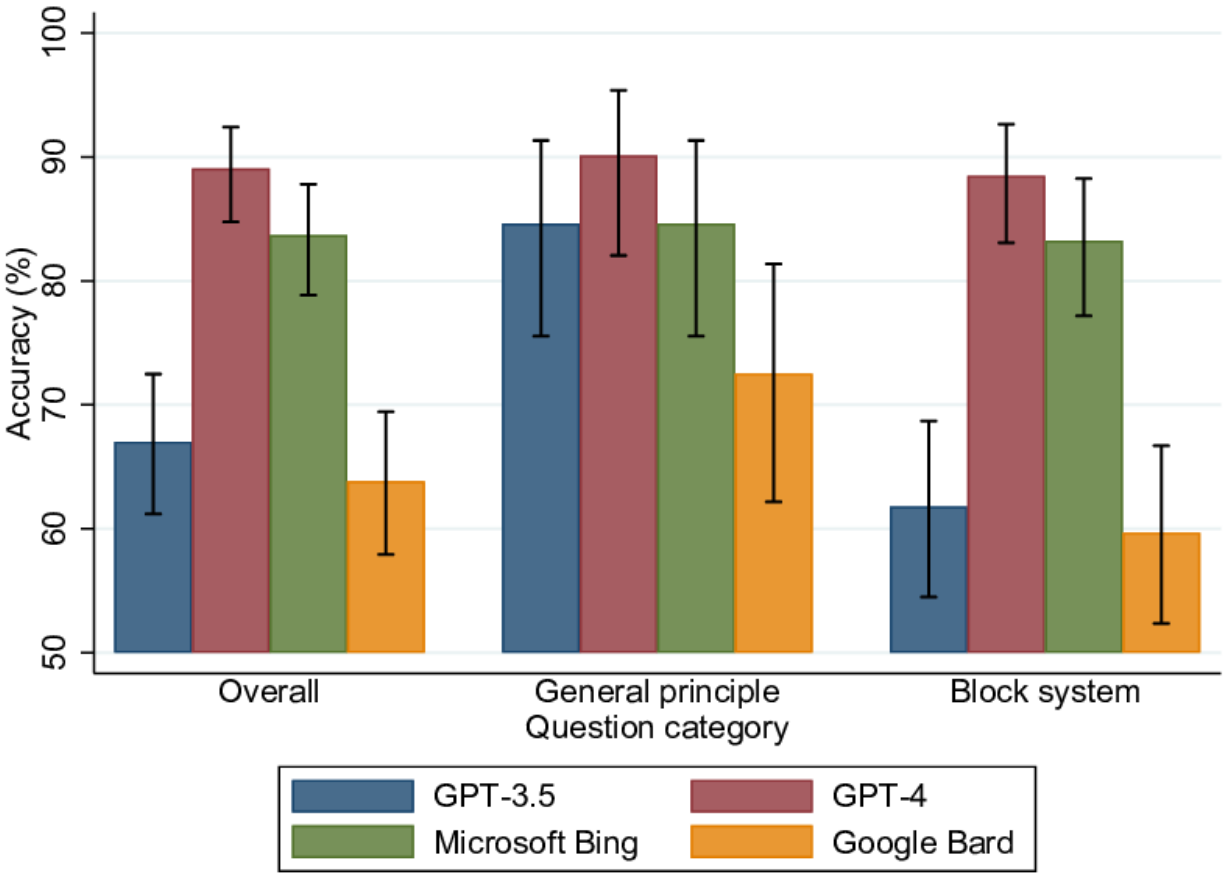


Table 3 presents the number of correct answers stratified by the block system alongside the proportion of correct answers relative to the total number of questions. The GPT-4 model exhibited the best performance, with its accuracy ranging from 84% to 95%. Following GPT-4, the Microsoft Bing model demonstrated block system accuracies between 68% and 91%. The accuracy of GPT-3.5 and Google Bard was comparable in this study, with GPT-3.5 achieving between 53% and 85%, and Google Bard ranging from 53% to 72%.

XSL•FO
RenderX

Table . Number of correct answers stratified by block system (n=282)

Topic	Correct answers, n (%)			
	GPT-3.5	GPT-4	Microsoft Bing	Google Bard
General principles (n=91)	77 (85)	82 (90)	77 (85)	66 (73)
Hematopoietic system (n=16)	8 (50)	14 (88)	13 (81)	11 (69)
Nervous system (n=23)	13 (57)	21 (91)	19 (83)	12 (52)
Skin and connective tissue (n=11)	8 (73)	10 (91)	9 (82)	5 (46)
Musculoskeletal system (n=12)	9 (75)	11 (92)	10 (83)	8 (67)
Respiratory system (n=22)	12 (55)	17 (77)	19 (86)	13 (59)
Cardiovascular system (n=25)	14 (56)	21 (84)	20 (80)	16 (64)
Gastrointestinal system (n=21)	15 (71)	20 (95)	18 (86)	14 (67)
Urinary system (n=19)	10 (53)	17 (90)	13 (68)	10 (53)
Reproductive system (n=20)	13 (65)	18 (90)	18 (90)	12 (60)
Endocrine system (n=22)	16 (73)	20 (91)	20 (91)	13 (59)

Table 4 illustrates the question characteristics associated with correct answers. There was a correlation between the difficulty index and the accuracy in all 4 models, with the strongest association observed in the GPT-4 model (OR 90.13, 95% CI 4.30 - 1887.54; *P*=.004). This was followed by GPT-3.5, which had an OR of 28.03 (95% CI 4.68 - 167.98; *P*<.001). Microsoft Bing and Google Bard demonstrated similar correlations with correct answers, with ORs of 18.9 (95% CI 1.84 - 195.42; *P*=.01) and 18.73 (95% CI 3.12 - 112.45; *P*=.001), respectively, as shown in Table 4. There was no statistically significant correlation between the accuracy of GPT-3.5, GPT-4, and Bing and question length, negative word questions, clinical case scenario questions, or the discrimination index.

Table . Multivariable logistic regression analysis showing question characteristics associated with correct answer of large language model artificial intelligence (n=282).

Variable	GPT-3.5		GPT-4		Microsoft Bing		Google Bard	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Question length (word)	0.99 (0.97 - 1.00)	.07	1.00 (0.98 - 1.02)	.96	1.00 (0.98 - 1.02)	.94	0.98 (0.97 - 1.00)	.02
Negative word question	0.55 (0.22 - 1.35)	.19	0.44 (0.15 - 1.30)	.14	0.46 (0.18 - 1.22)	.12	0.26 (0.10 - 0.69)	.007
Case scenario question	0.94 (0.50 - 1.77)	.85	1.57 (0.63 - 3.93)	.34	0.94 (0.43 - 2.04)	.87	0.56 (0.30 - 1.07)	.08
Difficulty index (*p*)	28.03 (4.68 - 167.98)	<.001	90.13 (4.30 - 1887.54)	.004	18.9 (1.84 - 195.42)	.01	18.73 (3.12 - 112.45)	.001
Discrimination index (*r*)	2.80 (0.34 - 23.32)	.34	4.85 (0.20 - 116.54)	.33	9.66 (0.67 - 140.06)	.10	9.31 (1.02 - 84.68)	.048

[a]OR: odds ratio.

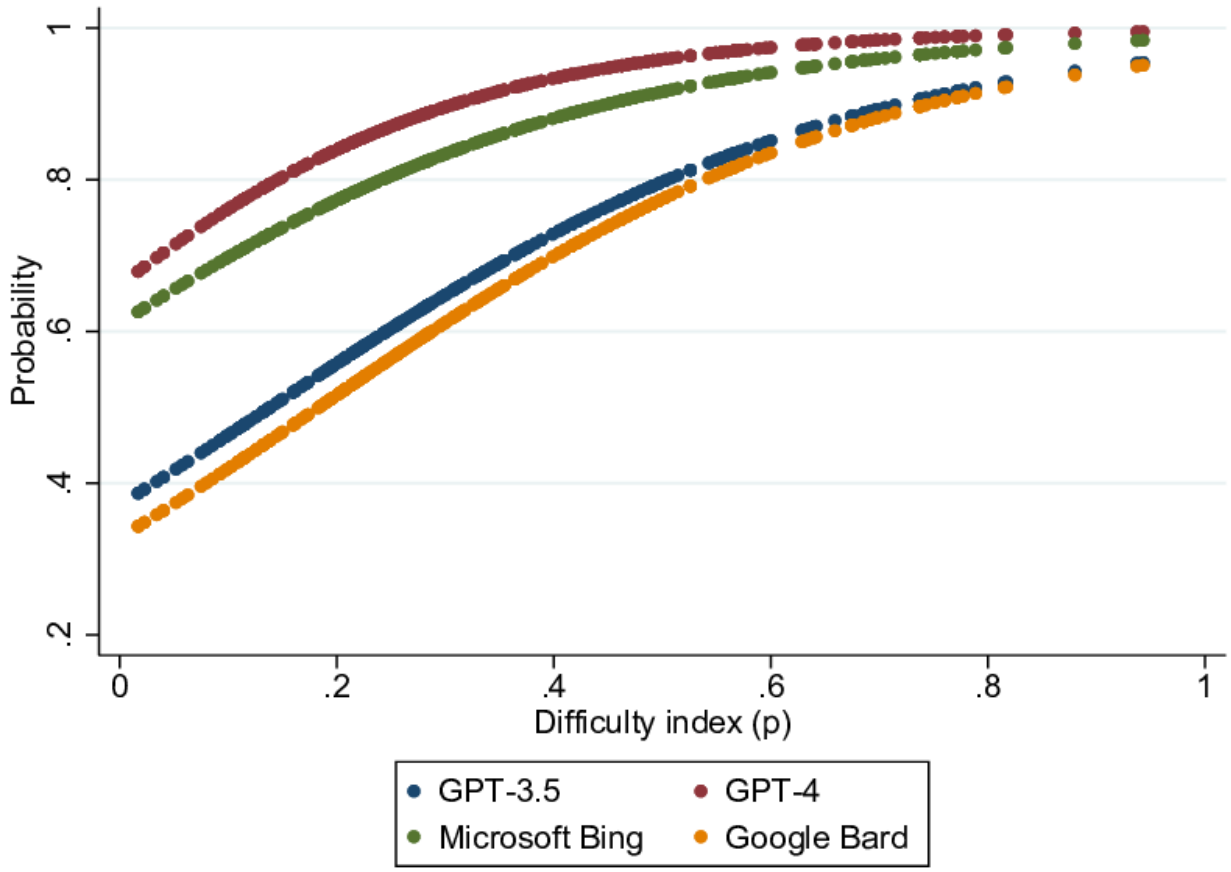
On the other hand, for Google Bard, longer questions had a higher OR, of 0.98 (95% CI 0.97 - 1.00; *P*=.02), for the model to provide the correct answer than shorter questions. The negative-word questions were less likely to be answered correctly by the model, with an OR of 0.26 (95% CI 0.10 - 0.69; *P*=.007), compared to those without negative words. Furthermore, questions with a higher discrimination index were more likely to be correctly answered with statistical significance by the model, with an OR of 9.31 (95% CI 1.02 - 84.68, *P*=.048), as compared to those with a lower discrimination index. No statistically significant correlation was observed between the accuracy of the AIs in answering clinical case scenario questions, as presented in Table 4.

The correlation between the difficulty index and the estimated accuracy of the various AI models, analyzed with binary logistic regression, is shown in Figure 3. The GPT-4 model consistently demonstrated the highest accuracy across all levels of question difficulty index (Figure 3). Google Bard, on the other hand, had the lowest estimated accuracy. The accuracy of the various

XSL•FO
RenderX

LLMs improved as the difficulty index increased, indicating that these models performed better on easier questions.

Figure 3. Accuracy of various artificial intelligence models estimated based on difficulty index.



Discussion

Accuracy of the LLMs on Basic Medical Science Examinations

This study compared the accuracy of LLMs in answering questions from a basic medical science examination related to the National Medical Licensing Examination, finding that GPT-4 had the highest accuracy, at 89.07%, and Google Bard had the lowest accuracy, at 63.83%, when tasked with answering questions in this context. The most frequently studied AI models were GPT-3.5 and GPT-4.

These results align with the 2023 findings of Yanagita et al [15], who used questions from the National Medical Licensing Examination in Japan, administered by the Japanese Ministry of Health, Labour and Welfare. When inputting Japanese questions into the prompt, they reported an accuracy for GPT-4 of 81.5%, significantly higher than GPT-3.5's accuracy of 42.8%, with GPT-4 surpassing the National Medical Licensing Examination passing standard of 72%.

Our results are similar to those of the study conducted by Gilson et al [6] in 2023, which found that the performance of GPT-3.5 on AMBOSS-Step1 and NBME-Free-Step1 was 44% and 64.4%, respectively. Flores-Cohaila et al [16] conducted a study on the accuracy of LLMs on the Peruvian National Licensing Medical Examination and discovered that GPT-4 had 86% accuracy, following by GPT-3.5 at 77%, with moderately difficult to difficult questions being associated with incorrect answers (the OR for GPT-3.5 was 6.6, 95% CI 2.73 - 15.95; for GPT-4, the OR was 33.23, 95% CI 4.3 - 257.12).

A literature review from China (Wang et al [17]) evaluated the performance of GPT-3.5 and GPT-4 on the China National Medical Licensing Examination and reported 56% and 84% accuracy for GPT-3.5 and GPT-4, respectively, demonstrating GPT-4's superiority over GPT-3.5 in terms of accuracy on basic medical science examinations.

The accuracy of GPT-4 and GPT-3.5 is influenced by the variety within the question dataset. This results in diverse outcomes across different countries, changing according to the environmental context, difficulty level of the examination, and the proportion of subcomponents within the examination question sets, which may vary from one country to another. Consequently, the estimated accuracy of AI models for each dataset is not constant.

Difficulty Index and the LLMs' Accuracy

In this study, we identified factors correlated with the accuracy of AI models in answering questions. We found that for every model, the difficulty index was associated with correctly answering questions. Moreover, across all models, there was a tendency to answer questions correctly as the difficulty index increased (indicating easier questions). Specifically, GPT-4

XSL•FO
RenderX

demonstrated the highest OR at 90.13 (95% CI 4.30 - 1887.54; *P*=.004), followed by GPT-3.5, with an OR of 28.03 (95% CI 4.68 - 167.98; *P*<.001).

This result aligns with findings from Antaki et al [18] showing that question difficulty was the most predictive factor of GPT-3.5's answer accuracy (likelihood ratio 24.05; *P*<.001) and that GPT-4 was more accurate than GPT-3.5. The current research reveals the accuracy of AI models in answering questions across various disciplines, particularly studies focusing on the renowned GPT-3.5 model. However, this study focused on the relationship between the difficulty index, derived from human examination observations, and the accuracy of every simple-to-access LLM that is widely used. There was also a variation in accuracy among all models, with GPT-4 being the most accurate, and there was an obvious correlation with the difficulty index for each model, indicating that easier questions had higher accuracy.

The Implication of LLMs for Medical Education

This study's findings hold significant implications for medical education, particularly regarding the use of LLMs such as GPT-4, Microsoft Bing, GPT-3.5, and Google Bard as educational tools [19]. There are 3 major ways that this study's findings can be applied to augment traditional study methods.

First, enhancing study efficiency: the high accuracy rates of LLMs, especially GPT-4, in answering medical examination questions suggest their utility as effective study aids. By providing immediate and accurate answers with explanations, these models can help students identify areas of weakness and reinforce their learning more efficiently than traditional study methods alone.

Second, supplementing traditional education methods: LLMs can act as supplementary tools in medical education, alongside lectures, textbooks, and clinical scenarios. Integrating LLMs into the curriculum provides students with an additional resource for study and review to enhance the overall educational experience.

Last, preparing for licensing examinations: given the study's focus on medical licensing examinations, LLMs could play a crucial role in preparing students for these critical assessments. The ability of LLMs to accurately answer examination questions, such as those tackled by GPT-4, and explain reasoning processes can assist students in better preparing for the format and content of licensing exams.

LLMs may have a negative impact on medical education. Excessive dependence on LLMs might impede the development of independent critical thinking skills. Students may become reliant on the model's suggestions instead of developing their own reasoning processes. LLMs can sometimes provide incorrect, incomplete, or biased information [20,21]. This can interfere with the development of critical appraisal skills, leading students to accept inaccurate information, which may hinder their critical thinking and medical reasoning abilities [22]. Additionally, reduced peer and mentor interaction can hinder the development of professional judgment, depriving students of diverse perspectives and collaborative problem-solving experiences.

To maximize the benefits while minimizing the negative impact of incorporating LLMs into medical education [23], 4 strategies can be considered. First, structured use: LLMs can be incorporated as supplementary tools in a structured curriculum rather than as primary sources of information. Second, critical appraisal training: the importance of critically appraising information provided by LLMs should be emphasized, and students should be taught how to cross-reference and validate information. Third, independent thought should be encouraged: environments should be fostered that encourage independent thinking and problem-solving, using LLMs to support (not replace) these processes. Fourth, monitoring and evaluation: the impact of LLMs on students' learning and reasoning skills should be assessed, and educational approaches should be adjusted based on these assessments.

Limitations

One significant limitation of this study is the LLMs' ability to accurately respond to complex medical examination questions. Moreover, despite GPT-4's high performance, the study's focus on a single culturally and geographically specific medical licensing examination (Thailand Step 1 National Medical Licensing Examination) may limit the generalizability of the findings to other medical examinations and educational contexts. The exclusion of questions containing images and those not in English restricted the comprehensiveness of the assessment, considering the importance of questions on visual diagnostics. Updates to LLMs can significantly affect their accuracy, leading to a potential increase in the capabilities of the models over time. Furthermore, different LLMs can respond differently to different prompts. They can generate different answers across independent sessions, even with identical prompts. Therefore, a sensitivity analysis of the accuracy of the LLMs' responses should be conducted with a variety of prompt and session settings.

Conclusion

Our results show a significant variation in performance among different LLMs, with the most accurate model being GPT-4. This study has shed light on the role of LLMs as supplementary tools in medical education, as well as the need for more research to increase the generalizability of the findings to different educational settings. We advocate for the ongoing development and modification of LLMs to match the unique demands of medical education internationally, which has important implications for the future integration of AI in medical training and test preparation.

Acknowledgments

This research received funding support from the Unit of Teaching and Learning Development and Research, Suranaree University of Technology.

XSL•FO
RenderX

Authors' Contributions

NK was responsible for the entire project and prepared the initial draft of the manuscript. JP conceptualized the project. GB and NP contributed by proofreading and editing the manuscript. All authors participated in interpreting the results and in preparing the final version of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Descriptive statistics of item analysis for each block system.
[DOCX File, 18 KB - mededu_v11i1e58898_app1.docx]

References

1. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science 2015 Jul 17;349(6245):255-260. [doi: 10.1126/science.aaa8415] [Medline: 26185243]
2. Chen L, Chen P, Lin Z. Artificial intelligence in education: a review. IEEE Access 2020;8:75264-75278. [doi: 10.1109/ACCESS.2020.2988510]
3. Novak LL, Russell RG, Garvey K, et al. Clinical use of artificial intelligence requires AI-capable organizations. JAMIA Open 2023 Jul;6(2):ooad028. [doi: 10.1093/jamiaopen/ooad028] [Medline: 37152469]
4. Sedaghat S. Early applications of ChatGPT in medical practice, education and research. Clin Med (Lond) 2023 May;23(3):278-279. [doi: 10.7861/clinmed.2023-0078] [Medline: 37085182]
5. Liu PR, Lu L, Zhang JY, Huo TT, Liu SX, Ye ZW. Application of artificial intelligence in medicine: an overview. Curr Med Sci 2021 Dec;41(6):1105-1115. [doi: 10.1007/s11596-021-2474-3] [Medline: 34874486]
6. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 8;9:e45312. [doi: 10.2196/45312] [Medline: 36753318]
7. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Dig Health 2023 Feb;2(2):e0000198. [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
8. Tolsgaard MG, Pusic MV, Sebok-Syer SS, et al. The fundamentals of artificial intelligence in medical education research: AMEE Guide No. 156. Med Teach 2023 Jun;45(6):565-573. [doi: 10.1080/0142159X.2023.2180340] [Medline: 36862064]
9. Subramani M, Jaleel I, Krishna Mohan S. Evaluating the performance of ChatGPT in medical physiology university examination of phase I MBBS. Adv Physiol Educ 2023 Jun 1;47(2):270-271. [doi: 10.1152/advan.00036.2023] [Medline: 36971685]
10. Oh N, Choi GS, Lee WY. ChatGPT goes to the operating room: evaluating GPT-4 performance and its potential in surgical education and training in the era of large language models. Ann Surg Treat Res 2023 May;104(5):269-273. [doi: 10.4174/astr.2023.104.5.269] [Medline: 37179699]
11. Weng TL, Wang YM, Chang S, Chen TJ, Hwang SJ. ChatGPT failed Taiwan's Family Medicine Board Exam. J Chin Med Assoc 2023 Aug 1;86(8):762-766. [doi: 10.1097/JCMA.0000000000000946] [Medline: 37294147]
12. Kumar D, Jaipurkar R, Shekhar A, Sikri G, Srinivas V. Item analysis of multiple choice questions: a quality assurance test for an assessment tool. Med J Armed Forces India 2021 Feb;77(Suppl 1):S85-S89. [doi: 10.1016/j.mjafi.2020.11.007] [Medline: 33612937]
13. Dhanvijay AKD, Dhokane N, Balgote S, et al. The effect of a one-day workshop on the quality of framing multiple choice questions in physiology in a medical college in India. Cureus 2023 Aug;15(8):e44049. [doi: 10.7759/cureus.44049] [Medline: 37746478]
14. Bhattacherjee S, Mukherjee A, Bhandari K, Rout AJ. Evaluation of multiple-choice questions by item analysis, from an online internal assessment of 6th semester medical students in a rural medical college, West Bengal. Ind J Community Med 2022;47(1):92-95. [doi: 10.4103/ijcm.ijcm_1156_21] [Medline: 35368481]
15. Yanagita Y, Yokokawa D, Uchida S, Tawara J, Ikusaka M. Accuracy of ChatGPT on medical questions in the National Medical Licensing Examination in Japan: evaluation study. JMIR Form Res 2023 Oct 13;7:e48023. [doi: 10.2196/48023] [Medline: 37831496]
16. Flores-Cohaila JA, García-Vicente A, Vizcarra-Jiménez SF, et al. Performance of ChatGPT on the Peruvian National Licensing Medical Examination: cross-sectional study. JMIR Med Educ 2023 Sep 28;9:e48039. [doi: 10.2196/48039] [Medline: 37768724]
17. Wang H, Wu W, Dou Z, He L, Yang L. Performance and exploration of ChatGPT in medical examination, records and education in Chinese: pave the way for medical AI. Int J Med Inform 2023 Sep;177:105173. [doi: 10.1016/j.ijmedinf.2023.105173] [Medline: 37549499]

XSL•FO
RenderX

18. Antaki F, Touma S, Milad D, El-Khoury J, Duval R. Evaluating the performance of ChatGPT in ophthalmology: an analysis of its successes and shortcomings. Ophthalmol Sci 2023 Dec;3(4):100324. [doi: 10.1016/j.xops.2023.100324] [Medline: 37334036]
19. Abd-Alrazaq A, AlSaad R, Alhuwail D, et al. Large language models in medical education: opportunities, challenges, and future directions. JMIR Med Educ 2023 Jun 1;9:e48291. [doi: 10.2196/48291] [Medline: 37261894]
20. Lucas HC, Upperman JS, Robinson JR. A systematic review of large language models and their implications in medical education. Med Educ 2024 Nov;58(11):1276-1285. [doi: 10.1111/medu.15402] [Medline: 38639098]
21. Alowais SA, Alghamdi SS, Alsuhebany N, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ 2023 Sep 22;23(1):689. [doi: 10.1186/s12909-023-04698-z] [Medline: 37740191]
22. Safranek CW, Sidamon-Eristoff AE, Gilson A, Chartash D. The role of large language models in medical education: applications and implications. JMIR Med Educ 2023 Aug 14;9:e50945. [doi: 10.2196/50945] [Medline: 37578830]
23. Shimizu I, Kasai H, Shikino K, et al. Developing medical education curriculum reform strategies to address the impact of generative AI: qualitative study. JMIR Med Educ 2023 Nov 30;9:e53466. [doi: 10.2196/53466] [Medline: 38032695]

Abbreviations

AI: artificial intelligence
LLM: large language model
USMLE: United States Medical Licensing Examination

Edited by B Lesselroth; submitted 29.03.24; peer-reviewed by AKD Dhanvijay, C Rose, GP Privitera, P Shingru, W Xu; revised version received 22.05.24; accepted 04.12.24; published 13.01.25.

Please cite as:
Kaewboonlert N, Poontananggul J, Pongsuwan N, Bhakdisongkhram G
Factors Associated With the Accuracy of Large Language Models in Basic Medical Science Examinations: Cross-Sectional Study
JMIR Med Educ 2025;11:e58898
URL: https://mededu.jmir.org/2025/1/e58898
doi:10.2196/58898

© Naritsaret Kaewboonlert, Jiraphon Poontananggul, Natthipong Pongsuwan, Gun Bhakdisongkhram. Originally published in JMIR Medical Education (https://mededu.jmir.org), 13.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Enhancing Medical Student Engagement Through Cinematic Clinical Narratives: Multimodal Generative AI–Based Mixed Methods Study

Tyler Bland, PhD

Department of Medical Education, University of Idaho, 875 Perimeter Drive MS 4061, WWAMI Medical Education, Moscow, ID, United States

Corresponding Author:
Tyler Bland, PhD
Department of Medical Education, University of Idaho, 875 Perimeter Drive MS 4061, WWAMI Medical Education, Moscow, ID, United States

Abstract

Background: Medical students often struggle to engage with and retain complex pharmacology topics during their preclinical education. Traditional teaching methods can lead to passive learning and poor long-term retention of critical concepts.

Objective: This study aims to enhance the teaching of clinical pharmacology in medical school by using a multimodal generative artificial intelligence (genAI) approach to create compelling, cinematic clinical narratives (CCNs).

Methods: We transformed a standard clinical case into an engaging, interactive multimedia experience called "Shattered Slippers." This CCN used various genAI tools for content creation: GPT-4 for developing the storyline, Leonardo.ai and Stable Diffusion for generating images, Eleven Labs for creating audio narrations, and Suno for composing a theme song. The CCN integrated narrative styles and pop culture references to enhance student engagement. It was applied in teaching first-year medical students about immune system pharmacology. Student responses were assessed through the Situational Interest Survey for Multimedia and examination performance. The target audience comprised first-year medical students (n=40), with 18 responding to the Situational Interest Survey for Multimedia survey (n=18).

Results: The study revealed a marked preference for the genAI-enhanced CCNs over traditional teaching methods. Key findings include the majority of surveyed students preferring the CCN over traditional clinical cases (14/18), as well as high average scores for triggered situational interest (mean 4.58, SD 0.53), maintained interest (mean 4.40, SD 0.53), maintained-feeling interest (mean 4.38, SD 0.51), and maintained-value interest (mean 4.42, SD 0.54). Students achieved an average score of 88% on examination questions related to the CCN material, indicating successful learning and retention. Qualitative feedback highlighted increased engagement, improved recall, and appreciation for the narrative style and pop culture references.

Conclusions: This study demonstrates the potential of using a multimodal genAI-driven approach to create CCNs in medical education. The "Shattered Slippers" case effectively enhanced student engagement and promoted knowledge retention in complex pharmacological topics. This innovative method suggests a novel direction for curriculum development that could improve learning outcomes and student satisfaction in medical education. Future research should explore the long-term retention of knowledge and the applicability of learned material in clinical settings, as well as the potential for broader implementation of this approach across various medical education contexts.

(JMIR Med Educ 2025;11:e63865) doi:10.2196/63865

KEYWORDS

artificial intelligence; cinematic clinical narratives; cinemeducation; medical education; narrative learning; AI; medical student; pharmacology; preclinical education; long-term retention; AI tools; GPT-4; image; applicability

Introduction

Background

Student and trainee engagement is a critical factor in medical education, influencing outcomes such as academic achievement, overall well-being, satisfaction, and reduced burnout [1,2]. High levels of engagement have been linked to increased motivation and better learning experiences, as active participation encourages deeper understanding and application of complex material [3]. In contrast, traditional lecture-based learning often results in passive absorption of information, limiting student engagement and negatively affecting the ability to interact meaningfully with content [4]. To address this, we developed a cinematic clinical narrative (CCN), an interactive multimedia learning experience designed to enhance student engagement by integrating cinematic storytelling and narrative-based learning techniques. This method builds upon the principles of

XSL•FO
RenderX

cinemeducation, a teaching approach that uses film to create emotional connections and foster active learning [5]. By using generative artificial intelligence (genAI) tools, we have further enhanced the learning experience and decreased the barrier to entry for instructors, making it more immersive and adaptable to current educational needs. GenAI has been recognized as a transformative tool in reshaping medical education, offering new opportunities for interactive, technology-driven learning environments that promote active student engagement [6,7].

The target audience for our CCN comprises first-year medical students learning pharmacology related to the immune system. Medical students often face a knowledge gap in understanding complex pharmacological interactions and the intricacies of immune responses largely due to the difficulty of the material [8,9]. Furthermore, there is speculated to be a skill gap in medical and other professional health science students in applying theoretical knowledge to clinical scenarios [10] and the real problem of burnout due to many factors, one of which is the large amount of knowledge required to retain in a short amount of time [11]. The CCN aims to address these issues by enhancing comprehension, clinical application skills, and empathy toward patients with autoimmune diseases.

The CCN used a unique instructional approach by merging cinemeducation [5] with multiple genAI platforms, tailored for first-year medical students in pharmacology. This method addresses the challenge of enhancing engagement and knowledge retention in complex subjects such as immune system pharmacology. Unlike traditional didactic teaching, our approach, supported by others advocating for innovative teaching strategies, uses storytelling to deepen understanding and empathy [12-14]. Use of genAI in medical training, particularly in personalizing learning experiences and competencies for genAI-based tools, is also a current area of active research [15,16]. This aligns with other researchers who highlight the importance of interactive and engaging content in medical education [17]. Our project also leverages the effectiveness of narrative-based learning, which offers an experiential learning environment over conventional teaching methods and is more accurate to real-world situations [18].

Medical students often struggle to engage with and retain complex pharmacological concepts, especially in preclinical education, where traditional teaching methods can lead to passive learning and poor knowledge retention. To address this challenge, we developed and implemented a novel instructional approach, CCNs, which leverages multimodal genAI tools to create immersive, engaging learning experiences. The aim of this study is to evaluate the effectiveness of these genAI-enhanced CCNs in increasing student engagement, interest, and knowledge retention in medical pharmacology concepts. We tested this intervention by assessing student interest using the Situational Interest Survey for Multimedia (SIS-M) and measuring examination performance on content covered by the CCNs. We hypothesize that students exposed to CCNs will report higher levels of engagement compared with traditional case-based learning and have passing examination grades on questions related to the CCN.

Theoretical Framework

The instructional method in the CCN uses contemporary educational theories emphasizing active, learner-centered approaches. Drawing inspiration from the Constructivist Learning Theory, which advocates for knowledge construction through experience [19], our approach uses an adaptation of cinemeducation to create an immersive learning environment [5]. This also aligns with Mayer's Cognitive Theory of Multimedia Learning, which suggests that learning is enhanced through multimodal presentations [20]. Furthermore, our multimodal use of various genAI platforms for content development is informed by the Technological Pedagogical Content Knowledge (TPACK) framework [21], ensuring an effective integration of technology in teaching. This methodology responds to identified needs in medical education for more engaging and effective teaching strategies, bridging theory and practice in a novel and impactful way.

Methods

Participants and CCN Design Overview

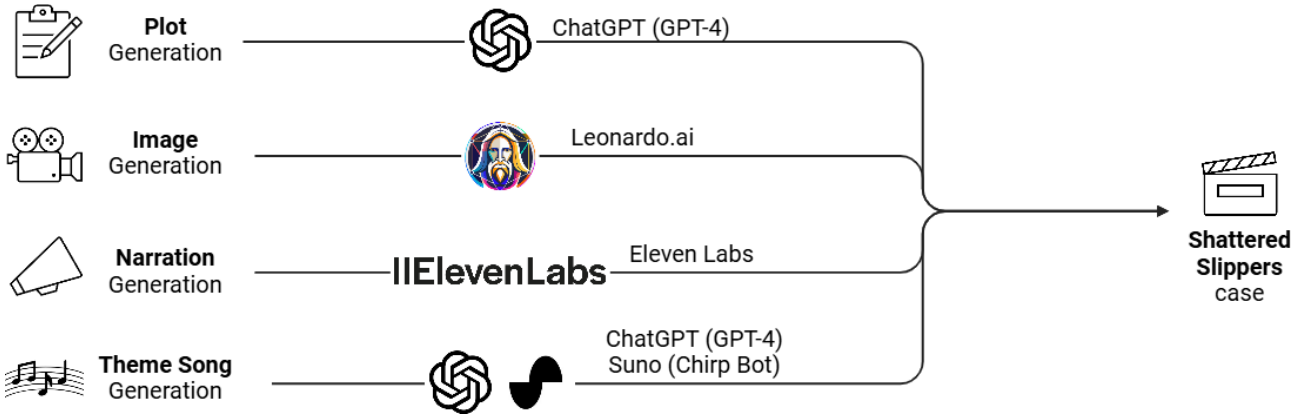
This study was conducted at the University of Idaho WWAMI Medical Education Program, which is part of a collaborative University of Washington School of Medicine program serving Washington, Wyoming, Alaska, Montana, and Idaho. The WWAMI program provides medical education to students across these states, offering them the opportunity to complete their first 2 preclinical years of medical school in their home states before transitioning to clinical training. The target learners for this study were first-year medical students in the WWAMI program enrolled in a 6-week foundational infections and immunity course, which included topics covering immune system pharmacology. Students in this course attend pharmacology lectures that culminate in clinical cases, allowing them to apply their newly acquired knowledge of medications to real-world patient scenarios.

We decided to reimagine one of these cases into "Shattered Slippers," a CCN that was presented as a fictional sequel to the movie "Another Cinderella Story" (Multimedia Appendices 1 and 2). This fictional sequel features the star from the original movie, Selena Gomez, which was purposeful, given her real-life battle with lupus and her experience receiving a kidney transplant. This choice not only provides a strong thematic link connecting the CCN to the source material but also serves to humanize and demystify the conditions under study.

The development of "Shattered Slippers" used a suite of genAI platforms to create an immersive and engaging learning experience (Figure 1). The plot was crafted using GPT-4, known for its language understanding and generation capabilities. For visual imagery, Leonardo.ai and Stable Diffusion were used to generate high-quality, contextually relevant images. Narration was produced using Eleven Labs, ensuring a coherent and captivating storytelling experience. Furthermore, the theme song, integral to setting the tone of the educational module, was composed using the combined efforts of GPT-4 and Suno.

XSL•FO
RenderX

Figure 1. Multimodal generative artificial intelligence (genAI) case generation approach. Each portion of the case used a different genAI platform for material generation. These included ChatGPT (GPT-4), Leonardo.ai, Eleven Labs, and Suno.



These artificial intelligence (AI)–generated materials were all integrated into 2 PowerPoint presentations. Part I of the CCN was presented at the end of a 1-hour pharmacology lecture on immunomodulatory drugs with specific focus on nonsteroidal anti-inflammatory drugs, glucocorticoids, and innate immune system inhibitors. Part II of the CCN was presented 4 weeks later at the end of a 1-hour pharmacology lecture on immunomodulatory and transplant drugs with specific focus on cytokine inhibitors, cytotoxic drugs, and antimetabolites. Both lectures were presented in-person with >90% of students attending both lectures. The combined CCN is provided as a supplemental file (Multimedia Appendix 2).

At the conclusion of the course, students were informed about Selena Gomez's actual medical journey. This revelation effectively bridged the gap between the fictional narrative of "Shattered Slippers" and real-world medical scenarios, thereby enhancing the educational impact and relevance of the clinical cases discussed.

Plot Development

The process of developing the plot for "Shattered Slippers" began with a reimagining of a clinical case initially presented in the first-year medical school curriculum. This original case centered around a ballerina struggling with rheumatoid arthritis, where students were tasked with diagnosing the sources of her pain and inflammation and selecting suitable immunomodulatory medications.

Using ChatGPT (GPT-4) [22], a large language model (LLM), we transformed this clinical scenario into a compelling narrative for "Shattered Slippers." The sequential steps of the medical case were input into GPT-4, with instructions to adapt these into a fictional storyline (Figure 2 and Multimedia Appendix 3). To enhance thematic resonance and real-world connection, the ballerina's diagnosis in the plot was altered from rheumatoid arthritis to lupus, mirroring the real-life medical condition of Selena Gomez, who stars in the CCN.

Further expanding the scope of the narrative, the plot incorporated a kidney transplant storyline. This addition served a dual purpose. First, it aligned with the second lecture on immunoregulatory pharmacology focusing on organ transplant pharmacology. Second, it resonated with Selena Gomez's personal medical history, as she has undergone a kidney transplant. This incorporation not only ensured continuity with the educational objectives of the course but also added depth and authenticity to the fictional narrative, making it more engaging and relatable for the students.

XSL•FO
RenderX

Figure 2. Excerpt of plot generation. The initial prompt in the conversation covered the development of a separate CCN. Prompt engineering techniques in this initial prompt included Persona Prompting [23,24] and a modified version of Zero-Shot CoT [25]. Excerpts of the first prompt and output related to the Shattered Slippers CCN are provided. CCN: cinematic clinical narrative; CoT: Chain of Thought; LLM: large language model.

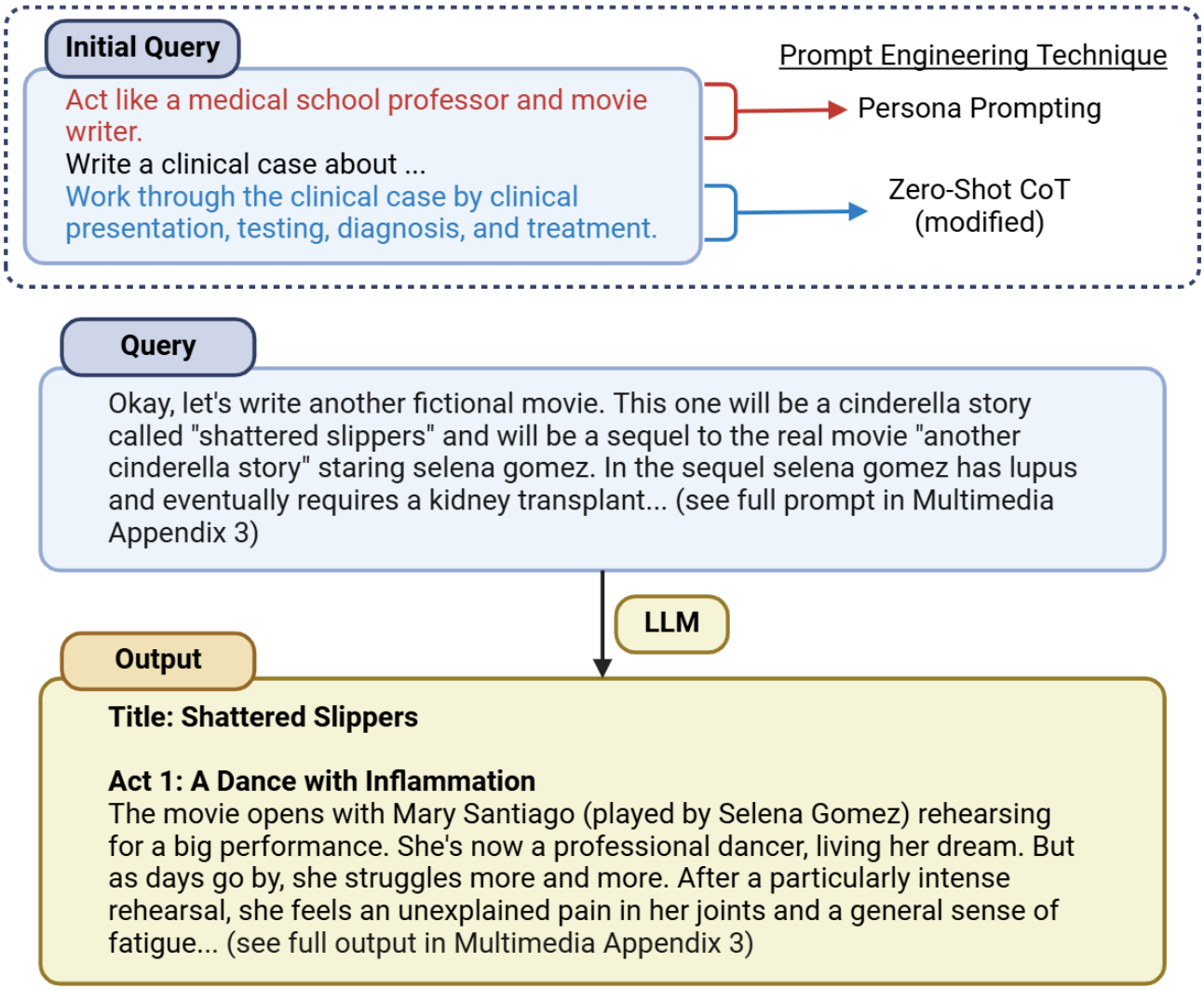


Image Generation

In order to create a more immersive educational experience, fictional images were integrated into the “Shattered Slippers” case study. These images were generated using the Leonardo.ai platform [26], which harnesses the capabilities of the Stable Diffusion XL image–generating technology (Figure 3 and Multimedia Appendix 4).

In an effort to maintain transparency and distinguish between real and AI-generated content, all images depicting real people were marked with an “AI-generated image” icon. This icon, chosen for its symbolic significance, is the spinning top from the movie “Inception.” The selection of this particular icon was purposeful; it serves as a metaphor for the increasingly blurred lines between reality and artificial constructs, mirroring the movie’s thematic exploration of distinguishing reality from illusion. This concept was explained to the students prior to their engagement with the case, setting the stage for a thoughtful consideration of the role and impact of genAI in content creation. This iconography not only helped in identifying AI-generated images but also subtly underscored the advanced capabilities of genAI in creating hyperrealistic images.

XSL•FO
RenderX

Figure 3. Artificial intelligence (AI)-generated image of Selena Gomez singing with Justin Bieber. The prompt used was "adult Selena Gomez and Justin Bieber singing together." The spinning top in the bottom right corner was added as a watermark to denote an AI-generated image. Generated with Leonardo.ai.

Narration Generation

Enhancing the immersive aspect of the CCN, an audio narration was incorporated to accompany the text on the PowerPoint slides. This element was designed to emulate the experience of listening to a movie narrator, thereby bringing the story of "Shattered Slippers" to life in an auditory format. To achieve this, the finalized script of the plot was submitted to the Eleven Labs platform [27], which specializes in converting text into lifelike audio narration (Multimedia Appendix 5).

Each of these audio narrations were incorporated into their corresponding PowerPoint slides. As each slide was presented during the course, the audio narration played automatically, further synchronizing the visual and auditory elements of the learning experience. This integration of audio narration with the visual content not only enriched the storytelling aspect of the module but also supported diverse learning styles, facilitating a more engaging and multisensory educational experience for the students.

Theme Song Generation

Although not directly educational, a theme song for "Shattered Slippers" was created to complete the immersive experience. The inclusion of a theme song aimed to add an additional layer of engagement and context to the fictional movie, contributing to a more comprehensive and cinematic learning environment.

The lyrics for the theme song were generated using GPT-4 [22]. Following the lyric generation, Suno Chirp Bot, a genAI tool for music composition [28], was used to create the melody and vocals for the theme song. This genAI-driven process allowed for a harmonious blend of lyrics and music, resulting in a fully rendered theme song (Multimedia Appendix 6).

Once completed, the theme song was embedded into the PowerPoint presentation. This musical addition served as a capstone to the multisensory educational module, further enriching the student's experience by providing a unique auditory element that complemented the visual and textual components of "Shattered Slippers."

Data Collection

The "Shattered Slippers" CCN was integrated into 2 distinct pharmacology lectures, both of which focused on medications used in immune system modulation. The target audience for this CCN was a class of 40 first-year medical students (n=40). This approach aimed not only to enrich their understanding of immunomodulatory pharmacology but also to engage them in a unique and memorable learning experience.

To evaluate student interest in the CCN as an educational tool, at the conclusion of the course, students were invited to participate in a feedback process using the SIS-M [29-31] (Table 1) of which 18 students responded (n=18). The SIS-M was developed by Dr Tonia Dousay, a professor in instructional design and educational technology, to assess various constructs of situational interest in multimedia-based learning environments. Originally created for the educational field, the SIS-M focuses on adult learners and measures constructs such

XSL•FO
RenderX

as triggered situational interest (initial engagement with multimedia), maintained interest, and value interest (perceived usefulness of the content). The survey was originally used to evaluate the effectiveness of multimedia in promoting engagement and motivation in higher education and adult learning settings [29,30] and has recently been used in medical education research [31], making it an appropriate tool for assessing learner engagement in this study. This survey was used to capture their views and opinions on the "Shattered Slippers" case, providing insights into student engagement, interest, and the overall impact of the CCN on their learning experience. The survey includes items to rank on a 1 - 5 scale (1=strongly disagree, 5=strongly agree), a question asking for preference of clinical case format, and an open-ended question asking, "Why do you think this is your preference." The CHERRIES report for this survey is supplied (Multimedia Appendix 7).

Table . SIS items.

SIS[a] type	Survey item
SI-triggered	The multimedia presentation was interesting.
SI-triggered	The multimedia presentation grabbed my attention.
SI-triggered	The multimedia presentation was often entertaining.
SI-triggered	The multimedia presentation was so exciting, it was easy to pay attention.
SI-maintained-feeling	What I learned in the multimedia presentation is fascinating to me.
SI-maintained-feeling	I am excited about what I learned in the multimedia presentation.
SI-maintained-feeling	I like what I learned in the multimedia presentation.
SI-maintained-feeling	I found the information in the multimedia presentation interesting.
SI-maintained-value	What I studied in the multimedia presentation is useful for me to know.
SI-maintained-value	The things I studied in the multimedia presentation are important to me.
SI-maintained-value	What I learned in the multimedia presentation can be applied to my job.
SI-maintained-value	I learned valuable things in the multimedia presentation.

[a]SIS: Situational Interest Survey.

Data Analysis

The research team used Microsoft Excel for the analysis of the SIS-M survey results. The average class pharmacology examination grades (n=40) from questions covered by the "Shattered Slippers" case study (n=2) were analyzed for achievement data. These included a multiple-choice question, selected by the course lead (not the study author) from a pool of questions that tested pharmacology content covered in each pharmacology lecture. The questions were administered during the students' weekly examinations, scheduled for the week immediately following the presentation of the material. Importantly, these questions were modeled after USMLE-style step 1 board questions, which assess students' ability to apply their pharmacological knowledge in a clinical context. Using this format provides a rigorous and standardized measure of student understanding of the material, ensuring that the assessment reflects the type of knowledge and critical thinking required for success on future board examinations.

The SIS-M survey's analysis focused on various dimensions of situational interest: triggered interest, maintained-value (MV), maintained interest, and maintained-feeling (MF). Thematic analysis was conducted using ChatGPT (GPT4o and o1-preview) and Claude 3.5 Sonnet. This involved generating initial codes and identifying themes, followed by the researcher combining and refining these themes for overlap and relevancy between the 3 LLMs [31]. Prompt engineering techniques used included Persona Prompting [23,24], Zero-Shot Chain of Thought (CoT) [25], and Self-Criticism [32]. The Zero-Shot Chain of Thought prompting was not used with the ChatGPT o1-preview model, as it has built-in Tree-of-Thought functionality in every output. The initial prompt was the following:

> *Act like a brilliant medical education researcher. I am doing a study on a Cinematic Clinical Narrative (CCN) which is an educational tool that combines clinical case studies with storytelling techniques typically seen in movies or TV shows. By embedding medical information within a compelling fictional storyline, CCNs help medical students retain complex medical concepts in an engaging, memorable way. The CCN in the study was called "Shattered Slippers," was a fictional sequel to the movie "Another Cinderella Story," and stars Selena Gomez. It covered the topics of immunomodulatory medications for treating lupus, and kidney transplants. I surveyed the participants on their preference of the CCN over traditional clinical cases and asked them to explain their preference. Please perform a thematic analysis on the below participant responses marked between <response> </response>. Let's work this out in a step by step way to be sure we have the right answer.*
>
> *<response>*
>
> *Participant responses here*
>
> *</response>*

XSL•FO
RenderX

This was then followed by the following Self-Criticism prompt: “Please reflect on your previous answer for any errors.”

Ethical Considerations

This educational research was approved as exempt by the institutional review board of the University of Idaho (21-223). As the CCN incorporated references to real celebrities and included AI-generated images of actual people, we consulted legal counsel to ensure compliance. The counsel advised that, given the educational context and the clear labeling of images as AI-generated rather than real, the usage was permissible. Furthermore, we end the CCN with a brief description of the real-life health struggles of the celebrities, which is all public information. However, since this remains a legally gray area, we recommend exercising caution in future projects that use similar techniques. The SIS-M was conducted anonymously to ensure the confidentiality of participants’ responses. No identifying information was collected, allowing students to provide honest feedback without concern for personal attribution.

Results

The quantitative assessment of the “Shattered Slippers” CCN using the SIS-M is summarized in Table 2. The results indicated high levels in participants’ interest with the “Shattered Slippers” CCN, with the majority of students (14/18) indicating a preference for the CCN over traditionally presented clinical cases, only 1 student preferring the traditional approach, and 3 expressing no preference (Table 3).

XSL•FO
RenderX

Table . Situational Interest Survey for Multimedia results (N=18): scores.

Question	Minimum[a]	Maximum[a]	Mean[a]	SD	Variance
The Shattered Slippers case was interesting.	4.00	5.00	4.61	0.49	0.24
The Shattered Slippers case grabbed my attention.	4.00	5.00	4.72	0.45	0.20
The Shattered Slippers case was often entertaining.	3.00	5.00	4.67	0.58	0.33
The Shattered Slippers case was so exciting, it was easy to pay attention.	3.00	5.00	4.33	0.58	0.33
What I learned from the Shattered Slippers case is fascinating to me.	4.00	5.00	4.39	0.49	0.24
I am excited about what I learned from the Shattered Slippers case.	4.00	5.00	4.39	0.49	0.24
I like what I learned from the Shattered Slippers case.	3.00	5.00	4.39	0.59	0.35
I found the information from the Shattered Slippers case interesting.	4.00	5.00	4.33	0.47	0.22
What I studied in the Shattered Slippers case is useful for me to.	4.00	5.00	4.50	0.50	0.25
The things I studied in the Shattered Slippers case are important to me.	3.00	5.00	4.28	0.56	0.31
What I learned from the Shattered Slippers case can be applied to my major/career.	3.00	5.00	4.44	0.60	0.36
I learned valuable things from the Shattered Slippers case.	4.00	5.00	4.44	0.50	0.25

[a]Rated on a 5-point scale (1=Strongly disagree, 5=Strongly agree).

Table . Situational Interest Survey for Multimedia results (N=18): preferences for case type.

Which case type do you prefer?	Count
Traditional case studies	1
Shattered Slippers case study	14
No preference	3

Participants indicated a high average triggered situational interest in the CCN (mean 4.58, SD 0.53), as well as high maintained interest scores indicated by the students (mean 4.40, SD 0.53).

The results for MF interest indicated high MF in students receiving the CCN (mean 4.38, SD 0.51). A feeling of educational value by the participants was supported by high scores for MV interest (mean 4.42, SD 0.54).

XSL•FO
RenderX

Bridging quantitative data with qualitative insights, the survey conducted among participants also provided an open-ended question for students to reflect on their opinion of the CCN. Thematic analysis of the responses revealed the following:

- *Enhanced engagement through storytelling and entertainment*: The combination of storytelling and entertainment in the CCN heightened student engagement, making the learning process more enjoyable and effective compared with traditional methods.
- *Improved memorability and recall of medical concepts*: The CCN's engaging narrative and multimedia elements enhanced memory retention, making complex medical information more accessible and memorable.
- *Relatability through pop culture and personal connection*: Leveraging familiar pop culture icons such as Selena Gomez helped students form a personal connection with the material, enhancing engagement and motivation to learn.
- *Preference for interactive and detailed learning*: Some students value interactive learning environments and detailed information, suggesting that while the CCN is engaging, it could be further enhanced by incorporating active learning elements and comprehensive content.
- *Suggestions for improvement*: Attention to technical elements, such as the use of genAI voice narration, could improve the overall effectiveness and reception of the CCN.

The thematic analysis reveals that the CCN "Shattered Slippers" was preferred over traditional case studies due to its engaging storytelling, enhanced memorability, and relatability through pop culture references. While students appreciated the innovative approach, some expressed a desire for more interactive learning methods and provided suggestions for technical improvements. Incorporating these insights can further refine the CCN as a valuable tool in medical education.

In addition to the survey feedback from the SIS-M, the success of the "Shattered Slippers" CCN was further demonstrated academically. Students displayed strong comprehension and knowledge of the material covered, achieving an average score of 88% on examination questions pertaining to the case study content. This high performance underscores the effectiveness of the CCN as a teaching tool, suggesting that it may also be useful in promoting academic performance as well as student preference and interest.

Discussion

Principal Findings

The "Shattered Slippers" CCN supports the pedagogical value of integrating innovative genAI-driven methods and culturally resonant themes into medical education. Our study shows the capacity of this approach to not only enhance student interest but also promote their understanding and retention of complex subject matter. Furthermore, it adds very little to no extra time to the lecture material, as it basically reskins the existing material into a more cinematic experience. This is particularly important, as many new active learning teaching methodologies either extend the amount of time students spend with the material or cause instructors to remove large amounts of material in order to incorporate novel active learning activities. We considered it ethical to clearly mark AI-generated images of real individuals to avoid confusion but did not deem it necessary to label AI-generated material such as text or audio that was not mimicking a real-world person. As genAI models continue to improve in generating realistic images and cloned voices, it will become increasingly important to label AI-generated materials that mimic real-world individuals to prevent confusion with reality and avoid potential legal issues.

This study shows the importance of engaging students beyond conventional didactic methods, suggesting that the inclusion of elements such as plot development, multimedia, and popular culture can make learning more relatable and impactful. The feedback from the SIS-M supports that this approach can effectively address the initial problem of student disengagement and the need for more effective educational strategies as identified in the introduction.

The process of creating CCNs with genAI tools is highly efficient and cost-effective. Designing the case outline took about a day, while plot and narration generation were completed in seconds using GPT-4 and Eleven Labs. Image and theme song generation took under an hour each, with slight delays due to iterative refinement. Overall, the time investment was minimal compared with traditional methods. The required technical skills are basic, involving familiarity with genAI platforms for text, image, and audio generation and standard project management skills to integrate these elements into a PowerPoint slide deck. In terms of cost, the only expense was a US $20 per month subscription to ChatGPT; other platforms were used on free tiers. This low cost, combined with fast production times, makes migrating to this format highly accessible and efficient for educators, offering significant time and cost savings compared with traditional content creation methods of this caliber.

Future directions of this work will explore how similar immersive educational experiences can be scaled and adapted for diverse student populations and learning environments. The versatility of genAI-enhanced CCNs extends beyond pharmacology, offering potential applications in other areas such as anatomy, pathology, and clinical skills. This pedagogical strategy can be adapted to various medical disciplines, making abstract topics more engaging and accessible to diverse learners. It also asks questions on how educational policies might evolve to integrate this type of AI-generated material into curricula systematically. As genAI becomes more integral to education, policies must address both the ethical use of genAI and the need for genAI literacy among educators and students. Personalized, genAI-driven learning experiences could revolutionize how content is delivered, providing flexibility and tailored learning opportunities. There is an opportunity to explore interdisciplinary collaborations, merging medical education with fields such as AI, storytelling, and multimedia design. These collaborations could further refine educational tools and help bridge the gap between traditional learning and modern health care technologies, fostering genAI literacy in future medical professionals. This promising pilot study shows potential for scalability and broad applicability of genAI-enhanced CCNs. The strategy offers a model for

XSL•FO
RenderX

transforming how complex medical topics are taught, providing a scalable, engaging solution that can be adapted across different medical content areas to meet evolving educational needs.

Limitations

Our project has limitations in terms of cultural adaptability due to its reliance on specific cultural references and celebrity figures, which may not resonate with all audiences. Furthermore, the use of genAI technologies presents challenges in environments with varying levels of technological resources and differing instructor familiarity with these platforms. While the skills required to effectively use genAI can vary depending on the model, these challenges are mitigated by the increasing availability of more user-friendly genAI platforms. These platforms are simplifying AI integration in educational contexts, expanding the potential for their broader application. For instance, prompt engineering, which is crucial for optimizing output from LLMs, is becoming less essential with newer versions such as ChatGPT's o1-preview model, which incorporates many of these strategies into the system itself. This reduces the need for advanced user expertise and lowers the barrier to efficient LLM use.

Another limitation of our study is the process of validity checking for AI-generated content. Although the materials were reviewed by medical professionals, including physicians, PhDs, and PharmDs, to ensure accuracy, the use of genAI introduces potential risks in content reliability, especially as AI-generated content may produce subtle inaccuracies or lack the nuanced context that a human expert might provide. Future implementations of this approach would benefit from a formalized validation process to ensure that the clinical and educational integrity of AI-generated materials is maintained.

The evaluation methodology, focusing on immediate reactions via the SIS-M, provides a single time point of the resource's impact but does not capture the longevity of knowledge retention or the applicability of the learned material in clinical settings. Furthermore, the study included a limited sample size, with only 18 respondents to the SIS-M survey, which may not provide a comprehensive view of the broader student population. Future research could explore longitudinal studies to measure the lasting educational benefits of such methodologies with a larger participant population.

Furthermore, our study lacked a control or comparison group, a common challenge in medical education research. All students in the study were exposed only to the CCN case, and without a traditional case-based learning comparison, it is difficult to isolate the exact impact of the CCN on student performance. While we acknowledge that a control group could provide valuable insights, the integration of such comparisons is often logistically difficult in medical school settings. Future studies could address this by designing more controlled experimental conditions or through the use of quasi-experimental designs to better understand the differential effects of various educational interventions on learning.

Conclusions

The "Shattered Slippers" CCN demonstrates the effectiveness of combining cinemeducation with genAI in medical education. This approach enhanced student engagement, promoted knowledge retention, and offered a novel perspective on complex pharmacological clinical cases. The application and positive student feedback suggest that this multimodal genAI approach to educational content creation has potential for broader application in medical education. Our project also highlights the need for continuous innovation and adaptation in teaching methodologies to meet the evolving demands of health care education. Future research and development in this area could further transform medical education, making it more engaging, effective, and aligned with modern technological advancements.

Acknowledgments

The author would like to extend his heartfelt gratitude to his students for their participation and invaluable contributions to the "Shattered Slippers" project. Their engagement and feedback were essential in shaping this educational endeavor and its success.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Shattered Slippers: cinematic clinical narrative.
[PDF File, 878 KB - mededu_v11i1e63865_app1.pdf]

Multimedia Appendix 2
Shattered Slippers full presentation.
[MP4 File, 134196 KB - mededu_v11i1e63865_app2.mp4]

Multimedia Appendix 3
ChatGPT plot generation.
[DOCX File, 20 KB - mededu_v11i1e63865_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
Leonardo.ai image generation.
[DOCX File, 3053 KB - mededu_v11i1e63865_app4.docx]

Multimedia Appendix 5
Eleven Labs narration generation and audio clips.
[DOCX File, 13 KB - mededu_v11i1e63865_app5.docx]

Multimedia Appendix 6
ChatGPT and Suno Chirp Bot theme song generation and audio clip.
[DOCX File, 15 KB - mededu_v11i1e63865_app6.docx]

Multimedia Appendix 7
Situations Interest Survey of Multimedia CHERRIES (Checklist for Reporting Results of Internet E-Surveys) report.
[DOCX File, 15 KB - mededu_v11i1e63865_app7.docx]

References

1. Kassab SE, Taylor D, Hamdy H. Student engagement in health professions education: AMEE Guide No. 152. Med Teach 2023 Sep;45(9):949-965. [doi: 10.1080/0142159X.2022.2137018] [Medline: 36306374]
2. Kassab SE, Al-Eraky M, El-Sayed W, Hamdy H, Schmidt H. Measurement of student engagement in health professions education: a review of literature. BMC Med Educ 2023 May 20;23(1):354. [doi: 10.1186/s12909-023-04344-8] [Medline: 37210491]
3. Hodges LC. Student engagement in active learning classes. In: Active Learning in College Science: The Case for Evidence-Based Practice: Springer Cham; 2020:27-41. [doi: 10.1007/978-3-030-33600-4_3]
4. Miller CJ, McNear J, Metz MJ. A comparison of traditional and engaging lecture methods in a large, professional-level course. Adv Physiol Educ 2013 Dec;37(4):347-355. [doi: 10.1152/advan.00050.2013] [Medline: 24292912]
5. Rueb M, Siebeck M, Rehfuess EA, Pfadenhauer LM. Cinemeducation in medicine: a mixed methods study on students' motivations and benefits. BMC Med Educ 2022 Mar 12;22(1):172. [doi: 10.1186/s12909-022-03240-x] [Medline: 35279156]
6. Narayanan S, Ramakrishnan R, Durairaj E, Das A. Artificial intelligence revolutionizing the field of medical education. Cureus 2023 Nov;15(11):e49604. [doi: 10.7759/cureus.49604] [Medline: 38161821]
7. Arango-Ibanez JP, Posso-Nuñez JA, Díaz-Solórzano JP, Cruz-Suárez G. Evidence-based learning strategies in medicine using AI. JMIR Med Educ 2024 May 24;10(1):e54507. [doi: 10.2196/54507] [Medline: 38801706]
8. Karim MY. Using clinical cases to restore basic science immunology knowledge in physicians and senior medical students. Front Immunol 2020;11:1756. [doi: 10.3389/fimmu.2020.01756] [Medline: 32973743]
9. Haidaris CG, Frelinger JG. Inoculating a new generation: immunology in medical education. Front Immunol 2019;10:2548. [doi: 10.3389/fimmu.2019.02548] [Medline: 31749807]
10. Ousey K, Gallagher P. The theory-practice relationship in nursing: a debate. Nurse Educ Pract 2007 Jul;7(4):199-205. [doi: 10.1016/j.nepr.2007.02.001] [Medline: 17689445]
11. Bhugra D, Molodynski A, Ventriglio A. Well-being and burnout in medical students. Ind Psychiatry J 2021;30(2):193-197. [doi: 10.4103/ipj.ipj_224_21] [Medline: 35017800]
12. Bhargava M, Naik PR, Hegde P, Navya N, Sachith M, Vineetha S. Integrating narrative medicine through story-telling: a feasibility study in a community medicine curriculum for undergraduate and postgraduate students. Cureus 2023 Jul;15(7):e41851. [doi: 10.7759/cureus.41851] [Medline: 37581154]
13. Kagawa Y, Ishikawa H, Son D, et al. Using patient storytelling to improve medical students' empathy in Japan: a pre-post study. BMC Med Educ 2023 Jan 27;23(1):67. [doi: 10.1186/s12909-023-04054-1] [Medline: 36707818]
14. Liao HC, Wang YH. Storytelling in medical education: narrative medicine as a resource for interdisciplinary collaboration. Int J Environ Res Public Health 2020 Feb 11;17(4):1135. [doi: 10.3390/ijerph17041135] [Medline: 32053911]
15. Cooper A, Rodman A. AI and medical education—a 21st-century Pandora's box. N Engl J Med 2023 Aug 3;389(5):385-387. [doi: 10.1056/NEJMp2304993] [Medline: 37522417]
16. Russell RG, Lovett Novak L, Patel M, et al. Competencies for the use of artificial intelligence-based tools by health care professionals. Acad Med 2023 Mar 1;98(3):348-356. [doi: 10.1097/ACM.0000000000004963] [Medline: 36731054]
17. Tuma F. The use of educational technology for interactive teaching in lectures. Ann Med Surg (Lond) 2021 Feb;62:231-235. [doi: 10.1016/j.amsu.2021.01.051] [Medline: 33537136]
18. Charon R. Narrative medicine: a model for empathy, reflection, profession, and trust. JAMA 2001 Oct;286(15):1897-1902. [doi: 10.1001/JAMA.286.15.1897] [Medline: 11597295]
19. Dennick R. Constructivism: reflections on twenty five years teaching the constructivist approach in medical education. Int J Med Educ 2016 Jun 25;7:200-205. [doi: 10.5116/ijme.5763.de11] [Medline: 27344115]

XSL•FO
RenderX

20. Mayer RE. Cognitive theory of multimedia learning. In: The Cambridge Handbook of Multimedia Learning: Cambridge University Press; 2012:31-48. [doi: 10.1017/CBO9780511816819.004]
21. Niess ML. Technological Pedagogical Content Knowledge (TPACK) Framework for K-12 Teacher Preparation: Emerging Research and Opportunities: Information Science Reference; 2016:1-173. [doi: 10.4018/978-1-5225-1621-7]
22. ChatGPT. OpenAI. URL: https://openai.com/chatgpt [accessed 2023-12-19]
23. Wang J, Liu Z, Zhao L, et al. Review of large vision models and visual prompt engineering. Meta Rad 2023 Nov;1(3):100047. [doi: 10.1016/j.metrad.2023.100047]
24. White J, Fu O, Hays S, et al. A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv. Preprint posted online on Feb 21, 2023 URL: https://arxiv.org/abs/2302.11382v1 [accessed 2024-09-23] [doi: 10.48550/arXiv.2302.11382]
25. Kojima T, Gu SS, Reid M, Matsuo Y, Iwasawa Y. Large language models are zero-shot reasoners. Presented at: 36th Conference on Neural Information Processing Systems (NeurIPS 2022); Nov 28 to Dec 9, 2022; New Orleans, Louisiana URL: https://arxiv.org/abs/2205.11916v4 [accessed 2023-09-23]
26. Leonardo.ai. URL: https://leonardo.ai/ [accessed 2023-12-19]
27. Text to speech & AI voice generator. ElevenLabs. URL: https://elevenlabs.io/ [accessed 2023-12-19]
28. SUNO AI. URL: https://www.suno.ai/ [accessed 2023-12-19]
29. Dousay TA. Effects of redundancy and modality on the situational interest of adult learners in multimedia learning. Education Tech Research Dev 2016 Dec;64(6):1251-1271. [doi: 10.1007/s11423-016-9456-3]
30. Dousay TA, Trujillo NP. An examination of gender and situational interest in multimedia learning environments. Brit J Educ Tech 2019 Mar;50(2):876-887. [doi: 10.1111/bjet.12610]
31. Bland T, Guo M, Dousay TA. Multimedia design for learner interest and achievement: a visual guide to pharmacology. BMC Med Educ 2024 Feb 5;24(1):113. [doi: 10.1186/s12909-024-05077-y] [Medline: 38317141]
32. Huang J, Gu S, Hou L, et al. Large language models can self-improve. Presented at: EMNLP 2023—Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing; Dec 6-10, 2023; Singapore p. 1051-1068. [doi: 10.18653/v1/2023.emnlp-main.67]

Abbreviations

AI: artificial intelligence
CCN: cinematic clinical narrative
genAI: generative artificial intelligence
LLM: large language model
MF: maintained-feeling
MV: maintained-value
SIS-M: Situational Interest Survey for Multimedia
TPACK: Technological Pedagogical Content Knowledge

Edited by B Lesselroth, G Eysenbach; submitted 01.07.24; peer-reviewed by M Guo, MI Knopp; revised version received 26.09.24; accepted 07.11.24; published 06.01.25.

Please cite as:
Bland T
Enhancing Medical Student Engagement Through Cinematic Clinical Narratives: Multimodal Generative AI–Based Mixed Methods Study
JMIR Med Educ 2025;11:e63865
URL: https://mededu.jmir.org/2025/1/e63865
doi:10.2196/63865

© Tyler Bland. Originally published in JMIR Medical Education (https://mededu.jmir.org), 6.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Perceptions and Earliest Experiences of Medical Students and Faculty With ChatGPT in Medical Education: Qualitative Study

Noura Abouammoh[1,2], MBBS, PhD; Khalid Alhasan[1,3,4], MBBS; Fadi Aljamaan[1,5], MD; Rupesh Raina[6], MD; Khalid H Malki[1,7], PhD; Ibraheem Altamimi[1], MBBS; Ruaim Muaygil[1,8], MBBS; Hayfaa Wahabi[2,9], MD, PhD; Amr Jamal[1,2,9], MBBS; Ali Alhaboob[1,3], MBBS; Rasha Assad Assiri[10], MBBS; Jaffar A Al-Tawfiq[11,12,13], MBBS; Ayman Al-Eyadhy[1,3], MD; Mona Soliman[1,8], MBBS, PhD; Mohamad-Hani Temsah[1,3,9], MD

[1]College of Medicine, King Saud University, Riyadh, Saudi Arabia

[2]Department of Family and Community Medicine, King Saud University Medical City, King Saud University, Riyadh, Saudi Arabia

[3]Pediatric Department, King Saud University Medical City, King Saud University, Riyadh, Saudi Arabia

[4]Department of Kidney and Pancreas Transplant, Organ Transplant Center of Excellence, King Faisal Specialist Hospital & Research Centre, Riyadh, Saudi Arabia

[5]Critical Care Department, King Saud University Medical City, King Saud University, Riyadh, Saudi Arabia

[6]Department of Nephrology, Cleveland Clinic Akron General and Akron Children Hospital, Akron, OH, United States

[7]Research Chair of Voice, Swallowing, and Communication Disorders, Department of Otolaryngology, College of Medicine, King Saud University, Riyadh, Saudi Arabia

[8]Medical Education Department, King Saud University Medical City, King Saud University, Riyadh, Saudi Arabia

[9]Evidence-Based Health Care & Knowledge Translation Research Chair, Family & Community Medicine Department, College of Medicine, King Saud University, Riyadh, Saudi Arabia

[10]Department of Basic Medical Sciences, College of Medicine, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

[11]Specialty Internal Medicine and Quality Department, Johns Hopkins Aramco Healthcare, Dhahran, Saudi Arabia

[12]Infectious Disease Division, Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, United States

[13]Infectious Disease Division, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States

Corresponding Author:
Mohamad-Hani Temsah, MD
Pediatric Department
King Saud University Medical City
King Saud University
King Abdullah Road
Riyadh, 11424
Saudi Arabia
Phone: 966 114692002
Email: mtemsah@ksu.edu.sa

Abstract

Background: With the rapid development of artificial intelligence technologies, there is a growing interest in the potential use of artificial intelligence–based tools like ChatGPT in medical education. However, there is limited research on the initial perceptions and experiences of faculty and students with ChatGPT, particularly in Saudi Arabia.

Objective: This study aimed to explore the earliest knowledge, perceived benefits, concerns, and limitations of using ChatGPT in medical education among faculty and students at a leading Saudi Arabian university.

Methods: A qualitative exploratory study was conducted in April 2023, involving focused meetings with medical faculty and students with varying levels of ChatGPT experience. A thematic analysis was used to identify key themes and subthemes emerging from the discussions.

Results: Participants demonstrated good knowledge of ChatGPT and its functions. The main themes were perceptions of ChatGPT use, potential benefits, and concerns about ChatGPT in research and medical education. The perceived benefits included collecting and summarizing information and saving time and effort. However, concerns and limitations centered around the potential lack of critical thinking in the information provided, the ambiguity of references, limitations of access, trust in the output of ChatGPT, and ethical concerns.

XSL•FO
RenderX

Conclusions: This study provides valuable insights into the perceptions and experiences of medical faculty and students regarding the use of newly introduced large language models like ChatGPT in medical education. While the benefits of ChatGPT were recognized, participants also expressed concerns and limitations requiring further studies for effective integration into medical education, exploring the impact of ChatGPT on learning outcomes, student and faculty satisfaction, and the development of critical thinking skills.

(JMIR Med Educ 2025;11:e63400) doi:10.2196/63400

KEYWORDS

ChatGPT; medical education; Saudi Arabia; perceptions; knowledge; medical students; faculty; chatbot; qualitative study; artificial intelligence; AI; AI-based tools; universities; thematic analysis; learning; satisfaction

Introduction

Artificial intelligence (AI) is a computer-based technology invented as a digital system to imitate and aid human intellect and skills. The wide use of AI technology is changing the medical field considerably, aiming for more efficient patient management. Medical education is one of the vital domains of health care practice, in which AI has a promising contribution by providing an alternative and efficient means of information access, achieving teaching goals and skills development. As an example, the integration of AI in simulated surgical skills learning showed comparable results compared to remote expert instructions [1], but it led to unintended outcomes in another study, which affected trainees' efficiency metrics on the cost of safer skills development [2]. Case-based learning is another potential field harnessing AI technology in medical education, which has shown promising results [3]. AI technology has also been used in teaching clinical examination skills, such as breast self-examination, yielding mixed results: high levels of student satisfaction paired with increased anxiety [4]. Such AI-driven interventions will be leading health care practice in the future, such as the introduction of machine-based surgical treatment with robotic surgery, which has effectively promoted diagnostic accuracy, achieving treatment goals and saving health care professionals' workload [5-7]. AI technology integration in medical education and medical research will not only contribute to patients' care but also improve if not revolutionize the medical education system [8,9]. All these changes of AI integration into the medical practice need to be accompanied by evolution in the medical teaching and training curricula [8,9], facing significant interest among educators and researchers recently on AI's rapid involvement in medical education [10-13].

One of the pioneer and popular generative AI-based tools is ChatGPT, a language model developed by OpenAI that uses natural language processing to generate humanlike responses to queries, with many potential applications in health care [14,15]. ChatGPT was perceived by health care workers to positively impact the future of health care systems by 76.7% in a recent study [16]. However, little is known specifically about the perceptions and experiences of faculty and students or trainees against the use of ChatGPT in the context of medical education within Saudi Arabia.

The health care sector in Saudi Arabia is experiencing dramatic growth and reformatting, with a strong emphasis on prioritizing medical education and digitizing the health systems. Therefore, using AI technology in the health care system is a promising strategy for substantial investments in medical, nursing, and other specialized educational disciplines [17]. As medical education evolves, the use of AI-based tools like ChatGPT could potentially transform the way medical education is delivered [18]. Literature has a gap in assessing the perceptions and attitudes of medical education stakeholders regarding integrating AI technology in curricula, clinical teaching, and simulation skills development. Most literature addressed specific AI technology adoption in medical practice or certain educational domains but did not assess it collectively in multiple domains related to medical education. Therefore, it is crucial to explore the medical faculty staff and students' knowledge, perceived benefits, concerns, and limitations of ChatGPT application in medical education.

This qualitative study seeks to explore the perception on the use of newly introduced AI chatbots, like ChatGPT3.5, in medical education from the perspective of faculty and medical students. By deepening our understanding of faculty and students' knowledge about ChatGPT and its applications in medical education, this study identifies both the facilitators and barriers to its use. The research offers valuable preliminary insights into the acceptance of AI-based tools in medical education and informs the development of effective strategies for integrating such tools within medical education systems in Saudi Arabia and similar contexts, as more AI models evolve.

Methods

Study Design

This study was conducted using a focus group technique at the College of Medicine, King Saud University, a leading university in Saudi Arabia [19]. The study included faculty and students from different levels.

The study aims to preliminarily explore and understand participants' perceptions of ChatGPT, a newly introduced large language model. A qualitative methodology was chosen, as it is well suited to exploring experiences, meanings, and perspectives from participants' viewpoints [20-22]. Examining the perceptions of both faculty and students enables a comparative head-to-head analysis of their viewpoints. Qualitative methodology provides a deep explanation of different viewpoints participants may have about ChatGPT use in medical education. It can also allow the authors to propose probing questions to understand and explore users' perceptions. Although individual interviews would elicit a more detailed picture of an issue, focus group discussion was used, as the aim

XSL•FO
RenderX

of the study is to explore different viewpoints using participants' dynamics and thought sharing to enrich the discussion [23]. Data source triangulation was applied to support the trustworthiness of the findings and allow prelude comparison.

Participants were recruited from the College of Medicine through purposive sampling. As a small number of faculty and students used ChatGPT at the time of data collection, a purposive sample was applied. A student was asked to announce the need to interview students who have ever used ChatGPT. Another announcement to faculty was made, and an invitation was sent to random faculties from 3 departments who use or want to share their ideas about ChatGPT in medical education. The sample included 6 medical faculty members (2 associate professors and 4 professors) and 6 medical students (2 second year, 2 third year, 1 fourth year, and 1 fifth year). Two focus group discussions were conducted in April 2023 on the Zoom platform (Zoom Video Communications), one with faculty members and the other with students, and each group consisted of 6 participants. The discussions were conducted in English language as preferred by the participants. Two of the authors (NA and MHT) served as moderators, and each discussion lasted for approximately 1 hour.

Using the Zoom platform in data collection facilitated gathering participants at the same time after working hours. As the team acknowledged that nonverbal cues may not be detected as participants refrained from opening their cameras, follow-up questions and probing were used to minimize subjectivity in understanding participants' responses. Not all participants knew each other; hence, the setting was more private to freely share opposing views.

A topic guide was prepared by the author (NA) to cover aspects such as participants' familiarity with ChatGPT, its uses, facilitators, and limiting factors of its incorporation in medical education. Probing and follow-up questions were allowed depending on participants' responses. The themes were saturated at that time after the second interview possibly due to the limited experience of participants in the early stages of ChatGPT launch. Thematic analysis was used to analyze the data using a priori themes and allowing new themes to emerge from the data [24]. The discussions were transcribed using Zoom's automatic transcription feature. This feature had the advantage of identifying the name that the participants chose for themselves in the discussion and linking it with the speaker.

The transcripts were revised and read multiple times to identify patterns and themes that emerged from the data. A coding framework was developed from the data by each coder (NA and MHT) and applied using NVivo software (version 12; QSR International) [25]. Themes were identified and refined through an iterative process of coding, reviewing, and discussing the data among the research team until a consensus was reached [24]. Initial codes were developed by 2 different authors (NA and MHT) and then, comparison and discussion were made to agree on the coding framework. Coding themes were similar, and no major changes were made in the thematic framework.

Ethical Considerations

This study received ethics approval from the Institutional Review Board at King Saud University (approval 23/0155/IRB). All participants provided verbal informed consent prior to their inclusion in the study, including consent for the audio recording of interviews. Participants were fully informed about the purpose of the study, the voluntary nature of their participation, and their right to withdraw at any time without any consequences. To ensure participant privacy and confidentiality, pseudonyms were assigned, and no identifying information was included in the transcripts or final report. The data were securely stored and accessible only to authorized members of the research team. No compensation was provided to participants for their participation in the study.

Results

Overview

In total, 6 medical faculty staff and 6 medical students with different experiences with ChatGPT participated in the study. Table 1 shows their demographic data. Figure 1 displays the thematic framework used to assess participants' perception of ChatGPT in general and in medical education.

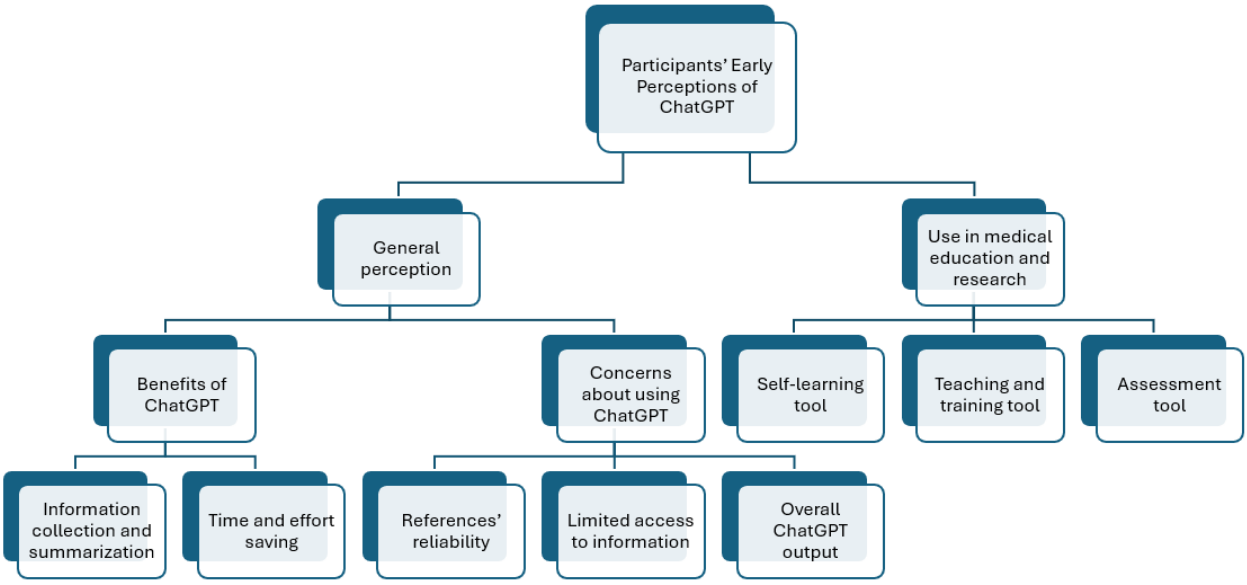
Analysis of the data from the discussion generated two main themes: (1) participants' general perception of ChatGPT and (2) ChatGPT use in medical education and research.

XSL•FO
RenderX

Table 1. Participants position, department, and frequency of ChatGPT use.

Participant code	Position	Department	Age (years) and sex	Using ChatGPT
Participant 1	Faculty	Critical Care Department, College of Medicine	44 and male	Regular user in medical education
Participant 2	Faculty	Ear, Nose, and Throat department, College of Medicine	56 and male	Regular user in medical education
Participant 3	Faculty	Family medicine, College of Medicine	60 and female	Not a user
Participant 4	Faculty	Pediatrics department, College of Medicine	38 and male	Not a user
Participant 5	Faculty	Pediatrics department, College of Medicine	41 and male	Regular user in medical education
Participant 6	Faculty	Medical Education Department, College of Medicine	58 and female	Not a user
Student 1	Student	College of Medicine	20 and male	Regular user for general search
Student 2	Student	College of Medicine	22 and male	Regular user for general search
Student 3	Student	College of Medicine	21 and male	Regular user for general search
Student 4	Student	College of Medicine	21 and male	Regular user for general search
Student 5	Student	College of Medicine	19 and male	Regular user for general search
Student 6	Student	College of Medicine	20 and male	Regular user for general search

Figure 1. Thematic framework of participants' perception on using ChatGPT.



Participants' General Perception of ChatGPT

Overview

All participants expressed good knowledge of ChatGPT's main goal and functions. One participant noted:

> *The idea from this software is that it will chat with you regarding any topic you will ask about...it chats with me in a human like manner, and collect for me the answers from all over resources, and display them.* [Participant 1]

One student described ChatGPT as an "assistant," and others elaborated:

> *Artificial intelligence helps me execute the command that I'm asked to execute.* [Student 2]

> *It's another way of searching for highly accurate information, depending on what I search for and how I search for it.* [Student 5]

Participants were challenged about ChatGPT compared to other traditional search engines: "It is not at similar to Google, even Google started invention of AI application to enrich its platform" (Participant 2). Most participants supported the use of ChatGPT but were not concerned about its information sources.

Benefits of ChatGPT

Two main subthemes emerged from the focus group discussions about the benefits of using ChatGPT.

Collects and Summarizes Information

The majority of the participants believed that searching for information through ChatGPT is more efficient compared to

XSL•FO
RenderX

standard search engines, as the former saves time by summarizing and textualizing the raw information output from the search: “ChatGPT is beautiful in collecting information and presenting it to me in a simplified text that I can easily comprehend” (Participant 2).

Few participants used ChatGPT to review scientific papers or provide ideas for new papers: “I used it to study limitations of studies and the future recommendations for studies I was asked to review...it gives me ideas” (Participant 2).

Students also noted:

> *I find ChatGPT more directive towards what I ask, and to the point, mostly because when I look for something on classical search engine, such as Google...I have to go into some sub web pages which has an answer and look between all the thousands of answers to find one. While ChatGPT will give it to me concisely like this is option A, option B, option C.* [Student 3]

Another faculty added, “It will do the search for me; then even with critically appraise it and give me the final result” (Participant 1). However, one faculty participant was more conservative in her comments about using AI in collecting data and did not perceive the information displayed by ChatGPT as reliable because it lacks the “critical thinking” skill to enable it to reach a final scientific plausible conclusion, “The problem of collecting all the information in one place is that collecting the information and giving it in a nutshell, in one place. This machine is not critically thinking” (Participant 3).

Saves Time and Efforts

Opinions varied in terms of whether using ChatGPT saves time and effort, considering the perceived benefits. One faculty mentioned: “It saves time when I’m stuck in generating exam question” (Participant 2). A student added: “It’s not accurate, but at least it saves me time. This is the most important point” (Student 4).

On the other hand, another faculty participant subtly disclosed her denunciation about the functionality of ChatGPT. She believed that ChatGPT helps partly in performing tasks, but that advantage is contradicted by paying time to verifying and authenticating the ChatGPT output.

> *Me as a researcher. When I search for information I’m putting it together, ChatGPT tries to put it for me. So far, I can’t see it superior to the human mind...It does some of the work for me, but I have to take it with a bunch of salt.* [Participant 3]

Concerns About Using ChatGPT

When the participants discussed the drawbacks of using ChatGPT, they mentioned expressions such as “hallucination” and “blinding euphorically.” The following subthemes emerged as perceived drawbacks of ChatGPT.

References Reliability

While few participants were not sure about the source of ChatGPT information, most participants believed that it is the internet: “It’s the same data retriever as Google” (Participant 3). Other participants had a deeper view: “ChatGPT generated references, and citations have to be taken with caution” (Participant 2).

Faculty experienced situations where they doubted the reference of the information provided by ChatGPT. For example, one faculty noted: “I’m not sure what are the sources used to extract information, even if I ask for references, it might not mention them...or at least it will not volunteer in mentioning them” (Participant 1). While others defended that: “If it doesn’t have access to the reference, it will tell that it doesn’t have access, but if the reference is online, it can refer to that” (Participant 2).

Similarly, a student commented: “It is multitasking, rather than searching for the source of information, it presents the answer and references” (Student 5). One participant pointed that the unreliability of ChatGPT sources supports her view of not relying on ChatGPT.

Limited Access to Information

Several participants acknowledged ChatGPT’s limitation in accessing all available information, driving caution while using ChatGPT:

> *One of the restrictions regarding medical search, it’s restricted to certain resources like PubMed...there are some other medical websites that it cannot access yet.* [Participant 4]

> *We don’t know the algorithm behind the search nor exactly how it looks for information.* [Participant 6]

One of the participants elaborated that ChatGPT is invented by humans; therefore, they may manipulate or restrict its search and output.

> *It’s not free of bias. If I am asking for something morally wrong or illegal. It will not answer because it is constrained. So, it is not fully free from human constraints.* [Student 3]

Some faculty participants raised an ethical concern that may affect the trust in ChatGPT information. One participant explained:

> *Can drug company pay ChatGPT to display answers that are in favour of certain medication? Could ChatGPT be manipulated? ChatGPT inventors are for sure looking for money somehow by anyway!* [Participant 1]

Overall ChatGPT Output

All participants believed that ChatGPT users should not fully trust the information presented and practice caution, while others elaborated that it is ideal for new topics as a jumpstart:

> *I should not take it (information from ChatGPT) for granted; I have to review what’s there, but it gives me a nice idea, very excellent ideas...It sheds the light on some certain angles that I was not looking for.* [Participant 2]

Some participants pointed that trusting ChatGPT output depends on your previous background about the topic:

XSL•FO
RenderX

> *I should have the ability to differentiate between what is reliable and what is not reliable...Myself, I am not well-versed in medical education. For example, I am highly qualified in research, but regarding education I take for granted whatever output from ChatGPT in that regard, while I can filter information regarding research and judge it well.* [Participant 3]

One student agreed:

> *It depends on what I am looking for. Sometimes it's very accurate. Sometimes it's not...But as a human mind I have an idea about what I am looking for, therefore, I can judge if its accurate or doubt the answer.* [Student 4]

All participants agreed that the unfamiliarity of ChatGPT users with its search algorithm enforced the participants' trust issue.

A faculty explained:

> *Do we know the ChatGPT searching methodology? is it scientific methodology? How it extracts the information from the paper, how it appraises it? What are the sources that this engine has access to? All this will augment the reliability of my experience.* [Participant 1]

One participant mentioned that ChatGPT cannot be used for critical thinking in certain contexts; thus, it cannot be fully trusted:

> *It cannot give me what is relevant to me, my community and population and my students...It might be dangerous to put ChatGPT superior to human intellect!* [Participant 3]

Another faculty participant defended the ChatGPT's reliability, noting that it declares its level of expertise and specialty ahead of each information presented:

> *If I ask ChatGPT about something in geology, it will start with "I am not a geologist" and then move on with the dialogue...and it finishes the response by "it is very important to refer to those sources."* [Participant 2]

ChatGPT Use in Medical Education and Research

Participants discussed ChatGPT use in medical education from 3 aspects as discussed below, but in general, they raised concerns about using it without appropriate and dedicated training.

Self-Learning Tool

The majority of faculty participants supported using ChatGPT in the teaching process. A faculty participant commented:

> *Students are no longer enjoying the usual long lectures, or didactic lectures but they enjoy more challenging aspects exploring a new experience, and living it...I think the ChatGPT could be used as a very good trigger for the students to go and read and find out more, discuss among themselves and go explore this with their seniors, with their educators.* [Participant 5]

Another faculty added that it should be used to get an idea about a topic, but further reading is important for students:

> *ChatGPT is like a short fast access to a topic, it helps to get the most important information...they (students) need to read the references.* [Participant 1]

However, another faculty participant raised concerns using ChatGPT for concluding opinions and summing debates:

> *If they (students) use ChatGPT just for recalling information then no problem...But if they want to make inferences, they should not use it.* [Participant 3]

ChatGPT methodology was raised by another faculty participant who did not support using it in learning at all because of its unclear methodology and unverified information sources.

On the other side, the majority of student participants did not support using ChatGPT to obtain information and felt the traditional search engines are more reliable and easier to use:

> *I do not perceive it as a search engine. I don't look up medical information on it, or anything, because I find the classic search engines easier.* [Student 2]

> *I know exactly where the reliable sources are. Then I can take the information from other sources with confidence, and more simple steps.* [Student 1]

Some participants, while supportive of ChatGPT's use in medical practice, emphasized its role in clinical medicine education. They raised concerns about its impact on decision-making, particularly due to ChatGPT's inadequate or unclear strategy for disclosing information sources:

> *If I look at the other search engines for which support medical information, they present like up-to-date information...ChatGPT is very complex, and the methodology and the algorithm it uses is not clear so, it is not a reliable source of information for decision making and for serious information.* [Participant 1]

The issue of updated sources in ChatGPT was also raised:

> *We need to be cautious about using the information...the medical field information is changing very quickly, so we have to be careful about this point.* [Participant 4]

Other participants debated that the information accuracy depends on user searching and prompt engineering skills:

> *Prompt questions will make the difference in getting the response, and I recommend digging into the prompts technology to get more accurate answers, and doing this is important to acquire the right answer.* [Participant 2]

> *You get the response according to the precision of the search.* [Student 1]

Interestingly, a faculty participant raised the concern of students and faculty losing their critical thinking skills if they depend on ChatGPT:

XSL•FO
RenderX

> *It is dangerous...because we are replacing critical thinking. We are prioritizing this thing over human intellect.* [Participant 3]

A student participant who expressed poor research skills was concerned about such skills being affected or even weakened by dependence on ChatGPT in research. In general, students did not support the use of ChatGPT as the primary source of information, especially for new topics, but as a collateral resource.

Teaching and Training Tool

Some participants believed that teaching modalities should change after the introduction of AI technology. They expressed optimism of more teaching methodology shifting from memorization to critical thinking; however, this aim was not perceived achievable through ChatGPT so far:

> *We must invest more in the skills of our medical students and problem-solving critical thinking analysis. These are the areas that is lacking in the ChatGPT, and that we need to focus more on.* [Participant 4]

Faculty participants raised concerns about students' replacement of traditional lectures with AI applications like ChatGPT, which might be risky in general, especially in the current stage of unverified and undedicated AI applications for medical education.

Another concern raised from one faculty regarding the lecturers and trainers:

> *Do our faculty have enough knowledge to use and recommend ChatGPT for their students and instruct them how to use it and get maximum benefit from it?* [Participant 3]

However, all students did not see themselves relying on ChatGPT for learning: "We just need to be familiar on how to use ChatGPT and use it as a tool that supports our search rather than completely relying on it" (Student 2).

Faculty participants differentiated between the needs of postgraduate and undergraduate students and their use of ChatGPT. One faculty (Participant 3) felt that using AI in training postgraduate trainees would be difficult because postgraduate training depends on building skills, while undergraduate depends on memorization as per him.

ChatGPT might be a tool to generate clinical scenarios and draw a framework for discussions with the students:

> *One problem would take weeks from our team and long hours of sitting together and creating the medical problems that we teach in the problem-based learning sessions. So, it would be interesting to see how ChatGPT deals with this.* [Participant 4]

An interesting point mentioned by some faculty is the inability of ChatGPT to teach students human, emotional, and social skills: "Using AI is not designed to help in teaching some skills such as Humanity and the communication, the teamwork" (Participant 4).

Assessment Tool

Most faculty participants mentioned using ChatGPT for academic assessment like examination questions generation:

> *I asked ChatGPT to generate questions for me with scenario and without scenario...it was good to Very good. It's not reaching to excellent level. I have to review and modify.* [Participant 2]

In addition, most faculty participants mentioned using ChatGPT for medical problems, clinical scenarios, and bedside teaching. Some faculty participants raised the idea of using AI applications like ChatGPT to assess the quality and objectives of examinations in order to guide certain questions to assess critical thinking rather than recall knowledge only. Cheating and plagiarism were one of the raised concerns by the faculty during the discussion: "We have to be very careful about cheating and misuse of ChatGPT by our medical students in medical assignments" (Participant 4).

In line with the former comment, one student defended his use of ChatGPT, raising a debatable point of using AI applications for academic assignments is ethical or not:

> *I mainly use it for writing, and then I just review it and edit it...mainly for research or some essays...for example I'd give it some data, and I ask it to write a paragraph that summarizes this data, or an introduction to something for example (Disease X).* [Student 2]

Another student mentioned that his use of ChatGPT in assignments is mainly for summarization. Others use it to collect information resources: "It can make my job way easier. For example, if I have a research assignment to just collect the resources about a topic" (Student 3).

Overall, all participants reached a conclusion of being open-minded and accepting for the ChatGPT intrusion into our lives: "I think it's coming in the near future, and we need to live in the reality to adjust and take the best out of it" (Participant 4).

Discussion

Principal Findings

This paper presents a general snapshot of the faculty and undergraduate medical students' perceptions of ChatGPT and its use in medical education. All participants demonstrated a good understanding of ChatGPT and its functionalities; some described its role as assistive, while others found it as a mere information search tool. Almost all participants were impressed by ChatGPT's ability to provide a concise summary of search results compared to traditional search engines, which is in line with the literature [26]. On the other hand, few students in our study perceived Google as a better tool for learning.

In line with other publications [27], our participants believe that ChatGPT provides a more user-appealing and faster solution for busy users by delivering a summarized, high-caliber textual output. One of the major challenges they mentioned regarding ChatGPT use is its sources of information, which is in line with previously published similar studies showing that students and

XSL•FO
RenderX

faculty are aware of the limitations of ChatGPT that influence its accuracy [26,27]. In a comparative study between platforms, ChatGPT-generated responses were considered to be reliable and beneficial, while others deemed them potentially risky [28]. For example, a study showed that there were concerns about ChatGPT advice regarding antimicrobial stewardship, general course lengths were accurate but the duration varied, and source control was either incorrectly cited as justification for prolonging therapy or ignored entirely [29]. Therefore, ChatGPT output should be dealt with skeptically and selectively, as poor users' baseline knowledge might lead to risky, dangerous, or suboptimal conclusions. Previous literature has shown that in comparison with Google, the majority of the participants tend to doubtfully trust ChatGPT output for reasons related to the novelty of AI and users, lack of understanding of its algorithm, and information sources as studied previously [14,30]. Notably, only 40% of these experts concluded that the perceived value of ChatGPT's responses outperformed those from Google [31].

Therefore, participants tend to trust ChatGPT responses if they have a previous background about the search topic. Participants suggested that while ChatGPT might be helpful in certain aspects of medical education, users should approach the information with caution and apply their medical judgment.

The participants' concerns about ChatGPT output also stemmed partly from the observed phenomenon of references' hallucinations, which raised serious concerns about its reliability and validity [32-34]. In addition, they stressed on the point of ChatGPT's limited access to updated medical literature. A previous study had cautioned authors regarding references generated by ChatGPT [35]. To overcome these limitations, developers should work on expanding the access of ChatGPT's resources, improving its search methodology, and ensuring a more comprehensive and reliable source of information.

Faculty participants explored the potential of ChatGPT in generating examination questions and clinical scenarios, enhancing bedside teaching, and reviewing assessments. Still, they emphasized the need for reviewing and modifying AI-generated content as well as the importance of developing policies and strategies to tackle potential academic misconduct related to ChatGPT use. Previous studies showed ChatGPT's excellent performance as it passed the American Heart Association examination with 84% accuracy, but it failed Taiwan's family medicine examination and fared poorly on the urology self-assessment examination [36-38]. A study concluded that ChatGPT responses were frequently incomplete and sometimes misleading [26]. However, a recent expletory review showed that ChatGPT has a potential impact on medical education, scientific research, and medical writing [14]. Thus, the ChatGPT's generated questions need to be carefully examined and revised especially regarding scientific content. Other research highlighted that generated output in that regard is not highly different among different AI platforms, as the multiple-choice question–based examination performance of ChatGPT was marginally better than that of Google's Bard [39].

Both faculty and students appreciated the time-saving advantage of ChatGPT and its fast access to information. Therefore, faculty used it in preparing lecture materials and examination questions. While students used it in their academic assignments, this mirrors a previous study about ChatGPT perception among students who used it for generating academic content, brainstorming ideas, and writing texts [40,41].

Faculty in our study and previous research raised concerns about students' ChatGPT overuse [13,27,42]. According to our participants, using it by students may interfere with their critical thinking, writing, and information retrieval skills. Faculty highlighted the students' need to critically review and modify the AI-generated content, ensuring it aligns with academic standards and expectations. Banerjee et al [11] reported that postgraduate trainee doctors have an overall positive perception of the impact of AI on clinical training; however, they found that AI will eventually reduce the trainees' clinical judgment and practical skills. In line with that, the faculty participants were concerned about students' self-reliance on AI applications on the cost of traditional teaching methods, which might deprive them from skills best learned in person or group teaching. One study listed the following as disadvantages: lack of originality, inaccurate content, or unknown data sources [14]. It is also uncertain how ChatGPT handles offensive material, false information, or plagiarism [34].

Ethical concerns, such as potential manipulation by pharmaceutical companies, were raised by participants. Maintaining transparency and integrity in AI-generated information is vital to address these concerns. Implementing measures such as third-party audits, strict guidelines, data transparency, and continuous monitoring of ChatGPT's information sources can help ensure the unmanipulated ethical use of ChatGPT in medical education [43-45].

We recommend creating guidelines for students on the appropriate use of AI applications, specifying tasks they should complete independently and the extent to which AI tools can assist. Additionally, we propose incorporating teaching sessions to help students critically evaluate AI-generated outputs. At this early stage of AI adoption [46], group teaching sessions comparing the critical appraisal of medical topics using AI tools versus traditional search methods would be beneficial. We also emphasize leveraging AI applications primarily as advanced search engines and using their summarization capabilities rather than relying entirely on their final outputs.

Participants emphasized the importance of being open-minded and adopting new technologies like AI chatbots including ChatGPT. As AI chatbots could have cultural bias, addressing cultural differences in learning styles is vital [46,47].

The potential implications of using ChatGPT in medical education include improved efficiency, streamlined information gathering, and time-saving benefits. However, future research is needed to explore the impact of AI-based tools on medical education in terms of quality, student and faculty satisfaction, and the development of critical thinking skills. Ongoing research and evaluation are essential to ensure the effective integration of AI-based tools like ChatGPT into medical education while addressing potential concerns and limitations.

In preparation for the future of medical education, educational institutions should be proactive in integrating AI technologies

XSL•FO
RenderX

like ChatGPT into their curricula and teaching methodologies [48,49]. Educators and policy makers need to remain vigilant about reliability concerns and actively take steps to be ready to address the ethical challenges and possibilities arising from the use of AI in health professions education [37,45]. This process should involve regular evaluations, ongoing improvements, and a strong emphasis on maintaining the essential human aspects of medical education, such as critical thinking, communication, and empathy.

Strengths

One of the strengths of this study is the qualitative design, which allowed for an in-depth exploration of participants' experiences, perceptions, and concerns related to the use of ChatGPT in medical education, revealing diverse viewpoints and generating valuable insights into the potential benefits and challenges of integrating ChatGPT into medical education [50]. Moreover, the study involved participants with varying levels of experience with ChatGPT, ensuring a comprehensive understanding of the perspectives of both novices and experienced users. The identification of themes and subthemes has laid a solid foundation for further research and exploration of AI-based tools like ChatGPT in medical education.

Limitations

There are some limitations to our study. The sample size was relatively small, and the participants were primarily drawn from a single institution, which may limit the generalizability of some findings to other medical education settings. The study did not quantitatively assess the impact of ChatGPT on learning outcomes, satisfaction, or other measurable aspects of medical education, which could in the future provide valuable data to supplement the qualitative findings. Additionally, since the study's focus was on understanding the perception of faculty and students, the perspectives of other stakeholders, such as administrators and policy makers, were not captured, and this could be explored in future research [51-53]. Furthermore, the study, which was conducted in the early phase of ChatGPT launching, did not explore the long-term implications and potential changes in perception and use of ChatGPT over time, as participants' experience with the tool may evolve, altering their views on its benefits and limitations [54].

Therefore, future research should incorporate larger and more diverse samples from multiple institutions as well as conduct quantitative studies to measure the impact of ChatGPT on various aspects of medical education in Saudi Arabia specifically and globally. Longitudinal studies could be conducted to assess the changes in perception and use of ChatGPT over time and evaluate the long-term effects of its integration into medical education.

Conclusions

Participants praised the advantages of ChatGPT, such as time-saving and excellent summarizing skills. However, concerns were raised regarding the accuracy and critical appraisal of information provided by ChatGPT and the need to approach the information with caution. ChatGPT-delivered information and cited references' hallucination were concerns seriously raised by participants, which needs urgent assessment and solution in addition to limited access to certain medical databases. This study highlights the need for ongoing research and evaluation to ensure that AI-based tools like ChatGPT are effectively integrated into medical education while addressing potential concerns and limitations. Educators and students must also maintain a strong foundation in critical thinking and judgment. As medical education continues to evolve, the integration of AI technologies like ChatGPT has the potential to transform the way medical education is delivered but must be done with a thoughtful and ethical approach.

Acknowledgments

We have used ChatGPT [55], an artificial intelligence chatbot developed by OpenAI, to improve some readability and language of this work, without replacing researchers' tasks. This was done with human oversight, and the authors then carefully reviewed and edited the generated text. We assure that the authors are ultimately responsible and accountable for the originality, accuracy, and integrity of their work. We would like to acknowledge the efforts in data curation in the focus groups, namely (listed with their permission): Abdulaziz Alomar, Faisal Alomri, Hadi Alhemsi, Homoud Algadhib, and Ibrahim Alhezam. The authors are grateful to Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R148), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Data Availability

The preprint of this research is available [56]. The datasets generated and analyzed during this study are not publicly available due to institutional review board privacy regulations but are available from the corresponding author on reasonable request after obtaining institutional review board approval.

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. Fazlollahi AM, Bakhaidar M, Alsayegh A, Yilmaz R, Winkler-Schwartz A, Mirchi N, et al. Effect of artificial intelligence tutoring vs expert instruction on learning simulated surgical skills among medical students: a randomized clinical trial. JAMA Netw Open 2022;5(2):e2149008 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.49008] [Medline: 35191972]
2. Fazlollahi AM, Yilmaz R, Winkler-Schwartz A, Mirchi N, Ledwos N, Bakhaidar M, et al. AI in surgical curriculum design and unintended outcomes for technical competencies in simulation training. JAMA Netw Open 2023;6(9):e2334658 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.34658] [Medline: 37725373]
3. Zhao C, Xu T, Yao Y, Song Q, Xu B. Comparison of case-based learning using Watson for oncology and traditional method in teaching undergraduate medical students. Int J Med Inform 2023;177:105117. [doi: 10.1016/j.ijmedinf.2023.105117] [Medline: 37301132]
4. Simsek-Cetinkaya S, Cakir SK. Evaluation of the effectiveness of artificial intelligence assisted interactive screen-based simulation in breast self-examination: an innovative approach in nursing students. Nurse Educ Today 2023;127:105857. [doi: 10.1016/j.nedt.2023.105857] [Medline: 37253303]
5. Sheng B, Chen X, Li T, Ma T, Yang Y, Bi L, et al. An overview of artificial intelligence in diabetic retinopathy and other ocular diseases. Front Public Health 2022;10:971943 [FREE Full text] [doi: 10.3389/fpubh.2022.971943] [Medline: 36388304]
6. Tosaki T, Yamakawa M, Shiina T. A study on the optimal condition of ground truth area for liver tumor detection in ultrasound images using deep learning. J Med Ultrason (2001) 2023;50(2):167-176 [FREE Full text] [doi: 10.1007/s10396-023-01301-2] [Medline: 37014524]
7. Alip SL, Kim J, Rha KH, Han WK. Future platforms of robotic surgery. Urol Clin North Am 2022;49(1):23-38. [doi: 10.1016/j.ucl.2021.07.008] [Medline: 34776052]
8. Han ER, Yeo S, Kim MJ, Lee YH, Park KH, Roh H. Medical education trends for future physicians in the era of advanced technology and artificial intelligence: an integrative review. BMC Med Educ 2019;19(1):460 [FREE Full text] [doi: 10.1186/s12909-019-1891-5] [Medline: 31829208]
9. Oh N, Choi G, Lee WY. ChatGPT goes to the operating room: evaluating GPT-4 performance and its potential in surgical education and training in the era of large language models. Ann Surg Treat Res 2023;104(5):269-273 [FREE Full text] [doi: 10.4174/astr.2023.104.5.269] [Medline: 37179699]
10. Malhotra A, Molloy EJ, Bearer CF, Mulkey SB. Emerging role of artificial intelligence, big data analysis and precision medicine in pediatrics. Pediatr Res 2023;93(2):281-283. [doi: 10.1038/s41390-022-02422-z] [Medline: 36807652]
11. Banerjee M, Chiew D, Patel KT, Johns I, Chappell D, Linton N, et al. The impact of artificial intelligence on clinical education: perceptions of postgraduate trainee doctors in London (UK) and recommendations for trainers. BMC Med Educ 2021;21(1):429 [FREE Full text] [doi: 10.1186/s12909-021-02870-x] [Medline: 34391424]
12. Lee J, Wu AS, Li D, Kulasegaram KM. Artificial intelligence in undergraduate medical education: a scoping review. Acad Med 2021;96(11S):S62-S70. [doi: 10.1097/ACM.0000000000004291] [Medline: 34348374]
13. Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P. Introducing artificial intelligence training in medical education. JMIR Med Educ 2019;5(2):e16048 [FREE Full text] [doi: 10.2196/16048] [Medline: 31793895]
14. Temsah O, Khan SA, Chaiah Y, Senjab A, Alhasan K, Jamal A, et al. Overview of early ChatGPT's presence in medical literature: insights from a hybrid literature review by ChatGPT and human experts. Cureus 2023;15(4):e37281 [FREE Full text] [doi: 10.7759/cureus.37281] [Medline: 37038381]
15. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst 2023;47(1):33 [FREE Full text] [doi: 10.1007/s10916-023-01925-4] [Medline: 36869927]
16. Temsah MH, Aljamaan F, Malki KH, Alhasan K, Altamimi I, Aljarbou R, et al. ChatGPT and the future of digital health: a study on healthcare workers' perceptions and expectations. Healthcare (Basel) 2023;11(13):1812 [FREE Full text] [doi: 10.3390/healthcare11131812] [Medline: 37444647]
17. Albejaidi F, Nair KS. Building the health workforce: Saudi Arabia's challenges in achieving Vision 2030. Int J Health Plann Manage 2019;34(4):e1405-e1416. [doi: 10.1002/hpm.2861] [Medline: 31402508]
18. Wartman SA, Combs CD. Medical education must move from the information age to the age of artificial intelligence. Acad Med 2018;93(8):1107-1109. [doi: 10.1097/ACM.0000000000002044] [Medline: 29095704]
19. List of 38 best universities in Saudi Arabia. EduRank. 2023. URL: https://edurank.org/geo/sar/ [accessed 2025-01-16]
20. Busetto L, Wick W, Gumbinger C. How to use and assess qualitative research methods. Neurol Res Pract 2020;2:14 [FREE Full text] [doi: 10.1186/s42466-020-00059-z] [Medline: 33324920]
21. Mukred M, Mokhtar UA, Hawash B. Exploring the acceptance of ChatGPT as a learning tool among academicians: a qualitative study. J Komunikasi Malays J Commun 2023;39(4):306-323. [doi: 10.17576/jkmjc-2023-3904-16]
22. Iqbal N, Ahmed H, Azhar KA. Exploring teachers' attitudes towards using ChatGPT. GJMAS 2022;3(4):97-111. [doi: 10.46568/gjmas.v3i4.163]
23. Bryman A. Social Research Methods. Oxford, United Kingdom: Oxford University Press; 2016:0199689458.
24. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Analyzing Qualitative Data. London, United Kingdom: Routledge; 2002:173-194.

XSL•FO
RenderX

25. Dhakal K. NVivo. J Med Libr Assoc 2022;110(2):270-272 [FREE Full text] [doi: 10.5195/jmla.2022.1271] [Medline: 35440911]
26. Shoufan A. Exploring students' perceptions of chatGPT: thematic analysis and follow-up survey. IEEE Access 2023;11:38805-38818. [doi: 10.1109/access.2023.3268224]
27. Gülhan Güner S, Yiğit S, Berşe S, Dirgar E. Perspectives and experiences of health sciences academics regarding ChatGPT: a qualitative study. Med Teach 2024:1-10. [doi: 10.1080/0142159X.2024.2413425] [Medline: 39392461]
28. van Bulck L, Moons P. What if your patient switches from Dr. Google to Dr. ChatGPT? A vignette-based survey of the trustworthiness, value, and danger of ChatGPT-generated responses to health questions. Eur J Cardiovasc Nurs 2024;23(1):95-98. [doi: 10.1093/eurjcn/zvad038] [Medline: 37094282]
29. Howard A, Hope W, Gerada A. ChatGPT and antimicrobial advice: the end of the consulting infection doctor? Lancet Infect Dis 2023;23(4):405-406. [doi: 10.1016/S1473-3099(23)00113-5] [Medline: 36822213]
30. Kamoun F, El Ayeb W, Jabri I, Sifi S, Iqbal F. Exploring students' and faculty's knowledge, attitudes, and perceptions towards ChatGPT: a cross-sectional empirical study. J Inf Technol Educ Res 2024;23(4):1-33. [doi: 10.28945/5239]
31. Deladem Kumordzie C. All you need to know about ChatGPT and why its a threat to Google. LinkedIn. URL: https://www.linkedin.com/posts/opheliamtsenuokpor_all-you-need-to-know-about-chatgpt-why-activity-7018610251717332993-zcHV/ [accessed 2025-01-19]
32. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. Cureus 2023;15(2):e35179 [FREE Full text] [doi: 10.7759/cureus.35179] [Medline: 36811129]
33. Masters K. Medical teacher's first ChatGPT's referencing hallucinations: lessons for editors, reviewers, and teachers. Med Teach 2023;45(7):673-675. [doi: 10.1080/0142159X.2023.2208731] [Medline: 37183932]
34. Aljamaan F, Temsah MH, Altamimi I, Al-Eyadhy A, Jamal A, Alhasan K, et al. AI chatbots' medical hallucination: innovation of references hallucination score and comparison of six large language models. JMIR Preprints Preprint posted online on November 6, 2023 [FREE Full text] [doi: 10.2196/preprints.54345]
35. Sanchez-Ramos L, Lin L, Romero R. Beware of references when using ChatGPT as a source of information to write scientific articles. Am J Obstet Gynecol 2023;229(3):356-357. [doi: 10.1016/j.ajog.2023.04.004] [Medline: 37031761]
36. Zhu L, Mou W, Yang T, Chen R. ChatGPT can pass the AHA exams: open-ended questions outperform multiple-choice format. Resuscitation 2023;188:109783. [doi: 10.1016/j.resuscitation.2023.109783] [Medline: 37349064]
37. Huynh LM, Bonebrake BT, Schultis K, Quach A, Deibert CM. New artificial intelligence ChatGPT performs poorly on the 2022 self-assessment study program for urology. Urol Pract 2023;10(4):409-415. [doi: 10.1097/UPJ.0000000000000406] [Medline: 37276372]
38. Weng TL, Wang YM, Chang S, Chen TJ, Hwang SJ. ChatGPT failed Taiwan's Family Medicine Board Exam. J Chin Med Assoc 2023;86(8):762-766. [doi: 10.1097/JCMA.0000000000000946] [Medline: 37294147]
39. Meo SA, Al-Khlaiwi T, AbuKhalaf AA, Meo AS, Klonoff DC. The scientific knowledge of Bard and ChatGPT in endocrinology, diabetes, and diabetes technology: multiple-choice questions examination-based performance. J Diabetes Sci Technol 2023:19322968231203987. [doi: 10.1177/19322968231203987] [Medline: 37798960]
40. Magalhães Araujo S, Cruz-Correia R. Incorporating ChatGPT in medical informatics education: mixed methods study on student perceptions and experiential integration proposals. JMIR Med Educ 2024;10:e51151 [FREE Full text] [doi: 10.2196/51151] [Medline: 38506920]
41. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell 2023;6:1169595 [FREE Full text] [doi: 10.3389/frai.2023.1169595] [Medline: 37215063]
42. Sedaghat S. Early applications of ChatGPT in medical practice, education and research. Clin Med (Lond) 2023;23(3):278-279 [FREE Full text] [doi: 10.7861/clinmed.2023-0078] [Medline: 37085182]
43. Mohammad B, Supti T, Alzubaidi M, Shah H, Alam T, Shah Z, et al. The pros and cons of using ChatGPT in medical education: a scoping review. Stud Health Technol Inform 2023;305:644-647. [doi: 10.3233/SHTI230580] [Medline: 37387114]
44. Aljamaan F, Temsah MH, Altamimi I, Al-Eyadhy A, Jamal A, Alhasan K, et al. Reference hallucination score for medical artificial intelligence chatbots: development and usability study. JMIR Med Inform 2024;12:e54345 [FREE Full text] [doi: 10.2196/54345] [Medline: 39083799]
45. Masters K. Ethical use of artificial intelligence in health professions education: AMEE Guide No. 158. Med Teach 2023;45(6):574-584 [FREE Full text] [doi: 10.1080/0142159X.2023.2186203] [Medline: 36912253]
46. Aljamaan F, Malki KH, Alhasan K, Jamal A, Altamimi I, Khayat A, et al. ChatGPT-3.5 System Usability Scale early assessment among healthcare workers: horizons of adoption in medical practice. Heliyon 2024 Apr 15;10(7):e28962 [FREE Full text] [doi: 10.1016/j.heliyon.2024.e28962] [Medline: 38623218]
47. Rodgers CM, Ellingson SR, Chatterjee P. Open data and transparency in artificial intelligence and machine learning: a new era of research. F1000Res 2023;12:387 [FREE Full text] [doi: 10.12688/f1000research.133019.1] [Medline: 37065505]
48. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT—reshaping medical education and clinical management. Pak J Med Sci 2023;39(2):605-607 [FREE Full text] [doi: 10.12669/pjms.39.2.7653] [Medline: 36950398]

XSL•FO
RenderX

49. Jamal A, Solaiman M, Alhasan K, Temsah M, Sayed G. Integrating ChatGPT in medical education: adapting curricula to cultivate competent physicians for the AI era. Cureus 2023;15(8):e43036 [FREE Full text] [doi: 10.7759/cureus.43036] [Medline: 37674966]
50. Gupta R, Pande P, Herzog I, Weisberger J, Chao J, Chaiyasate K, et al. Application of ChatGPT in cosmetic plastic surgery: ally or antagonist? Aesthet Surg J 2023;43(7):NP587-NP590. [doi: 10.1093/asj/sjad042] [Medline: 36840479]
51. Shahsavar Y, Choudhury A. User intentions to use ChatGPT for self-diagnosis and health-related purposes: cross-sectional survey study. JMIR Hum Factors 2023;10:e47564 [FREE Full text] [doi: 10.2196/47564] [Medline: 37195756]
52. Das D, Kumar N, Longjam LA, Sinha R, Deb Roy A, Mondal H, et al. Assessing the capability of ChatGPT in answering first- and second-order knowledge questions on microbiology as per competency-based medical education curriculum. Cureus 2023;15(3):e36034 [FREE Full text] [doi: 10.7759/cureus.36034] [Medline: 37056538]
53. Hasanein AM, Sobaih AEE. Drivers and consequences of ChatGPT use in higher education: key stakeholder perspectives. Eur J Investig Health Psychol Educ 2023;13(11):2599-2614 [FREE Full text] [doi: 10.3390/ejihpe13110181] [Medline: 37998071]
54. Temsah MH, Altamimi I, Jamal A, Alhasan K, Al-Eyadhy AA. ChatGPT surpasses 1000 publications on PubMed: envisioning the road ahead. Cureus 2023;15(9):e44769 [FREE Full text] [doi: 10.7759/cureus.44769] [Medline: 37809155]
55. ChatGPT. OpenAI. URL: https://chatgpt.com/ [accessed 2025-01-23]
56. Abouammoh N, Alhasan K, Raina R, Malki KA, Aljamaan F, Tamimi I, et al. Exploring perceptions and experiences of ChatGPT in medical education: a qualitative study among medical college faculty and students in Saudi Arabia. medRxiv Preprint posted online on July 16, 2023. [doi: 10.1101/2023.07.13.23292624]

Abbreviations

AI: artificial intelligence

Edited by B Lesselroth; submitted 18.06.24; peer-reviewed by F Kamoun, SM Araujo, KA Morbitzer; comments to author 02.09.24; revised version received 03.11.24; accepted 02.01.25; published 20.02.25.

Please cite as:
Abouammoh N, Alhasan K, Aljamaan F, Raina R, Malki KH, Altamimi I, Muaygil R, Wahabi H, Jamal A, Alhaboob A, Assiri RA, Al-Tawfiq JA, Al-Eyadhy A, Soliman M, Temsah MH
Perceptions and Earliest Experiences of Medical Students and Faculty With ChatGPT in Medical Education: Qualitative Study
JMIR Med Educ 2025;11:e63400
URL: https://mededu.jmir.org/2025/1/e63400
doi:10.2196/63400
PMID:39977012

©Noura Abouammoh, Khalid Alhasan, Fadi Aljamaan, Rupesh Raina, Khalid H Malki, Ibraheem Altamimi, Ruaim Muaygil, Hayfaa Wahabi, Amr Jamal, Ali Alhaboob, Rasha Assad Assiri, Jaffar A Al-Tawfiq, Ayman Al-Eyadhy, Mona Soliman, Mohamad-Hani Temsah. Originally published in JMIR Medical Education (https://mededu.jmir.org), 20.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Detecting Artificial Intelligence–Generated Versus Human-Written Medical Student Essays: Semirandomized Controlled Study

Berin Doru[1]; Christoph Maier[1]; Johanna Sophie Busse[1]; Thomas Lücke[1]; Judith Schönhoff[2]; Elena Enax- Krumova[3]; Steffen Hessler[4]; Maria Berger[5*]; Marianne Tokic[6*]

[1]University Hospital of Paediatrics and Adolescent Medicine, St. Josef-Hospital, Ruhr University Bochum, Bochum, Germany
[2]Departement of German Philology, General and Comparative Literary Studies, Ruhr University Bochum, Bochum, Germany
[3]Department of Neurology, BG University Hospital Bergmannsheil gGmbH Bochum, Ruhr University Bochum, Bochum, Germany
[4]German Department, German Linguistics, Ruhr University Bochum, Bochum, Germany
[5]German Department, Digital Forensic Linguistics, Ruhr University Bochum, Bochum, Germany
[6]Department for Medical Informatics, Biometry and Epidemiology, Ruhr University Bochum, Bochum, Germany
[*]these authors contributed equally

Corresponding Author:
Berin Doru
University Hospital of Paediatrics and Adolescent Medicine, St. Josef-Hospital
Ruhr University Bochum
Alexandrinenstraße 5
Bochum, 44791
Germany
Phone: 49 234 509 2611
Email: Berin.Doru@rub.de

Abstract

Background: Large language models, exemplified by ChatGPT, have reached a level of sophistication that makes distinguishing between human- and artificial intelligence (AI)–generated texts increasingly challenging. This has raised concerns in academia, particularly in medicine, where the accuracy and authenticity of written work are paramount.

Objective: This semirandomized controlled study aims to examine the ability of 2 blinded expert groups with different levels of content familiarity—medical professionals and humanities scholars with expertise in textual analysis—to distinguish between longer scientific texts in German written by medical students and those generated by ChatGPT. Additionally, the study sought to analyze the reasoning behind their identification choices, particularly the role of content familiarity and linguistic features.

Methods: Between May and August 2023, a total of 35 experts (medical: n=22; humanities: n=13) were each presented with 2 pairs of texts on different medical topics. Each pair had similar content and structure: 1 text was written by a medical student, and the other was generated by ChatGPT (version 3.5, March 2023). Experts were asked to identify the AI-generated text and justify their choice. These justifications were analyzed through a multistage, interdisciplinary qualitative analysis to identify relevant textual features. Before unblinding, experts rated each text on 6 characteristics: linguistic fluency and spelling/grammatical accuracy, scientific quality, logical coherence, expression of knowledge limitations, formulation of future research questions, and citation quality. Univariate tests and multivariate logistic regression analyses were used to examine associations between participants' characteristics, their stated reasons for author identification, and the likelihood of correctly determining a text's authorship.

Results: Overall, in 48 out of 69 (70%) decision rounds, participants accurately identified the AI-generated texts, with minimal difference between groups (medical: 31/43, 72%; humanities: 17/26, 65%; odds ratio [OR] 1.37, 95% CI 0.5-3.9). While content errors had little impact on identification accuracy, stylistic features—particularly redundancy (OR 6.90, 95% CI 1.01-47.1), repetition (OR 8.05, 95% CI 1.25-51.7), and thread/coherence (OR 6.62, 95% CI 1.25-35.2)—played a crucial role in participants' decisions to identify a text as AI-generated.

Conclusions: The findings suggest that both medical and humanities experts were able to identify ChatGPT-generated texts in medical contexts, with their decisions largely based on linguistic attributes. The accuracy of identification appears to be independent of experts' familiarity with the text content. As the decision-making process primarily relies on linguistic attributes—such as stylistic features and text coherence—further quasi-experimental studies using texts from other academic disciplines should be

XSL•FO
RenderX

conducted to determine whether instructions based on these features can enhance lecturers' ability to distinguish between student-authored and AI-generated work.

(JMIR Med Educ 2025;11:e62779) doi:10.2196/62779

KEYWORDS

artificial intelligence; ChatGPT; large language models; textual analysis; writing style; AI; chatbot; LLMs; detection; authorship; medical student; textual analysis; linguistic quality; decision-making; logical coherence

Introduction

The rapid development of artificial intelligence (AI) and the emergence of large language models (LLMs), such as ChatGPT, have increasingly blurred the lines between human-written and AI-generated text. This has created a significant challenge in identifying the authorship of written work, especially as the use of AI has become ubiquitous since chatbots have become freely available [1,2]. Consequently, critical concerns have arisen in the educational and academic sectors, where the reliability and authenticity of written work are fundamental.

According to a recent nationwide survey, nearly two-thirds of the German students reported using AI-based tools for their studies, with ChatGPT being the most commonly used chatbot [3]. ChatGPT, developed by OpenAI, an American AI research laboratory, is a state-of-the-art AI chatbot capable of assisting users with a wide range of tasks, from text generation to problem-solving [4,5]. Its capabilities have opened up significant opportunities for educational and academic contexts. For example, AI-based tools like ChatGPT can support tasks such as text analysis, translation, and proofreading for research purposes [6]. They also can provide support to students by enhancing their understanding of scientific methods, improving and refining written work, and assisting with examination preparation [3,7].

However, the widespread use of such tools raises concerns about their impact on students' development of critical and independent thinking skills [8,9]. In addition, it is possible that ChatGPT could provide incomplete or inaccurate information, potentially leading to misunderstandings of academic concepts and topics [10,11]. Further concerns arise from the potential for academic dishonesty and plagiarism, particularly in the context of written assignments and academic essays [9,12].

The lack of clarity on how to handle such cases is putting universities in a quandary, leading to the first court cases and, more recently, to the University of Munich being vindicated in its decision to reject such written work [13]. In this case, an essay submitted as part of a Master's application was rejected because it was "too well written," raising suspicions that the text was likely generated by an AI tool such as ChatGPT [13]. This case highlights that AI-generated texts are often characterized by their seemingly perfected style of formulation that refers to the linguistic level [13,14], a characteristic that is known to be particularly pronounced and even more nuanced in English output texts compared with other languages such as German [15-18].

The relevance of the problem for medical studies is not obvious at first glance because medical students generally do not have to write long scientific texts on a regular basis during their studies, but usually only for their doctoral thesis. However, the increasing use of AI tools such as ChatGPT also poses challenges in the medical field, where assessments rely not only on linguistic quality but also on content accuracy. The potential misattribution of authorship in medical texts—such as research articles, patient information, or promotional materials—has particularly serious implications, as errors or inaccuracies in these contexts can have grave consequences. It can be assumed that medical texts do not fall as much within the scope of ChatGPT and are therefore more difficult to reproduce accurately, especially because AI authors have no "moral scruples" about concealing ignorance and replacing verified sources with falsified ones [19-22]. However, a few studies have addressed the problem of AI-generated content on medical texts [21-26], reporting that ChatGPT has, at times, managed to mislead medical professionals [26], which may suggest that familiarity with content plays a minor role in authorship identification.

Existing research on LLM-based text generators such as ChatGPT frequently focuses on their role in assisting the writing process rather than evaluating the quality and detectability of longer scientific texts [27-29]. In addition, studies often investigate the detection of texts written by chatbots using automatic tools or even detectors specifically designed for this purpose [18,25,27-30]. The detection rate of these detectors is often higher than that of human reviewers, but the accuracy can vary greatly depending on the text genre and the classifier used [14,31]. Moreover, linguistic features appear to be the most important subset of features influencing the performance of feature-based classifiers [2,5]. In the academic domain, educators still face the challenge of qualitatively assessing the authenticity of student texts, often without the aid of automated detection tools.

It is therefore of particular interest to determine how well human readers from the academic field can detect differences when directly comparing 2 texts on the same topic—an original student text and an AI-generated text—and which features stand out as particularly conspicuous and decisive for them. To better distinguish between the relevance of content-related and text-analytical attributes, we assembled a group of language experts from the humanities field alongside a group of medical experts specializing in pediatrics and neurology. Another key novelty and prerequisite of our study is the use of fully reproduced, longer scientific texts on medical topics written in German by medical students.

Therefore, we conducted a study to determine whether medical experts and humanities lecturers could distinguish between texts written by medical students and those generated by ChatGPT

XSL•FO
RenderX

(specifically ChatGPT version 3.5, March 23). Unlike the interactive "Turing Test" [32], this task was not performed through a dialogue with a machine but rather through an internal, personal evaluation of 2 texts. We hypothesize that, in line with the Turing prophecy [32], the correct identification rate for AI-generated texts within a German medical sample is approximately 70%, with content familiarity playing a secondary role, while formal and linguistic features exert a greater influence.

Through a prospective analysis of longer German-language scientific texts written by students in the specialized health-related field of medicine, this study aims to provide new insights into AI-generated texts and the influence of content familiarity and linguistic expertise. The findings are intended to inform the development of guidelines to help lecturers (and others) recognize AI-generated texts, even in the absence of a comparable "original" text, and to contribute to future projects addressing the challenges posed by AI tools in academia.

Methods

Recruitment Process

This semirandomized controlled trial was conducted between May and August 2023 at the University Hospital of Pediatrics and Adolescent Medicine, St. Josef-Hospital, Ruhr University Bochum (RUB), Germany. To recruit participants, an open call was issued to the Department of Pediatrics and Adolescent Medicine and the Department of Neurology at the University Hospital Bergmannsheil (both RUB). Senior physicians and members of scientific working groups were invited to participate, ensuring the involvement of clinical experts familiar with the content of the texts. Participation was voluntary, with clinical employment or medical expertise serving as key inclusion criteria, along with an interest in scientific texts. In the next phase of recruitment, a call was made to the Faculty of Humanities at Ruhr University to include participants with experience in text reception and analysis, with teaching experience as an additional inclusion criterion.

Design

Each participant received 2 pairs of texts, totaling 4 printed texts. Each pair consisted of 1 of 18 available term papers written by a medical student and a corresponding text generated by ChatGPT (version 3.5, March 2023), with the order of presentation randomized. The medical experts received 1 pair of texts on a topic closely related to their specialty and another pair on a less familiar topic. For example, a pediatrician received the text "Autoantibodies in Diabetes Mellitus," while a neurologist received "Measurement of Aδ and C Fibers in Electrophysiology." The second pair of texts covered a topic less directly related to their field of expertise.

Given the exploratory nature of the study, it became increasingly evident—only after completing the experimental phase with the medical experts—that the extent to which content familiarity influences the identification process needed further examination, particularly in comparison to formal and linguistic aspects.

As a result, a second phase of the study was initiated, involving a new group of experts with greater expertise in formal and linguistic analyses. This allowed for a comparison of results and evaluations across groups. In this group of humanities experts, each participant received the same 2 pairs of texts to ensure better comparability and verification of subject unfamiliarity. The first pair addressed a more general topic also familiar to nonmedical fields: "Iodine Deficiency." The second pair analyzed a more specialized medical topic: "Autoantibodies in Diabetes Mellitus."

Participants were asked to read a pair of texts and, based on their personal experience with student-written texts, decide within a week—without extensive research—which of the 2 they believed was generated by ChatGPT. To ensure the blinding of both interviewers and participants, the headers of the texts contained only a randomly generated 3-digit number and a "chatbot or student" checkbox. Before the subsequent interview, participants documented their decision by ticking the corresponding box for their chosen text version. They were also instructed not to discuss the task or the texts with one another.

About 1-2 weeks after the texts were distributed, participants were invited to a semistructured interview—conducted in person, by telephone, or via Zoom (Zoom Communications/Qumu Corporation)—to discuss their decisions, reasoning, and evaluations of the texts. Unblinding occurred after the interviews.

Creation of the ChatGPT-Generated Versions

Eighteen German-language medical essays served as templates for the ChatGPT-generated texts. These essays were written by doctoral students from the University Pediatric Clinic and the Clinic for Neurology at Bergmannsheil University Hospital, RUB, Germany. They originated from the Doctoral Colloquium pool at the University Hospital of Pediatrics and Adolescent Medicine in Bochum. As part of the colloquium, each doctoral student is encouraged to write a scientific essay thematically related to their announced dissertation topic. To provide an initial experience with scientific research and writing—and to give the reviewing study coordinator a first impression of their academic level and skills—students do not receive specific instructions. For this study, all available German texts from this pool that were written before the general introduction of ChatGPT were considered, provided their authors consented to their use.

We used ChatGPT version 3.5, March 14 to replicate the texts. To generate a version with the same title and outline as the original papers while avoiding text breaks, 2 separate prompts were required to produce a continuous text from the introduction to the conclusion (Table 1).

For the main part, depending on the type of original paper, several commands were necessary. For example, see Table 2.

We then merged the individual sections to create a complete term paper, supplementing it with a bibliography that listed the sources provided by ChatGPT in sequential order. To ensure consistency, we harmonized the formatting of both ChatGPT-generated and student-written texts as much as possible, using the Arial font (size 11 for body text and size 12 for headings) with justified alignment. Sections or sentences specific to a student's individual dissertation project were

XSL•FO
RenderX

removed to maintain general applicability. However, we did not alter the choice of words, sentence structure, punctuation, spelling, or citation style.

Table 1. Prompts used to create ChatGPT text.

German (original prompt)	English (translation)
• "Schreibe bitte einen Abschnitt über das Thema [TITEL DES TEILTHEMAS] der [wissenschaftlichen/medizinischen] Hausarbeit [TITEL DER HAUSARBEIT]." • "Belege Deine Aussagen mit Quellen, die bei Pubmed auffindbar sind."	• "Please write a section on the topic [NAME OF SUBTOPIC] of the [scientific/medical] term paper [NAME OF TERM PAPER]." • "Support your statements with sources that can be found on PubMed."

Table 2. Additional instructions used to create ChatGPT text.

German (original prompt)	English (translation)
"Schreibe einen Abschnitt zum Thema 'Potenziell reversible Pathomechanismen als mögliche Ursachen von Hyposmie oder Anosmie bei Kindern' der wissenschaftlichen Hausarbeit 'Ursachen und Diagnostik von Riechstörungen bei Kindern und Jugendlichen', der an den vorherigen Abschnitt anknüpft."	"Write a section on the topic 'Potentially reversible pathomechanism as possible causes of hyposmia or anosmia in children' of the scientific paper 'Causes and diagnostics of olfactory dysfunction in children and adolescents', which ties in with the previous section."

Data Assessment During the Interview

The medical expert group was interviewed by 2 blinded interviewers (BD and JSB), while the humanities expert group was interviewed by a partially blinded interviewer (CM). Initially, participants provided demographic information, including age, experience in academic and student teaching, academic qualifications, publication history, and prior experience with ChatGPT. They were then asked to assess how well the following questions were addressed in each text, using the German grading system from 1 (very good) to 6 (unsatisfactory) (see Multimedia Appendix 1 for details). How would you rate (1) linguistic fluency, (2) scientific quality (eg, are the re-definitions scientifically derived and are studies cited that lead to certain conclusions?), (3) internal logic, (4) description of the limitations of current knowledge, (5) future research questions, and (6) citations and references of the text? Participants were then asked to identify which text version, using the corresponding 3-digit number, they had categorized as being generated by ChatGPT and to list the key reasons for their decision, which the interviewer recorded using keywords. Next, they rated their confidence in their decision on a scale from 1 (very confident) to 6 (not confident at all). After this initial assessment, participants were unblinded and informed which text had been written by a student and which by ChatGPT. In cases of misidentification, they were asked about their suspected reasons, which the interviewer also documented using keywords.

Construction of Categories

Beyond the identification rate and the text evaluations by each group, a qualitative analysis of participants' statements regarding their reasoning for assigning authorship proved essential. This deeper analysis aimed to examine the influence of content-related versus formal-linguistic aspects and to better attribute global features to either student or chatbot authorship. For this purpose, the free-text responses (recorded by interviewers using keywords) were first thematically clustered based on the terms mentioned (see sample statements in the *Free-Text Analysis* section). Subsequently, through multiturn discussions between medical and linguistic experts, these thematic clusters were refined into distinct, nonoverlapping categories that encompassed all interviewee statements while reducing redundancy and multiple classifications.

Many of these categories align with standard text-analytical frameworks, which typically cover a broad range of attributes, including morphology, syntax, style, structure, coherence and cohesion, content quality, form, and even sociolinguistic aspects [33-35]. However, the categories derived in this study are directly based on the text types used in the experiments. As a result, they provide a more precise representation of the emerging and still undefined text type "AI-generated" and are therefore preferable (see Multimedia Appendix 2 for an overview of the categories).

Statistical Analysis

Data were analyzed using Microsoft Excel, SPSS version 29.3 (IBM Corp.), and R-4.1.2 (R Foundation). Descriptive statistics are presented as numbers (n) and percentages or as means (SD), where appropriate. Univariate odds ratios and 95% CIs from the Fisher exact test were used to examine the association between demographic markers, participants' field (medicine or humanities), and their expertise with the likelihood of correctly identifying a text's source. The relationship between interview scores and response accuracy was assessed using the 2-sided Wilcoxon signed rank test for paired values. Additionally, correlations among all 5 responses were tested using a Friedman 2-way analysis of variance for ranks with Bonferroni correction.

To analyze how participants attempted to identify the machine-generated text, we modeled the association of the derived categories (items) from the interviews based on their likelihood of being mentioned in the context of a chatbot-generated text. For each participant and interview, we recorded whether an item was cited in reference to a perceived chatbot text or a perceived student text. This association was analyzed using repeated-measures logistic regression, incorporating a random participant and sequence effect. The

XSL•FO
RenderX

model was further adjusted for age group, the expert group (medical vs humanities), and prior experience with ChatGPT (binary).

Ethical Considerations

An application for the study project was submitted to the Ethics Committee of the Medical Faculty at RUB (reference number 23-7837; April 2023). As the study did not involve direct research on human participants or patient data, the committee informed us that ethical approval was not required.

Results

Interviewee Sample

The biographical data of the 22 participating physicians (14 pediatricians, 3 nutritionists, 4 neurologists, and 1 neuroscientist) and 13 humanities scholars (8 literary scholars, 3 Germanists or linguists, 1 classical philologist, and 1 Romance philologist) are presented in Table 3.

As there were more participating experts than available term papers, 3 pairs of texts were each assessed by 3 or 4 medical experts.

At the time of the survey, only one-fifth of the participants reported having prior experience with ChatGPT. As the number of participating experts exceeded the number of available term papers, 3 pairs of texts were each assessed by 3 or 4 medical experts.

Table 3. Interviewee sample.

Characteristics	All participants (N=35)	Medical experts (n=22)	Humanities experts (n=13)
Age (years), n (%)			
<40	17 (49)	13 (59)	4 (31)
≥40	18 (51)	9 (41)	9 (69)
Experience in academic teaching[a] (years), n (%)			
None	2 (6)	2 (9)	N/A[b]
<5	8 (23)	8 (36)	N/A
≥5	25 (71)	12 (55)	13 (100)
PhD/professorship, n (%)			
Yes	28 (80)	18 (82)	10 (77)
Authorship in a publication, n (%)			
Yes	32 (91)	20 (91)	12 (92)
Experience with ChatGPT[a], n (%)			
Yes	7 (20)	2 (9)	5 (38)
Only a little	7 (20)	4 (18)	3 (23)
No	21 (60)	16 (73)	5 (38)

[a]Self-assessed.

[b]N/A: not applicable.

Detection Rate

With 35 participants evaluating 2 text pairs each—excluding 1 misaligned and omitted case—a total of 69 decision rounds were conducted. In 48 out of 69 (70%) decision rounds, participants correctly identified the authorship of the texts. Medical and humanities experts showed a slight but nonsignificant difference in detection rates, with medical experts correctly identifying 31 out of 43 (72%) decision rounds compared with 17 out of 26 (65%) decision rounds by humanities experts (odds ratio 1.37, 95% CI 0.5-3.9). Among the 35 participants, 21 (60%) misidentified the authorship of at least one text pair, including 12 medical experts. Additionally, 5 (14%) participants, including 3 physicians, misidentified both text pairs.

Notably, familiarity with the topic did not significantly impact identification accuracy (Table 4), nor did personal characteristics such as age, academic qualifications, or years of teaching experience. However, younger participants without advanced academic titles showed a slight tendency to better identify ChatGPT-generated texts. Confidence in participants' own judgments did not differ significantly across groups (odds ratio 0.6, 95% CI 0.2-1.79).

XSL•FO
RenderX

Table 4. Characteristics of the participants with correct and incorrect decisions about the authorship of the respective text.

Characteristics	All participants			Medical experts			Humanities experts		
	Decision			Decision			Decision		
	Correct	False	OR[a] (95% CI)	Correct	False	OR (95% CI)	Correct	False	OR (95% CI)
Tests, n (%)	48 (70)	21 (30)	N/A[b]	31 (72)	12 (28)	N/A	17 (65)	9 (35)	
Age (years), n (%)									
<40	24 (73)	9 (27)	N/A	17 (68)	8 (32)	N/A	7 (88)	1 (12)	N/A
≥40	24 (67)	12 (33)	0.75 (0.27-2.11)	14 (78)	4 (22)	1.65 (0.41-6.63)	10 (56)	8 (44)	0.18 (0.02-1.77)
PhD/professorship, n (%)									
Yes	37 (67)	18 (33)	N/A	25 (71)	10 (29)	N/A	12 (60)	8 (40)	N/A
No	11 (78.6)	3 (21.4)	1.78 (0.44-7.2)	6 (75)	2 (25)	1.2 (0.21-6.98)	5 (83.3)	1 (16.7)	3.33 (0.33-34.12)
Experience in academic teaching[c] (years), n (%)									
≥5	30 (67)	15 (33)	N/A	16 (70)	7 (30)	N/A	14 (64)	8 (36)	N/A
<5 or none	18 (75)	6 (25)	1.5 (0.49-4.56)	15 (75)	5 (25)	1.31 (0.34-5.05)	3 (75)	1 (25)	1.71 (0.15-19.36)
Authorship in a publication, n (%)									
Yes	44 (70)	19 (30)	N/A	28 (72)	11 (28)	N/A	16 (67)	8 (33)	N/A
No	4 (67)	2 (33)	0.86 (0.15-5.12)	3 (75)	1 (25)	1.18 (0.11-12.59)	1 (50)	1 (50)	0.5 (0.03-9.08)
Experience with ChatGPT[c], n (%)									
Yes	20 (74)	7 (26)	N/A	9 (82)	2 (18)	N/A	11 (69)	5 (31)	N/A
No	28 (67)	14 (33)	0.7 (0.24-2.05)	22 (69)	10 (31)	0.49 (0.09-2.69)	6 (60)	4 (40)	0.68 (0.13-3.55)
Text pair (sequence), n (%)									
First	23 (66)	12 (34)	N/A	16 (73)	6 (27)	N/A	7 (54)	6 (46)	N/A
Second	25 (74)	9 (26)	0.75 (0.27-2.09)	15 (71)	6 (29)	0.94 (0.25-3.56)	10 (77)	3 (23)	2.86 (0.53-15.47)
Familiar with the topic, n (%)									
More	25 (68)	12 (32)	N/A	18 (75)	6 (25)	N/A	7 (54)	6 (46)	N/A
Less	23 (72)	9 (28)	1.01 (0.36-2.8)	13 (68)	6 (32)	0.72 (0.19-2.75)	10 (77)	3 (23)	2.86 (0.53-15.47)
Self-confidence in the decision[c], n (%)									
Rather sure	35 (73)	13 (27)	N/A	22 (76)	7 (24)	N/A	13 (68)	6 (32)	N/A
Unsure	13 (62)	8 (38)	0.6 (0.2-1.79)	9 (64)	5 (36)	0.57 (0.14-2.29)	4 (57)	3 (43)	0.62 (0.1-3.66)

[a]OR: odds ratio for the correct decision.

[b]N/A: not applicable.

[c]Self-assessed.

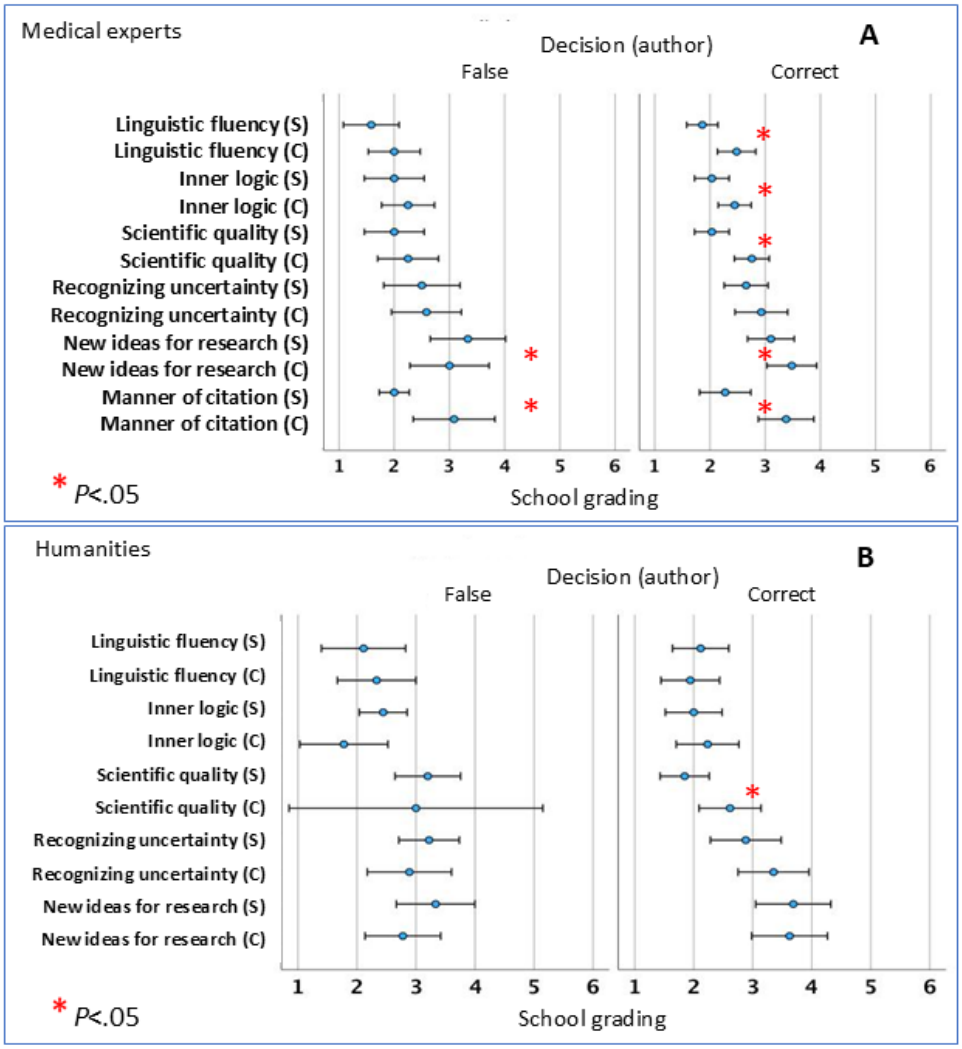
Interview Analysis

When authorship was correctly identified, texts written by medical students received notably higher ratings for stylistic fluency, the internal logic of argumentation, and scientific quality (see Figure 1A, right-hand side). These differences were less pronounced in the assessment of knowledge limitations and future research directions. Regardless of correct identification, the way sources were cited was consistently rated higher in student-authored texts (Figure 1). Among humanities experts, score differences between correctly and incorrectly classified texts were minimal, though the academic quality of student-written texts was still rated significantly (r=0.336, P=.009) higher overall (Figure 1B, right-hand side). Many participants who misidentified the authorship attributed their errors to either underestimating the quality of student work or overestimating ChatGPT's capabilities—particularly in terms

XSL•FO
RenderX

of logical coherence, the presentation of scientific knowledge limitations, and the formulation of new research ideas.

Figure 1. Association of mean school grade (using German school grades 1=very good to 6=unsatisfactory) and correctness of authoring identification in (A) medical experts and (B) humanities experts. Participants were not yet unblinded at the time of assessment (see Multimedia Appendix 1 for details). Left side: incorrect attribution, right side: correct attribution of authorship. *P<.05 (2-sided Wilcoxon signed rank test). C: chatbot-generated text; S: student text.



Free-Text Analysis

We categorized the 187 freely formulated reasons participants provided for their decisions into 1 of 12 derived categories (see Multimedia Appendix 2). Three categories were excluded from statistical analysis due to their low frequency: inconsistency of writing style (n=4), other issues (n=4), and errors in content (n=3). Notably, all 3 content-error attributions came from the medical expert group. Of the remaining 176 statements, 88 (50%) were contributed by medical experts and 88 (50%) by humanities scholars (Table 5).

The experiment revealed that significantly more statements were made about (suspected) ChatGPT-generated texts (130/176, 73.9%) than about student-written texts, regardless of whether the suspicion was correct (Table 5). Sample statements from both groups are provided in Tables 6 and 7. Medical experts' explanations were often concise, frequently critiquing a "superficial" style with "unnecessary additional information." By contrast, humanities experts tended to provide more detailed justifications, describing characteristics such as "smooth style" and "redundancies."

We analyzed the likelihood of specific categories being mentioned in reference to texts suspected to be generated by ChatGPT. The results indicate that "redundancy" (12/14, 86%, associated with GPT vs 2/14, 14%, with student texts), "repetition" (20/22, 91% vs 2/22, 9%), and "common thread and coherence" (21/24, 88% vs 3/24, 13%) were the most frequently cited characteristics (Figure 2).

XSL•FO
RenderX

Table 5. The remaining 9 categories and item frequency by presumed nature of the text (for a detailed explanation in German and English, see Multimedia Appendix 2).

Item	Category	How often mentioned				
		Overall, n (%)	Humanities		Medical experts	
			Chatbot, n (%)	Students, n (%)	Chatbot, n (%)	Students, n (%)
		176 (100)	59 (33.5)	29 (16.5)	71 (40.3)	17 (9.7)
1	Differentiated content	16 (9.1)	4 (6.8)	4 (13.8)	4 (5.6)	4 (23.5)
2	Superficial content	13 (7.4)	2 (3.4)	2 (6.9)	8 (11.3)	1 (5.9)
3	Redundancy	14 (8.0)	7 (11.9)	2 (6.9)	5 (7.0)	N/A[a]
4	Monotonous structure of sentences	14 (8.0)	6 (10.2)	4 (13.8)	4 (5.6)	N/A
5	Repetition	22 (12.5)	8 (13.6)	1 (3.4)	12 (16.9)	1 (5.9)
6	Common thread coherency	24 (13.6)	10 (16.9)	2 (6.9)	11 (15.5)	1 (5.9)
7	Distinctive literature style	16 (9.1)	3 (5.1)	2 (6.9)	8 (11.3)	3 (17.6)
8	Form	25 (14.2)	6 (10.2)	6 (20.7)	9 (12.7)	4 (23.5)
9	Distinctive Wording	32 (18.2)	13 (22.0)	6 (20.7)	10 (14.1)	3 (17.6)

[a]N/A: not applicable.

Table 6. Excerpt from the statements of the medical expert group on the main reasons for choosing ChatGPT as the author.

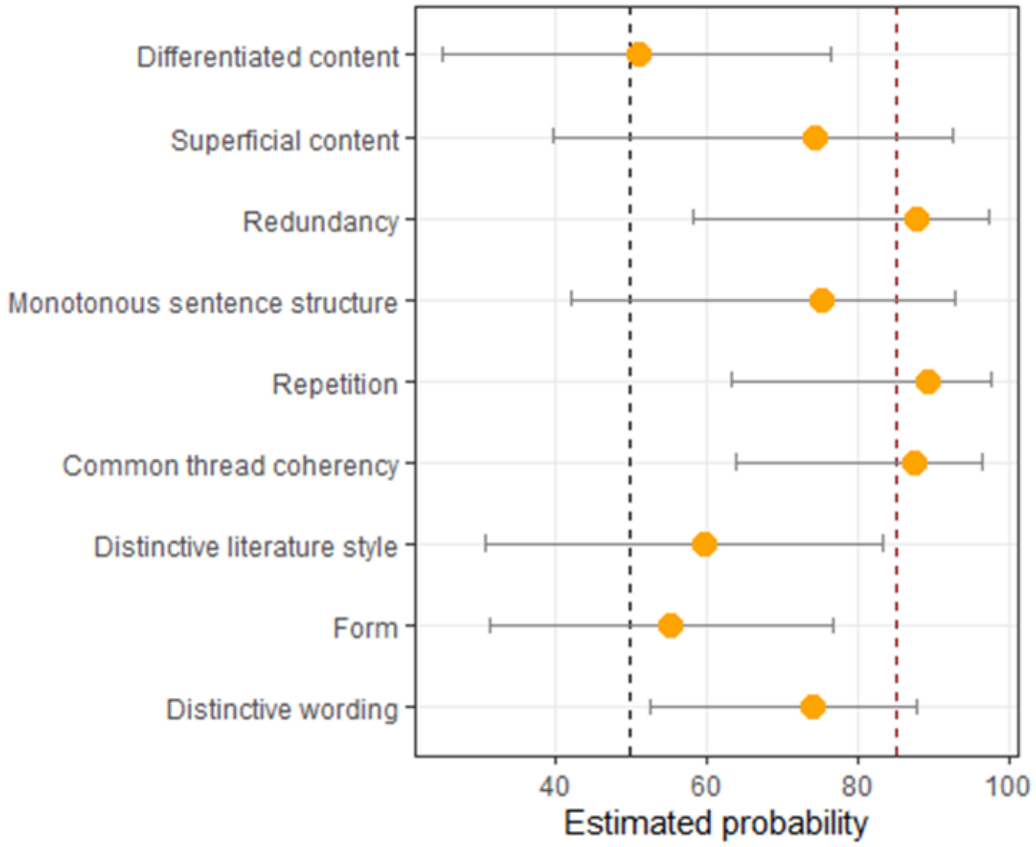
German (original statement)	English (translation)
“,erschien zu perfekt geschrieben, oberflächlich‘”	“,seemed too perfectly written, superficial’”
“,unnützes Wissen, Zusatzinfos, die nicht notwendig für die Arbeit wären‘”	“,useless knowledge, additional information that is not necessary for the work‘”
“,fehlender roter Faden, fehlende Kontinuität der Logik’”	“,lack of common thread, lack of continuity of logic’”

Table 7. Excerpt from the statements of the humanities expert group on the main reasons for choosing ChatGPT as the author.

German (original statement)	English (translation)
“,fehlende Kohärenz, Redundanz, Monotonie […] ist sehr redundant, wiederholt Formulierungen teils mehrfach in Variationen. Der Definitionsteil wirkt reihend, stilistisch homogen, […] teils erfährt man, was man sich hätte denken können, […]. Der letzte Absatz wiederholt in etwa, was vorher da stand - was so wirkt, als hätte er diesen schon 'vergessen'.”	“,lack of coherence, redundancy, monotony [...] is very redundant, repeats formulations sometimes several times in variations. The definition section appears to be sequential, stylistically homogeneous, [...] partly one learns what one could have imagined, [...]. The last paragraph roughly repeats what was there before - which makes it seem as if he had already 'forgotten' it.”
“,Smooth, fließende Übergänge, aber übertextlich schlechter, d.h. gesamt Kohärenz schlechter (Wiederholung), habe nichts gelernt‘.”	“,Smooth, fluent transitions, but overtextually worse, i.e. overall coherence worse (repetition), 'haven't learned anything'.”
“,Wiederholung vieler Sätze und Inhalte; ausgeprägte Tendenz zu bestimmten schablonenartigen Formulierungen im Sinne einer ‘Anmoderation’, z.T. phrasenhaft ohne Inhalt. Optisch gute Gliederung (vorgegeben durch die Überschriften), jedoch roter Faden nicht gut erkennbar, vieles wirkt lediglich wie aufgelistete Einzelformationen. […] Nennung vieler Quellen, deren Zuordnung zu einzelnen Aussagen ist oft nicht konkret.”	“,Repetition of many sentences and contents; pronounced tendency towards certain template-like formulations in the sense of a ‘presentation’, sometimes phrase-like without content. Visually well structured (given by the headings), but a common thread is not easily recognizable, many things appear to be merely a list of individual formations. [...] Mention of many sources, their assignment to individual statements is often not concrete.”

XSL•FO
RenderX

Figure 2. Estimated probability of thinking of ChatGPT as authorship.



Discussion

Overview

This analysis offers insights into the current identification rate of AI-generated texts and their evaluation compared with medical student texts by 2 different expert groups. It also provides an initial overview of the decision-making processes of medical and humanities experts during these assessments. Our findings suggest that both medical and humanities experts can effectively identify ChatGPT-generated texts in medical contexts and that linguistic and stylistic features play a significant role in distinguishing AI-generated from human-written texts, regardless of content familiarity. This supports the broader notion that linguistic analysis is crucial in identifying AI-generated text, aligning with foundational theories in human-robot interaction, such as Turing's predictions [32].

Identification Rate

In the 1950s, Alan Turing [32] predicted that within 50 years, AI would advance to the point where the likelihood of identifying a machine as nonhuman in a dialogue or an "imitation game" would be no more than 70% [32]. With a slight delay of about 20 years, his prediction was almost precisely fulfilled in an online game inspired by the Turing Test [5]. Unlike Turing's method and the large-scale Israeli study, our research did not involve direct dialogue between humans and machines [5,32]. However, when participants were presented with 2 texts of different authorship, an internal dialogue was essential for making an authorship determination. Ultimately, our study's main finding aligns almost exactly with Turing's prediction: only in 48 out of 69 (70%) decision rounds, participants correctly identified the ChatGPT-generated text. This accuracy rate remained consistent regardless of whether participants were experts in the content of the text or in linguistic analysis, and irrespective of their prior experience with ChatGPT. Notably, familiarity with the subject matter did not appear to be a decisive factor, as humanities experts performed similarly to medical experts who specialized in the respective topics. Moreover, at the individual participant level, no significant differences were found between the 2 expert groups in terms of their proximity to the text's subject matter.

A Chinese study by Ma et al [2], which also examined the identification rate of chatbot-generated texts, reported similar findings, with approximately 66% of texts correctly identified. This study analyzed around 40 scientific texts, including 20 scientific paper abstracts and 20 wiki item descriptions, assessed by 2 PhD students with a background in computer science [2]. Ma et al [2] also highlighted notable differences in writing style between AI-generated and human-written scientific texts, a conclusion that aligns with our findings. In our study, participants primarily based their decisions on text-analytical features, while content errors influenced their judgment in only 3 instances.

The study by Waltzer et al [36], which closely resembles our research in design, reported similar findings. In their study, 140 college instructors were presented with pairs of essays and correctly identified the ChatGPT-generated text 70% of the time. Like our results, Waltzer et al [36] found that neither prior

XSL•FO
RenderX

experience with ChatGPT nor subject-specific expertise—measured by self-reported familiarity with the topic—significantly improved identification accuracy [36]. However, a key difference is that their study analyzed English-language essays written for a psychology program, whereas our research focused on German-language texts authored by medical students [36].

Performance

The evaluation of a text can focus on different levels and aspects, often emphasizing either content or linguistic features. Currently, AI programs such as ChatGPT are recognized for their seemingly perfected linguistic style [13,14]. A notable case at a German university (TU Munich) illustrates this: an essay submitted as part of a Master's application was rejected—and this decision was upheld by a court—on the grounds that it was "too well written," strongly suggesting AI authorship [13]. However, it is important to note that this essay was written in English [13]. While ChatGPT is also proficient in translating languages such as German and Chinese [15], its performance in German differs from English. Research suggests that AI-generated texts tend to be more nuanced and varied in English than in German [16,17]. This discrepancy is likely due to the greater availability of digital data in English, which results in more refined and contextually accurate outputs. Nevertheless, AI language models continuously improve as they interact with users, enhancing their capabilities in non–English languages over time.

Interestingly, the humanities group, despite their focus on linguistic features, identified ChatGPT-generated texts less accurately than the medical expert group—though this difference was not statistically significant. Notably, humanities experts rated the linguistic quality of ChatGPT texts higher than those written by medical students, a contrast that was significant compared with the evaluations of the medical experts. The decision-making process behind text identification revealed key patterns: participants were more likely to suspect a human author when encountering spelling and grammatical errors, greater variation in sentence structure, medical-specific terminology, a writing style aimed at a professional readership, or shifts in citation style.

An "AI author," by contrast, was suspected if there was a monotonous sentence structure, partly "English" grammar, a "smooth" wording style, that is, good readability/understandability, but overall more superficial, an intended less professional readership, better overall formal structure of the text (derivation, outline, weft), frequent repetitions, and a lack of supra-textual coherence of the argumentation in contrast to the coherent and easily comprehensible sequence of arguments within individual paragraphs.

Many studies explored the identifiability of chatbot-generated text using machine learning–based detectors, a subset of AI technologies [27-29]. These detectors often achieve higher identification rates than human evaluators. However, direct human comparison is rarely included, and accuracy and F_1-scores vary significantly depending on the text genre and the specific machine learning classifier used. For instance, when various LLM-based classifiers are applied to different data sets, their accuracy ranges from 70% (DetectGPT classifier on Wikipedia articles) to 97% (GPT-Pat classifier on COVID-19–related question-answer data sets). Similarly, perplexity-based classifiers achieve around 70% accuracy on ACL paper abstracts, whereas RoBERTa (Robustly Optimized BERT Pretraining Approach)–based classifiers reach up to 97% on COVID-19–related data sets [37].

Another challenge is that while these tools are generally reliable in detecting AI-generated text, they are not always sufficiently accurate in identifying human-authored text. This suggests that the tools may struggle with the complexity of human writing while also highlighting a key limitation—especially in cases where a lecturer, for example, must evaluate a single piece of writing without comparison [14,31].

Our study also compares the performance of medical students and ChatGPT. Notably, the texts written by medical students received higher professional evaluations. However, the humanities experts specifically rated the linguistic quality of ChatGPT-generated texts more favorably. Additionally, when comparing ChatGPT's performance with that of medical students in an examination setting—such as in the study by Huh [38]—ChatGPT performed worse than medical students [31]. In a parasitology examination, ChatGPT correctly answered 60.8% of the questions, whereas the average score among 77 medical students was significantly higher at 90.8% [38]. In comparison, a German study by Friederichs et al [39] found that ChatGPT correctly answered two-thirds of all multiple-choice questions at the level of the German state licensing examination in the Progress Test Medicine. It even outperformed most medical students in their first 3 years and performed comparably to students in the later stages of their studies [39]. Our study also revealed that participants who overestimated ChatGPT's writing capabilities and underestimated those of the students were more likely to misidentify the author. This misconception was particularly evident in cases where participants misclassified texts in both sessions, suggesting that their biased perception significantly influenced their decisions.

Interpretation

Our study demonstrates that the identification rate predicted by Turing holds within a group primarily engaged in student teaching and academic writing. Our findings confirm the expectation that linguistic features play a more significant role in identifying AI-generated texts than content familiarity or specialized expertise. In both expert groups, text-analytical features were the primary factors influencing their decisions. This aligns with the emerging field of stylometric analysis, which is increasingly being applied to the detection of AI-generated content [40]. ChatGPT-generated text, especially in comparison to authors from the (fictional) literature domain, exhibits limited stylistic variety [41]. Notably, there was no significant difference in the identification rate between the 2 expert groups, despite 1 being more familiar with the subject matter. Higher proximity to the topic was also not a predictive factor at the individual participant level. Instead, certain linguistic characteristics played a key role in the

XSL•FO
RenderX

decision-making process and were consistently associated with AI-generated texts. In particular, redundancy, repetition, and a lack of coherence were distinctive features attributed to ChatGPT-generated texts. While these traits influenced the perception of AI authorship, they ultimately did not prove to be reliable predictors for correct identification. The linguistic features of ChatGPT-generated texts are often perceived as superior due to their smoother wording and better structural organization. This aligns with findings from [42], which indicate that AI-generated texts tend to exhibit relatively low lexical density, high reading ease, and frequent use of the simple present tense. Whether these linguistic characteristics, if systematically outlined in a manual and provided to participants beforehand, could enhance identification accuracy remains an open question. However, this presents an intriguing avenue for future research.

Limitations

While numerous studies are currently investigating the performance of LLM-based text generators such as ChatGPT, many focus primarily on their assistive role in the writing process rather than assessing the quality of fully generated long-form scientific texts. A key contribution of our study is that it examines complete, AI-generated scientific texts rather than partial outputs. Additionally, instead of relying on specialized AI detection tools, we analyze how individuals working in academia recognize such texts without assistance and how they evaluate their performance while identifying distinct linguistic features. This study enables the compilation of categorical features that could aid in identifying AI-generated text in both academic and everyday reading. However, limitations in generalizability arise due to the relatively small sample size and the exclusive use of a single AI model, ChatGPT version 3.5. Nevertheless, for an exploratory study, we do not consider this a critical issue. Additionally, participants were aware that 1 of the texts had to be AI-generated, raising the question of whether they would have identified an AI-authored text without this prior knowledge. A further limitation arises from the use of different interviewers for the 2 expert groups, who also differed methodologically in terms of blinding. However, it should be noted that the decision to identify the authors was always made before the interview process. Additionally, the interview was transcribed in bullet points, so some information may have been lost in this process. Finally, the dynamic nature of development should also be acknowledged, as ChatGPT, like other AI programs, is continuously being developed and improved.

Conclusion

Our study shows that linguistic and text-analytical features, in particular, play a role in the decision-making process for correctly identifying a chatbot author. In our sample, both nonspecialists and specialists identified AI-generated texts with an accuracy rate of approximately 70% (48/69). Further quasi-experimental studies using texts from other academic disciplines should be conducted to determine whether instructions based on these features can enhance lecturers' ability to distinguish between student-authored and AI-generated work.

A follow-up study could be conducted in a few years to track the evolution of AI-generated text identification and examine whether identification success changes as AI technology and tools advance.

Acknowledgments

The authors thank the students for providing their work. We also extend our gratitude to our colleagues at the Paediatric and Neurological University Clinics of the Ruhr-University of Bochum, as well as the teachers at the Department of Philology, for their voluntary participation, contributions, and suggestions. Additionally, we thank the Medical Dean of Studies, Prof. T. Schäfer, for the initial exchange, and Dr. A. Lucke for her critical endorsement of our work.

Authors' Contributions

BD and JSB contributed equally to conceptualization and data acquisition, with BD leading the qualitative analysis and the original draft writing. CM played a leading role in conceptualization, supervision, and writing—review and editing—while contributing equally to formal analysis and data acquisition. TL and MB shared equal responsibilities in supervision and writing—review and editing. JS contributed equally to conceptualization, qualitative analysis, and writing, whereas EEK was involved in writing—review and editing on an equal basis. SH took the lead in proofreading and contributed equally to review and editing. MT led the formal analysis while sharing equal responsibilities in supervision and writing—review and editing. All authors have reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire with English translation and abbreviation label.
[DOCX File , 15 KB - mededu_v11i1e62779_app1.docx]

Multimedia Appendix 2

XSL•FO
RenderX

Categories for the qualitative analysis, created from the various reasons given in free form for the decision on the authorship of a text.
[DOCX File , 18 KB - mededu_v11i1e62779_app2.docx]

References

1. Jannai D, Meron A, Lenz B, Levine Y, Shoham Y. Human or not? A gamified approach to the Turing Test. arXiv Preprint posted online on January 30, 2023 [FREE Full text]
2. Ma Y, Liu J, Yi F, Cheng Q, Huang Y, Lu W, et al. AI vs. human -- differentiation analysis of scientific content generation. arXiv. Preprint posted online on January 21, 2023 URL: https://doi.org/10.48550/arXiv.2301.10416 [accessed 2023-05-01]
3. von Garrel J, Mayer J, Mühlfeld M. Hochschule Darmstadt. 2023. URL: https://doi.org/10.48444/h_docs-pub-395 [accessed 2024-11-01]
4. ChatGPT: optimizing language models for dialogue. OpenAI. URL: https://openai.com/blog/chatgpt/ [accessed 2023-07-01]
5. Taecharungroj V. “What can ChatGPT do?” Analyzing early reactions to the innovative AI chatbot on Twitter. BDCC 2023 Feb 16;7(1):35 [FREE Full text] [doi: 10.3390/bdcc7010035]
6. Berdejo-Espinola V, Amano T. AI tools can improve equity in science. Science 2023 Mar 10;379(6636):991 [FREE Full text] [doi: 10.1126/science.adg9714] [Medline: 36893248]
7. Marx JPS. ChatGPT im studium: die top 10 befehle für effektives Lernen. Shribe. 2023. URL: https://shribe.de/chatgpt-studium/ [accessed 2024-11-01]
8. Arif TB, Munaf U, Ul-Haque I. The future of medical education and research: is ChatGPT a blessing or blight in disguise? Med Educ Online 2023 Dec 21;28(1):2181052 [FREE Full text] [doi: 10.1080/10872981.2023.2181052] [Medline: 36809073]
9. Choi EPH, Lee JJ, Ho MH, Kwok JYY, Lok KYW. Chatting or cheating? The impacts of ChatGPT and other artificial intelligence language models on nurse education. Nurse Educ Today 2023 Jun;125:105796. [doi: 10.1016/j.nedt.2023.105796] [Medline: 36934624]
10. Flanagin A, Bibbins-Domingo K, Berkwits M, Christiansen SL. Nonhuman "authors" and implications for the integrity of scientific publication and medical knowledge. JAMA 2023 Feb 28;329(8):637-639 [FREE Full text] [doi: 10.1001/jama.2023.1344] [Medline: 36719674]
11. Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, et al. ChatGPT and other large language models are double-edged swords. Radiology 2023 Apr;307(2):e230163 [FREE Full text] [doi: 10.1148/radiol.230163] [Medline: 36700838]
12. O'Connor S. Open artificial intelligence platforms in nursing education: tools for academic progress or abuse? Nurse Educ Pract 2023 Jan;66:103537 [FREE Full text] [doi: 10.1016/j.nepr.2022.103537] [Medline: 36549229]
13. Zenthöfer J. Erstes urteil zu ChatGPT an hochschulen. FAZ. 2024. URL: https://www.faz.net/aktuell/karriere-hochschule/student-nutzt-chatgpt-fuer-bewerbung-erstes-urteil-zu-ki-an-hochschulen-19564795.html [accessed 2024-05-01]
14. Abburi H, Roy K, Suesserman M, Pudota N, Veeramani B, Bowen E, et al. A simple yet efficient ensemble approach for AI-generated text detection. arXiv Preprint posted online on January 23, 2023 [FREE Full text]
15. Jiao W, Wang W, Huang J, Wang X, Tu Z. Is ChatGPT a good translator? A preliminary study. arXiv Preprint posted online on January 14, 2023 [FREE Full text]
16. Richtscheid W. Some hints for the use of “chatGPT” when using with German input language. Medium. 2023. URL: https://medium.com/@walter.richtscheid_93860/notes-on-the-use-of-chatgpt-in-german-language-c4941a14da54 [accessed 2023-07-01]
17. De Vries A. Computergenerierter Zufall als kreatives Moment in Malerei und Literatur. Potentiale und Grenzen von Machine-Learning-Modellen am Beispiel von GPT. In: Lucke A, Alexa H, editors. Literaturwissenschaft und Informatik. Transdisziplinäre Perspektiven, digitale Methoden und selbstlernende Algorithmen. Bielefeld, Germany: Universität Bielefeld; 2024:93-121.
18. Corizzo R, Leal-Arenas S. One-GPT: a one-class deep fusion model for machine-generated text detection. 2023 Presented at: IEEE International Conference on Big Data (BigData); December 15-18, 2023; Sorrento, Italy p. 5743-5752.
19. Parviainen J, Rantala J. Chatbot breakthrough in the 2020s? An ethical reflection on the trend of automated consultations in health care. Med Health Care Philos 2022 Mar 04;25(1):61-71 [FREE Full text] [doi: 10.1007/s11019-021-10049-w] [Medline: 34480711]
20. Alturaiki AM, Banjar HR, Barefah AS, Alnajjar SA, Hindawi S. A smart chatbot for interactive management in beta thalassemia patients. Int J Telemed Appl 2022;2022:9734518 [FREE Full text] [doi: 10.1155/2022/9734518] [Medline: 35601050]
21. Bin Sawad A, Narayan B, Alnefaie A, Maqbool A, Mckie I, Smith J, et al. A systematic review on healthcare artificial intelligent conversational agents for chronic conditions. Sensors (Basel) 2022 Mar 29;22(7):1-20 [FREE Full text] [doi: 10.3390/s22072625] [Medline: 35408238]
22. Cao X, Liu X. Artificial intelligence-assisted psychosis risk screening in adolescents: practices and challenges. World J Psychiatry 2022 Oct 19;12(10):1287-1297 [FREE Full text] [doi: 10.5498/wjp.v12.i10.1287] [Medline: 36389087]
23. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. JMIR Ment Health 2019 Oct 18;6(10):e14166 [FREE Full text] [doi: 10.2196/14166] [Medline: 31628789]

XSL•FO
RenderX

24. Schick A, Feine J, Morana S, Maedche A, Reininghaus U. Validity of chatbot use for mental health assessment: experimental study. JMIR Mhealth Uhealth 2022 Oct 31;10(10):e28082 [FREE Full text] [doi: 10.2196/28082] [Medline: 36315228]
25. Liao W, Liu Z, Dai H, Xu S, Wu Z, Zhang Y, et al. Differentiating ChatGPT-generated and human-written medical texts: quantitative study. JMIR Med Educ 2023 Dec 28;9:e48904 [FREE Full text] [doi: 10.2196/48904] [Medline: 38153785]
26. Khairatun Hisan U, Miftahul Amri M. ChatGPT and medical education: a double-edged sword. Journal of Pedagogy and Education Science 2023 Mar 11;2(01):71-89 [FREE Full text] [doi: 10.56741/jpes.v2i01.302]
27. Hayawi K, Shahriar S, Mathew S. The imitation game: detecting human and AI-generated texts in the era of ChatGPT and BARD. Journal of Information Science 2024 Feb 14;The Imitation Game:1-36 [FREE Full text] [doi: 10.1177/01655515241227531]
28. Orenstrakh MS, Karnalim O, Suarez CA, Liut M. Detecting LLM-generated text in computing education: a comparative study for ChatGPT cases. arXiv Preprint posted online on January 17, 2023 [FREE Full text] [doi: 10.1109/compsac61105.2024.00027]
29. Mitrovic S, Andreoletti D, Ayoub O. ChatGPT or human? Detect and explain. Explaining decisions of machine learning model for detecting short ChatGPT-generated text. arXiv. Preprint posted online on January 11, 2023 URL: https://doi.org/10.48550/arXiv.2301.13852 [accessed 2023-05-01]
30. Fröhling L, Zubiaga A. Feature-based detection of automated language models: tackling GPT-2, GPT-3 and Grover. PeerJ Comput Sci 2021;7:e443 [FREE Full text] [doi: 10.7717/peerj-cs.443] [Medline: 33954234]
31. Elkhatat A, Elsaid K, Almeer S. Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. Int J Educ Integr 2023 Sep 01;19(1):17 [FREE Full text] [doi: 10.1007/s40979-023-00140-5]
32. Turing AM. Computing machinery and intelligence. Mind 1950 Oct 01;LIX(236):433-460 [FREE Full text] [doi: 10.1093/mind/lix.236.433]
33. Schäer R. Einführung in Die Grammatische Beschreibung des Deutschen: Dritte. Berlin, Germany: Language Science Press; 2018.
34. Fobbe E. Forensische Linguistik: Eine Einführung. Tübingen, Germany: Narr Francke Attempto Verlag; 2011.
35. Hessler S. Autorschaftserkennung und Verstellungsstrategien: Textanalysen und-vergleiche im Spektrum forensischer Linguistik, Informationssicherheit und Machine-Learning (Volume 585). Tübingen, Germany: Narr Francke Attempto Verlag; 2023.
36. Waltzer T, Pilegard C, Heyman GD. Can you spot the bot? Identifying AI-generated writing in college essays. Int J Educ Integr 2024 Jul 08;20(1):11. [doi: 10.1007/s40979-024-00158-3]
37. Yu X, Qi Y, Chen K, Chen G, Yang X, Zhu P, et al. LLM paternity test: generated text detection with LLM genetic inheritance. arXiv Preprint posted online on January 30, 2024 [FREE Full text]
38. Huh S. Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: a descriptive study. J Educ Eval Health Prof 2023 Jan 11;20:1 [FREE Full text] [doi: 10.3352/jeehp.2023.20.01]
39. Friederichs H, Friederichs WJ, März M. ChatGPT in medical school: how successful is AI in progress testing? Med Educ Online 2023 Dec;28(1):2220920 [FREE Full text] [doi: 10.1080/10872981.2023.2220920] [Medline: 37307503]
40. Uzun L. ChatGPT and academic integrity concerns: detecting artificial intelligence generated content. Language Education & Technology. 2023. URL: https://www.langedutech.com/letjournal/index.php/let/article/view/49/36 [accessed 2024-02-01]
41. Zwilling L, Berger M. ChatGPT does not speak style!. 2024 Presented at: Computational Linguistics Poster Session at the 46th Annual Meeting of the German Linguistic Society (DGfS); February 28, 2024; Bochum, Germany.
42. AlAfnan MA, Mohdzuki SF. Do artificial intelligence chatbots have a writing style? An investigation into the stylistic features of ChatGPT-4. Journal of Artificial Intelligence and Technology 2023 May 12;3:3 [FREE Full text]

Abbreviations

AI: artificial intelligence
LLM: large language model
RoBERTa: Robustly Optimized BERT Pretraining Approach
RUB: Ruhr University Bochum

XSL•FO
RenderX

Edited by B Lesselroth; submitted 31.05.24; peer-reviewed by C Bach, L Zhu, R Corizzo; comments to author 12.08.24; revised version received 28.11.24; accepted 16.01.25; published 03.03.25.

Please cite as:
Doru B, Maier C, Busse JS, Lücke T, Schönhoff J, Enax- Krumova E, Hessler S, Berger M, Tokic M
Detecting Artificial Intelligence–Generated Versus Human-Written Medical Student Essays: Semirandomized Controlled Study
JMIR Med Educ 2025;11:e62779
URL: https://mededu.jmir.org/2025/1/e62779
doi:10.2196/62779
PMID:40053752

©Berin Doru, Christoph Maier, Johanna Sophie Busse, Thomas Lücke, Judith Schönhoff, Elena Enax- Krumova, Steffen Hessler, Maria Berger, Marianne Tokic. Originally published in JMIR Medical Education (https://mededu.jmir.org), 03.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

AIFM-ed Curriculum Framework for Postgraduate Family Medicine Education on Artificial Intelligence: Mixed Methods Study

Raymond Tolentino[1], BHSc, MSc; Fanny Hersson-Edery[1], MD; Mark Yaffe[1,2], MD; Samira Abbasgholizadeh-Rahimi[1,3,4,5], BEng, PhD

[1]Department of Family Medicine, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada

[2]Department of Family Medicine, St. Mary's Hospital Center, Integrated University Centre for Health and Social Services of West Island of Montreal, Montreal, QC, Canada

[3]Mila-Quebec, Montreal, QC, Canada

[4]Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, QC, Canada

[5]Faculty of Dental Medicine and Oral Health Sciences, McGill University, Montreal, QC, Canada

Corresponding Author:
Samira Abbasgholizadeh-Rahimi, BEng, PhD
Department of Family Medicine
Faculty of Medicine and Health Sciences
McGill University
5858 chemin de la Côte-des-Neiges
Montreal, QC, H3S 1Z1
Canada
Phone: 1 514 399 9218
Email: samira.rahimi@mcgill.ca

Abstract

Background: As health care moves to a more digital environment, there is a growing need to train future family doctors on the clinical uses of artificial intelligence (AI). However, family medicine training in AI has often been inconsistent or lacking.

Objective: The aim of the study is to develop a curriculum framework for family medicine postgraduate education on AI called "Artificial Intelligence Training in Postgraduate Family Medicine Education" (AIFM-ed).

Methods: First, we conducted a comprehensive scoping review on existing AI education frameworks guided by the methodological framework developed by Arksey and O'Malley and Joanna Briggs Institute methodological framework for scoping reviews. We adhered to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist for reporting the results. Next, 2 national expert panels were conducted. Panelists included family medicine educators and residents knowledgeable in AI from family medicine residency programs across Canada. Participants were purposively sampled, and panels were held via Zoom, recorded, and transcribed. Data were analyzed using content analysis. We followed the Standards for Reporting Qualitative Research for panels.

Results: An integration of the scoping review results and 2 panel discussions of 14 participants led to the development of the AIFM-ed curriculum framework for AI training in postgraduate family medicine education with five key elements: (1) need and purpose of the curriculum, (2) learning objectives, (3) curriculum content, (4) organization of curriculum content, and (5) implementation aspects of the curriculum.

Conclusions: Using the results of this study, we developed the AIFM-ed curriculum framework for AI training in postgraduate family medicine education. This framework serves as a structured guide for integrating AI competencies into medical education, ensuring that future family physicians are equipped with the necessary skills to use AI effectively in their clinical practice. Future research should focus on the validation and implementation of the AIFM-ed framework within family medicine education. Institutions also are encouraged to consider adapting the AIFM-ed framework within their own programs, tailoring it to meet the specific needs of their trainees and health care environments.

(JMIR Med Educ 2025;11:e66828) doi:10.2196/66828

KEYWORDS

artificial intelligence; family medicine; curriculum; framework; postgraduate education

XSL•FO
RenderX

Introduction

The College of Family Physicians of Canada (CFPC) establishes standards for postgraduate family medicine training and its accreditation [1]. It promotes a competency-based curriculum model known as Triple C (comprehensive, continuous, and centered in family medicine) [2] based on the Canadian Medical Education Directives for Specialists (CanMEDS)—Family Medicine framework [3] and on assessment objectives for certification in family medicine [4]. To ensure that medical curricula respond to new developments in health care, education, and societal trends, they must undergo periodic review, modification, and renewal [5-9]. Accordingly, a number of new content areas have been introduced in the recent past into the family medicine curricula. They include leadership [10], social determinants of health [11], ethics [12], global health [13-15], and physician wellness and burnout [16-18]. The increasing complexity of the medical needs of an aging population, the exponential growth in medical knowledge, and an increasingly digitalized environment suggest the need for digital-mediated solutions to support medical practitioners.

Artificial intelligence (AI) and its applications have made a rapid impact on many segments of society, including medicine [19] and notably, in primary health care [20]. While there is no universal consensus on the definition of AI, the World Health Organization [21] describes it as "the performance by computer programs of tasks that are commonly associated with intelligent beings." The introduction, integration, and implementation of AI-based tools and systems into family medicine education and practice assume an adequately trained cohort of users, but to date, training of family physicians on relevant aspects of AI to ensure effective and safe implementation has been absent or inconsistent [20,22]. As such, the CFPC's Outcomes of Training project has identified digital care and health informatics as a training gap and an area for educational enhancement requiring priority attention across the 17 family medicine postgraduate programs in Canada [23,24]. There have been efforts to include AI education globally within each level of medical training. These efforts are led by national medical associations such as the UK National Health Service, the US American Medical Association, and Canada's Royal College of Physicians and Surgeons. They have released documents recommending policies for integrating AI within their respective medical educational institutions [25-27].

Initiatives of AI teaching directed at physicians already in practice include the development of a continuing professional development 3-module CFPC Learn e-course titled, "Artificial Intelligence for Family Medicine" [28]. The first module of this course reviews the basic functionality of AI with applications in family medicine, while the second module focuses on core terminology and related concepts as well as potential harms or risks associated with AI. The last module reviews the concepts of the first 2 and focuses on learning how to tell if an AI-based tool is working well [28].

Competency about a particular subject has been described as the ability to carry out a certain task or action at a basic or acceptable level [29]. Liaw et al [30] have recently proposed six competency domains for family medicine training in AI: (1) foundational knowledge (What is this tool?), (2) critical appraisal (Should I use this tool?), (3) medical decision-making (When should I use this tool?), (4) technical use (How do I use this tool?), (5) patient communication (How should I communicate with patients regarding the use of this tool?), and (6) awareness of unintended consequences (What are the "side effects" of this tool?)." These authors suggest that such competencies can be integrated within current residency training during existing sessions on health informatics or evidence-based medicine but emphasize that these competencies are a "point of departure" and must be further worked on [30].

A curriculum framework can be described as "a core policy document that describes a range of requirements, regulations and advice which should be respected by all stakeholders in the education system, and which should guide the work of schools, teachers and the developers of other curriculum documents" [31]. Curriculum frameworks allow for a visual and detailed roadmap to develop and implement a curriculum [32]. Input from an interdisciplinary team of medical educators, AI experts, end users, researchers, and curriculum designers [33] can effectively support the development of a curriculum framework for teaching AI in family medicine postgraduate training programs. Our comprehensive review of the available curriculum frameworks [34,35] highlighted that there is no framework designed specifically for family medicine residency and no paper that described a systematic approach to design one. From the 2 frameworks uncovered, one framework was incomplete, while the other framework was brief and focused on ophthalmology [34]. The ophthalmology curriculum framework lacks adaptability, as it may prove inadequate for family medicine residency due to the diverse, community-based nature of family medicine, which differs significantly from the highly technological and hospital-based focus of ophthalmology.

Considering the gaps mentioned previously and the foundational importance of curriculum frameworks in the creation of new educational structures, our objective was to design and develop a curriculum framework for AI family medicine education, that is, Artificial Intelligence Training in Postgraduate Family Medicine Education (AIFM-ed), ensuring alignment with current competencies and educational goals. To achieve this, a combination of validated methods including 2 national expert panel discussions were conducted, supplemented by a previous comprehensive systematic scoping review [34]. Developing a framework based on expert insights would help address gaps in AI education and provide an adaptable guide for family medicine educators, curriculum designers, postgraduate residency program directors, medical education researchers, and policy makers in health care education. Due to the systematic approach in designing this framework, audiences can adopt this framework to other fields and specialties, considering that our review did not find any systematically developed frameworks.

XSL•FO
RenderX

Methods

Study Design

For the construction of an AIFM-ed framework, we followed the analysis, design, development, implementation, and evaluation model for instructional design, using the first 3 activities to guide our work. We followed a two-step approach suggested by Redwood-Campbell et al [36] for framework development, wherein (1) a review of the literature was made focusing on curriculum frameworks and core competencies for AI education in medicine [34,35] and (2) a working group used qualitative or consensus methods for final development of the framework.

Our scoping review aimed to synthesize knowledge from the literature on curriculum frameworks and current educational programs that focus on the teaching and learning of AI for medical students, residents, and practicing physicians, and adhered to PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines. Details of this comprehensive study have been published elsewhere [34,35]. Our review specifically identified several AI educational curricula programs (eg, courses, workshops, webinars, and projects) and 2 curriculum frameworks for AI education, one outlining a broad framework for any level of education [37], while the other described a complete framework for ophthalmology residency education [38].

The outcome of our review was the identification of early concepts that could be applied to elements of the curriculum framework for family medicine and AI [34,35]. This initial curriculum framework was later used during the panel discussion as part of the co-development and redesigning of the framework. This discussion applied the curriculum framework structure described by Obadeji [39], which examines six common elements: (1) the need and the purpose of a curriculum or a program, (2) learning objectives and outcomes, (3) course content that will facilitate the accomplishment of the objectives or learning outcomes, (4) organization of the content, (5) implementation of curriculum, and (6) curriculum evaluation and refinement. This study examined all elements except the final element (curriculum evaluation and refinement). The initial framework was deemed successful by the expert team based on the following indicators: relevance to medical educators and curriculum designers, alignment to current family medicine competencies and educational goals, clarity of AI-specific content, and its potential for further validation. However, we acknowledge that further studies are needed.

Qualitative Methodology

The expert panel methodology follows the SRQR (Standards for Reporting Qualitative Research) checklist [40]. Expert panels help to attempt to reach consensus on controversial subjects [41,42] such as the risk of AI tools leading to reduced proficiency in independent critical thinking and clinical judgment among physicians. The use of qualitative consensus methods for curriculum development facilitates input from a wide range of stakeholders (eg, physicians and curriculum developers) in order to assess and validate expert knowledge [43]. The use of expert panel discussions to assist in creating curricula has become established in pedagogical research and development [44]. Examples within the field of medicine include discussions around social determinants of health for undergraduate medical education [45], telemedicine opportunities for postgraduate medical education [46], and geriatric oncology in continuing medical education [47].

Participant Recruitment and Sampling Strategy

Our panel size fell within the recommended average of 8 members or a median of 6 [48]. The definition of an expert in our case is flexible due to the limited knowledge and experience on this emerging topic; this is emphasized by Duncan et al [49], who state that, “[t]oo narrow a definition, however, can restrict the number of potential participants.” In our case, we chose experts according to the definitions of Fink et al [41], which state that they must be, “representative of their professional group, with either sufficient expertise not to be disputed or the power required to instigate the findings.” This was reinforced by Mead and Mosely [50], which state that, “experts can be defined in a number of ways, such as their position in a hierarchy [...] or as recommended by other participants in a study.” Therefore, from these definitions, we selected panelists based on their academic qualifications, their number and relevance of AI-related publications, professional experience within the development, implementation or research of AI, and finally, any participation in AI-specific projects or conferences.

For panel 1, we reached out to family medicine (clinical) educators from affiliated universities and professional organizations across Canada via email. Snowballing by this initial group generated the names of others known as family medicine educators. For panel 2, family medicine residents were invited from an initial group of residents who were knowledgeable and aware of AI, and that initial group helped to recruit relevant residents for this study.

Each participant voluntarily participated in the study by providing their explicit consent and agreement, which was confirmed through email correspondence. To uphold confidentiality, data were safeguarded through limited, secure data access, the disposal of audiotapes after transcription, and the anonymous analysis of transcripts.

Ethical Considerations

This study involved a panel discussion with experts, which does not require formal ethics board approval under the Economic and Social Research Council Framework for Research Ethics guidelines [51]. According to these guidelines as well as Canada’s Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans, research that presents minimal risk and does not involve sensitive information may be exempt from formal ethics review [52]. This study adhered to these recognized guidelines, ensuring that all participants were treated in accordance with principles of research integrity, voluntary participation, and informed consent.

Participant Eligibility Criteria

Input from 2 different types of panelists was desired, and they were included as participants within 2 distinct expert panels.

XSL•FO
RenderX

The first included family medicine educators practicing in Canada who were somewhat knowledgeable or have expertise in AI education. No limitations were placed on years of practice experience, years of knowledge or experience in AI, language proficiency, work setting, or the types of patients for whom they provided care. The second expert panel included participants who were at the time of the study family medicine residents at McGill University, and who were somewhat knowledgeable in AI. No limitations were placed on language proficiency, years of knowledge or experience in AI, work settings, or the types of patients they provided care for.

Data Collection

We conducted a recorded session of each expert panel via Zoom (version 5.16.10; Zoom Video Communications). The use of a web-based expert panel minimizes costs associated with travel; it also mitigates potential biases linked to panelists [53]. Each web-based expert panel discussion was approximately 2 hours long, followed the same format, used congruent discussion guides, and was facilitated by 2 members of the research team (RT and SAR). The discussions began with a brief presentation given by RT on the results of the first step of the project, that is, the comprehensive scoping review in the field [34,35]. Following the presentation, each of the five elements of the curriculum framework: (1) the need and the purpose of a curriculum or a program, (2) learning objectives and outcomes, (3) course content that will facilitate the accomplishment of the objectives or learning outcomes, (4) organization of the content, and (5) implementation of the curriculum were discussed sequentially and at length. When presenting each element, participants were invited to respond and discuss their opinions and thoughts related to each element, allowing for the co-development and redesigning of the framework together.

Data Analysis

Expert panel discussion data were analyzed using content analysis strategies [54,55] as previously used in a study developed for a training model for nurses using a literature review and expert panel discussions, in which data were analyzed using a descriptive qualitative approach that includes content analysis [56]. In our work, the preparation phase included transcribing the data, immersing in the data, and obtaining a sense of whole through reading the transcript multiple times. In our study, once the recordings from the expert panel discussions were received, one of the authors (RT) listened to the entire recording and subsequently transcribed it verbatim. The next stage of data analysis was the organizing phase, in which open coding and the creating of categories were conducted along with the grouping of codes under higher-order headings. These were carried out by one of the authors (RT) and verified by the senior author (SAR).

As the analysis of data used an inductive approach, no prior coding systems were used, such that coded categories were derived directly from the data [55]. Sentences and phrases from the panelists were captured. In vivo coding was used to prioritize participants' language and perspectives, while descriptive coding aided in categorizing key themes. Two independent coders reviewed the data (RT and SAR), with discrepancies resolved through discussion between coders and the research team. Saturation was achieved when no new themes emerged during the coding of the final transcript. The final step included the presentation of the final curriculum framework, which resulted from the incidence of codes and categories and its relation to the literature. Codes and categories derived were prioritized and highlighted with how frequently they appeared during the panel discussion as well as the overlap between both groups. These highlighted findings were then compared with existing literature to either support or challenge them. If these codes and categories were supported by the literature, they were subsequently integrated into the framework.

Results

Panelists Characteristics

A total of 37 educator and resident experts were invited, 14 for the educator group and 23 for the resident group. Ultimately, 8 from the former and 6 from the latter group participated, for a total of 14 participants. Scheduling problems were the most common reasons for nonparticipation. The characteristics of those included in the expert panel discussion are displayed in Table 1.

Table 1. Characteristics of expert panel participants included.

	Educator experts (n=8), n (%)	Resident experts (n=6), n (%)
Sex		
Male	3 (38)	4 (66)
Female	5 (62)	2 (33)
Educational background		
Doctoral (PhD)	7 (88)	0 (0)
Master	1 (22)	2 (33)
Bachelor or MD only	0 (0)	4 (66)
Affiliation		
McGill University	5 (62)	6 (100)
Other academic institution	3 (38)	0 (0)

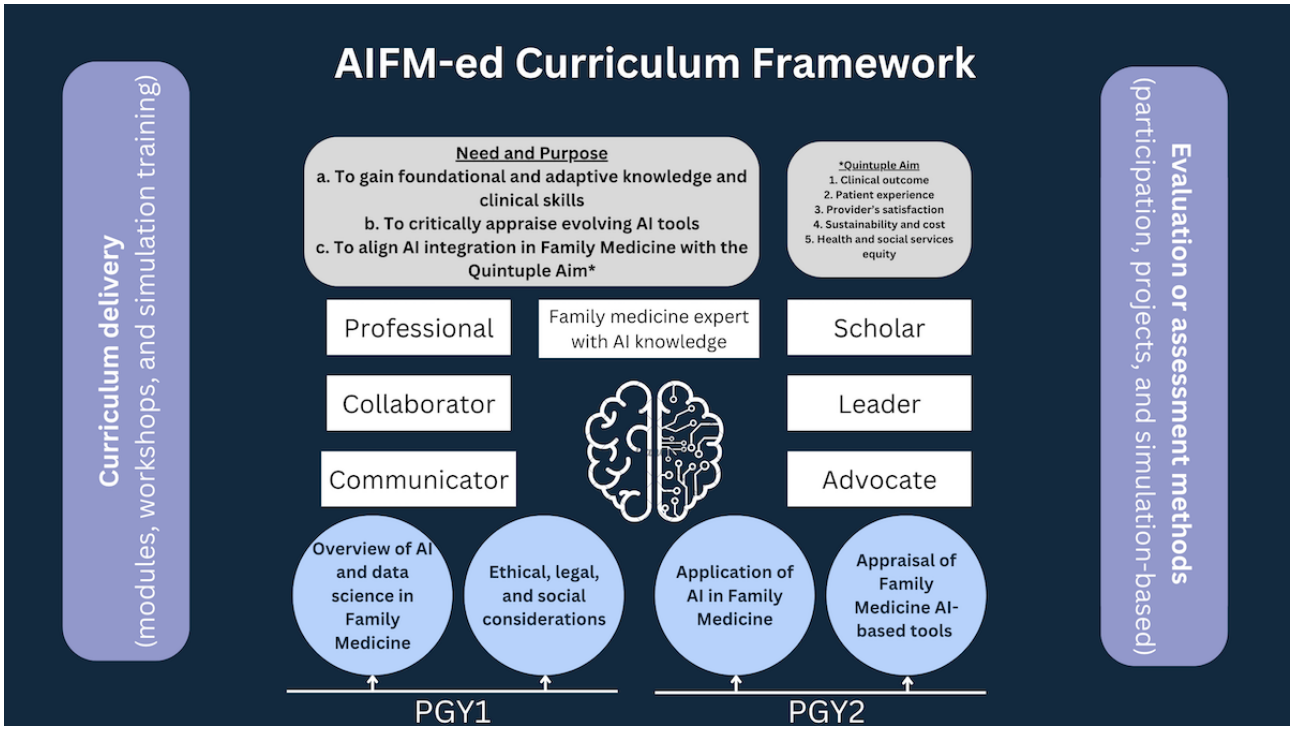
XSL•FO
RenderX

Curriculum Framework for AIFM-ed

Overview

Our project has identified five elements of the curriculum framework for AI training in postgraduate family medicine education: (1) need and purpose of the curriculum, (2) learning objectives, (3) curriculum content, (4) organization of curriculum content, and (5) implementation of the curriculum. A condensed visual representation of the AIFM-ed curriculum framework is displayed in Figure 1, while each element is discussed in detail below.

Figure 1. Representation of the AIFM-ed curriculum framework. AI: artificial intelligence; AIFM-ed: Artificial Intelligence Training in Postgraduate Family Medicine Education; PGY: postgraduate year. * refers to the Quintuple Aim.



Element 1: Need and Purpose of the AIFM-ed Curriculum

When modifying a curriculum in family medicine postgraduate training, it is important to understand why it must be changed and for what purpose. Both panels discussed the current low priority of AI curricula. Residents emphasized a lack of exposure in training and practice. Both panels agreed that the integration of an AI curriculum will inevitably become imperative, recognizing its potential as an essential toolset in practice. One educator summarized this thought by saying, "AI will continue to evolve quickly, so a curriculum must be built for the future."

To describe the need and purpose for AI education in family medicine, we co-developed the following: "The purpose of an AI curriculum for family medicine residents is for future family physicians to: (a) gain foundational and adaptive knowledge and clinical skills, (b) to critically appraise evolving AI tools, and (c) to align AI integration in Family Medicine with the quintuple aim for health care improvement (i.e., improving population health, improving the provider and patient experience, reducing costs, and advancing health equity [57,58])."

Using various definitions of AI, the educator panel debated what constitutes AI specifically in family medicine. The term "AI-based tools" is used throughout the results of this paper as a way of describing technologies empowered or enabled with AI algorithms to support clinical practice. This term has been used in previous literature on AI in the context of family medicine training [20,30].

Element 2: AIFM-ed Learning Objectives

Learning objectives are statements that describe significant and essential learning that learners need to be familiar with and reliably demonstrate at the end of a course or educational program [59]. The following outlines the learning objectives for AI training, aligning with CanMEDS and family medicine roles. Table 2 presents each CanMEDS role on the left column [3], with their affiliated learning objectives for AI family medicine education as structured by participants during the panel on the right column. Although the learning objectives are comprehensive and their practical application for most family doctors may be limited, they are ideal for advancing the knowledge and skills of AI-empowered family physicians.

XSL•FO
RenderX

Table 2. Learning objectives discussed during panels about artificial intelligence (AI) in relation to Canadian Medical Education Directives for Specialists (CanMEDS) roles.

CanMEDS roles	The learner engaged in AI education will be able to
Family medicine expert with AI knowledge	
Family physicians are skilled generalists who should be able to understand and use technology including AI tools to provide high-quality, responsive, community-adaptive care across the lifecycle, from prevention to palliation, in multiple settings, and for diverse populations.	• Explain a basic understanding of AI and basic concepts in relation to family medicine. • Demonstrate the use of AI-based tools for family medicine by showing how to use the tool and understand the output. • Critique and decide on when to use an AI-based tool over another health care resource. • Recognize AI-based tools' perceived biases and discriminatory behavior (eg, an AI-based tool diagnosing skin conditions mainly trained on images of lighter skin tones may be less accurate in detecting conditions in individuals with darker skin tones) and the results demonstrated by AI-based tools where the learner will be able to solve and prevent further effects.
Communicator	
Family physicians foster life-long therapeutic relationships with patients and their families. This incorporates the dynamic exchanges that occur before, during, and after the medical encounter that facilitates gathering and sharing essential information for effective patient-centered health care [3].	• Explain to patients the current AI-based tool they are using and its results. • Address relevant gaps of understanding of AI tools among patients such as differing cultural perspectives and digital health literacy.
Collaborator	
Family physicians work with patients, families, communities, and other health care providers to provide safe, high-quality, patient-centered care [3].	• Practice a collaborative team-based approach, including establishing positive and continuing working relationships with relevant stakeholders in relation to developing, implementing, and improving the quality of AI-based tools.
Leader	
Family physicians must actively contribute to implementing and maintaining a high-quality health care system and take responsibility for delivering excellent patient care. This includes prioritizing and using health care resources efficiently, executing tasks collaboratively with colleagues, and contributing to ongoing quality improvement initiatives within their own practice and its management [3].	• Identify which AI-based tools are appropriate for the clinical practice of family physicians. • Allocate AI-based tools, when available, to specific tasks (eg, administrative work) in order for optimal patient care and practice management. • Analyze incidents of use of AI-based tools, appraise AI-based tools, and resolve any issues to avoid patient injury.
Advocate	
Family physicians leverage AI-driven insights to advocate for patients and communities, using their expertise to identify health needs, drive meaningful change, and mobilize resources for improved care outcomes [3].	• Extend AI-based tools and resources, when available and known, with other family physicians and family medicine communities. • Advocate for established AI-based tools, when available, to patients with the aim of improving their health outcomes.
Scholar	
Family physicians demonstrate a lifelong commitment to excellence in practice through continuous learning and teaching others; gather, combine, and evaluate evidence; and contribute to the creation and dissemination of knowledge [3].	• Participate in scholarly activities related to AI that benefit professional growth, clinical practice, and patients. • Maintain or enhance one's own knowledge and skills through professional educational activities related to AI and ongoing self-directed learning.
Professional	
Family physicians are committed to the health and well-being of their patients and society through competent medical practice; accountability to their patients, the profession, their colleagues, and society; profession-led regulation; ethical behavior; and maintenance of personal well-being [3].	• Recognize and appropriately respond to ethical, legal, and social issues encountered in practice, as it relates to AI-based tools and family medicine by communicating to the proper channels and resources (eg, AI and data experts, information technology specialists, ethics boards, and lawyers).

Element 3: AIFM-ed Curriculum Content

When developing a curriculum, a crucial task is to identify relevant subject knowledge, skills, attitudes, and behaviors that will help form the learning objectives [39]. Currently, there is no required AI education in Canadian undergraduate medical education. However, both educators and residents in our study agreed that for AI to be efficiently introduced in family medicine residency, it must be preceded by education in undergraduate

XSL•FO
RenderX

medical education. This earlier introduction of principles and concepts of AI will facilitate learning the more difficult material that is to come. The panels envisaged a basic stream of education in residency for those who had no exposure in undergraduate years. This would address fundamentals and basic knowledge of AI (eg, history, AI model development process, and core algorithms). A more advanced stream of AI education for residents would summarize the fundamentals and focus on how to use AI-based tools (applications) along with how to decide when to use and evaluate them (critical appraisal).

Residents noted that understanding how AI-based tools are used in clinical practice was the preferred content area for study, with less attention devoted to ethical, legal, and social considerations of AI. A resident put this in context, noting that they "do not need or want to learn the history of ChatGPT, but rather how to write effective prompts within this natural language processing chatbot." Table 3 summarizes the key concepts and areas of interest that family physicians should learn and content to include in the curriculum, as viewed by the participants.

Table 3. The curricular concepts and topics of relevance to family physicians.

Main curricular topic	Subtopics
Overview of AI[a] and data science in family medicine	
Providing an overview of AI definitions and concepts including machine learning as well as topics related to data science (eg, mathematics and statistics) and clinical epidemiology for family medicine.	• Review of AI (definitions and concepts) as it relates to family medicine • Introduction to AI and fundamentals of data science in family medicine • Strength and limitations of AI-based tools
Ethics, legal, and social considerations	
Understanding the ethical, legal, and social concerns of AI as it impacts family medicine clinical practice.	• Ethics, patient rights, data security, and confidentiality • Liabilities and regulatory and policy considerations • Equity, diversity, and inclusion of AI
Application of AI in family medicine	
Understanding how to choose and engage with AI-based tools in clinical settings and workflows with the ability to understand, interpret, and apply results of AI systems in clinical practice.	• Clinical practice management and operation • Preventative care and risk profiling (eg, mental health and chronic disease) • Patient self-management • Physician decision support • Physician wellness and resilience • Social determinants of health
Appraisal of family medicine AI-based tools	
Assessing and reviewing AI-based tools to ensure safe and effective integration and use in clinical practice.	• Identification of potential AI adverse effects and potential solutions • Quality improvement

[a]AI: artificial intelligence.

Element 4: Organization of AIFM-ed Curriculum Content

Family medicine postgraduate training is 24 months long in Canada. Given that the current curriculum is considered very heavy, educators and residents emphasized that the addition of another competency could be a burden to both educators and resident learners. They nonetheless agreed that AI curricula will eventually need to be added to that and an organized teaching structure would need to be established. Residents favored incorporating the teaching within the existing, already tight, 24-month core teaching, so that the benefits of longitudinal learning could be taken advantage of. The educators saw AI knowledge–based training during the first postgraduate year, followed by the development of AI-based clinical skills in the second postgraduate year. Educators proposed that if deeper AI education is needed, an additional third-year training program could be introduced for a select group of interested trainees to develop advanced AI skills in family medicine.

Element 5: Implementation of AIFM-ed Curriculum

Curriculum implementation will require the identification of appropriate resources (eg, educators and materials) along with educational strategies that will facilitate teaching activities and learner evaluation.

Curriculum Delivery

Residents highlight that AI education must be longitudinal, as it must be built upon throughout the medical education continuum. Furthermore, educators emphasized that residency is student-centered with learners coming from diverse backgrounds where they must replicate the actual tasks performed during in practice. Therefore, the learning theory of constructivism appears to be a sound and advantageous choice. This learning theory posits that learners actively construct their own learning by drawing upon their prior experiences [60].

There are several methods to implement an AI education curriculum to family medicine residents; however, there are

XSL•FO
RenderX

certain methods that are recommended by both educators and residents. In terms of learning about the knowledge and background of AI (eg, review of AI concepts or the ethical, legal, and social considerations of AI), hybrid (web-based and in-person) courses with asynchronous web-based modules, and in-person workshops, problem-solving sessions could be applied. Residents emphasized that didactic large group lectures especially in regard to a novel topic such as AI would be less engaging. The learning of such content should be considered a refresher with emphasis on the context of AI in family medicine. Both educators and residents then suggest that the in-person sessions would serve as a space for questions and answers and problem-solving activities.

To execute these educational methods, human resources (eg, AI medical educators) and material resources (eg, existing AI-based tools) are pertinent. Educators and residents highlighted that experts in the field of AI and family medicine would be ideal; however, educators emphasized that the faculty challenges such as the current number of experts are limited to provide this education. To overcome this, residents suggested that once an AI curriculum is established, further educators could be sourced from recently graduated residents who completed the AI in family medicine curriculum. With respect to material resources such as family physician–focused AI-based tools, both groups emphasized that they must be validated before use in educational settings.

Assessment and Evaluation Methods

Residents emphasize that the assessment and evaluation methods for the curriculum should be simple in context and focus on learners' participation and exposure. More specifically, learners should be able to have the capacity to demonstrate how to use AI-enabled tools and techniques in a health care setting. This can be seen through the completion of projects and problem-based and simulation-based assessments. Educators on the other hand emphasized taking into account Kirkpatrick's 4 levels of training evaluation model [61], where assessments should be directly related to the activity's learning objectives.

Discussion

The First Curriculum Framework for AI in Family Medicine (AIFM-ed)

In this study, we introduced a novel and evidence-based initial curriculum framework, that is, AIFM-ed developed for AI literacy education in family medicine postgraduate training. This systematically co-developed framework used a combination of validated methods including a comprehensive scoping review, resident and educator panel discussions, and the involvement of interdisciplinary experts in the field. During the development and cocreation of this framework, several key findings emerged. These include the crucial role of multiple resource partners and innovative practices when integrating AI educational content in family medicine education. For example, AI technology vendors specializing in health care, upcoming startups, and AI-focused organizations.

Furthermore, educators and residents stressed the importance of learning about the application of AI-based tools and simulating their use as a method of learning. Several innovative practices have already been implemented including case-based learning and flipped classroom models. Moreover, the adoption of AI-based tools can be diverse depending on its context (eg, teaching and learning and clinical practice) with several barriers and enablers. Additionally, the study identified several challenges in effectively integrating an AI curriculum framework into existing educational structures. These include the lack of AI definition standardization, the reduced urgency in practice due to the lack of time and resources, as well as the capacity to balance theoretical and practical curricular content.

Interprofessional Collaboration and Resources

During the development of the AIFM-ed curriculum framework, several resource partners were identified when discussing the implementation of AI education in family medicine. Interprofessional collaboration within multidisciplinary teams is essential in order for an AI curriculum to be effective [62]. Other researchers emphasize this sentiment when listing their recommendations of ensuring a responsible integration of AI technologies in medical education [63]. This multidisciplinary team and resource partners may include several stakeholders such as nurses, social workers, epidemiologists, AI experts, data engineers, software developers, and patients [64]. Other resource partners identified included AI technology vendors specializing in health care, upcoming startups, and AI-focused organizations. Residents brought up the concept that AI-based tools and AI in general will substantially change in the future (eg, improved tools, systems, and integrations) and thus stressed the importance of continuous partnerships with other professionals in order for relevant information and AI tools.

Educators emphasized that they were unaware of many AI-based tools for patient support and were thus apprehensive in advocating for AI-based tools. Therefore, family physicians and other primary care team members (eg, administrative staff and nurses) should share AI-based tools and resources, when available and known, with other family physician and family medicine communities. Additionally, residents have suggested that before advocating or suggesting AI-based tools, a list of recommended AI-based tools must be developed and released from a medical organization such as the CFPC. Currently, there is a scoping review and inventory that has identified and evaluated published studies that have tested or implemented AI in primary care settings [20,65]. This can be a starting point for such a list of recommended AI-based tools.

Application and Simulation of AI-Based Tools

Both educators and residents emphasize that a curriculum should focus on how to use AI-based tools (application) along with how to decide when to use and evaluate them (critical appraisal). Residents are already doing this comparatively as seen through their discussions of using ChatGPT, an AI-based chatbot launched by OpenAI that can be used as a digital consultant (eg, simple inquiries about diagnoses and treatment plans). One resident stressed that although they use ChatGPT at times for inquiries related to patient care, they are cautious of the information, as they are aware that ChatGPT can make mistakes and always consult other resources. As ChatGPT rises in

XSL•FO
RenderX

prominence, its impact on medical education has been evident through the resident panel discussion and the literature [66,67].

The incorporation of AI content in medical education has already begun with innovative practices, which include case-based learning and flipped classroom models. Case-based learning incorporates real-world AI use cases, where AI is used in clinical practice as examples for physicians [68]. Through this learning approach, students have a better understanding of the technical aspects of AI, as it allows physicians to compare their thought processes with other students and critically reflect or challenge their assumptions and biases of AI and clinical practice [68]. One study assessed the capabilities of ChatGPT within the framework of a preclerkship case-based active learning curriculum. Although the AI chatbot is not comprehensive enough to serve as a textbook, it was shown to answer questions, generate test questions, and appropriately respond to prompts in case-based learning scenarios [69]. According to a scoping review of teaching AI ethics in medical education, 5 publications reported in using case-based learning when understanding ethical challenges [70]. Resident panelists believe that simulation of these tools is beneficial, as it allows residents to enjoy the learning process and realize how these AI-based tools would operate in actual clinical settings. During these simulation sessions or case-based learning approaches, educator panelists highlighted reviewing the capabilities and basic functions of AI-based tools.

Another practical example of incorporating AI content through innovative practices is the flipped classroom model approach. Flipped classroom models can consist of web-based content supplemented by in-person classroom sessions [71], a key observation reinforced by residents of the panel discussion. One study designed and evaluated a novel AI course for medical students using a flipped classroom model, and they found that attending the course can increase self-perceived AI readiness in medical students [71]. In addition, educators have also commented on facilitating AI learning by integrating family medicine AI-based tools in quality improvement projects, which has been emphasized and recommended by other researchers [72].

Adopting AI in Education and Clinical Practice

Family physicians use AI, when implemented, primarily for diagnosis, detection, or surveillance purposes [20]. Although educators have flexibility in choosing from a wide range of AI tools, certain tools have proven to be particularly essential for effective integration. These include AI-enabled chatbots, clinical documentation support, and diagnostic decision support, which have shown to improve physicians' efficiency and accuracy in their work [73-75]. However, there have been several barriers identified in previous reviews, which have made the adoption of AI-based tools difficult [76-78]. These issues include a lack of trust among educators, students, and clinicians; insufficient training and digital literacy; and resistance to change [77].

Additional challenges include data privacy and patient safety concerns, ethical and legal issues, interoperability issues, lack of funding, and inequities in access to AI tools—particularly between rural and urban settings [79]. In contrast, several strategies and enablers have been identified in order to better facilitate the adoption of AI and its continued use. These strategies include strategies fostering interdisciplinary collaboration between educators, clinicians, and AI developers; providing targeted training programs to build AI literacy; developing high-quality datasets for diverse use cases; and creating supportive regulatory frameworks [77]. Establishing national or local community networks to share resources and best practices, while leveraging trusted relationships within these networks, can also significantly enhance confidence in and adoption of AI-based tools. To identify relevant enablers and barriers to AI adoption of a certain audience, a comprehensive, stakeholder-centered approach is essential. For example, researchers in Canada conducted in-depth interviews with primary health care and digital health stakeholders and were able to ascertain their current barriers and potential facilitators of AI [80].

It is important to note that AI systems exist in diverse contexts and content with distinct implications, risks, and ethical and legal challenges depending on their application and domain. For example, in education, AI-enabled tools using large language models may offer personalized education, but biases may be propagated, inaccurate information may be generated, or students may overrely on AI, undermining their critical thinking skills [63,81]. In addition, there is potential for the exacerbation of inequities in accessing AI tools as well as the misuse of AI-generated content. In comparison, AI-enabled tools in clinical practice, such as decision-support systems, could carry risks of incorrect or biased recommendations that may directly impact patient outcomes [82,83], thus, raising ethical concerns about patient autonomy and safety as well as legal liability in cases of harm. Therefore, the differences of AI in each domain are important to understand in order to identify appropriate safeguards. Future research should conduct comparative analyses of AI's risks, implications, and ethical and legal dimensions in educational versus clinical settings, examining factors such as accuracy, equity, accountability, and trust. These studies can inform best practices and policies to optimize AI's potential while mitigating domain-specific risks.

Curriculum Framework Challenges

During the development of this curriculum framework, there were several challenges in effectively integrating an AI curriculum framework into a family medicine residency training program. During the expert panel discussions, many experts emphasized the issue regarding the lack of standardization with the definition of AI. Although a definition of AI was chosen for the purpose of the panel, a specific and committed definition of AI within medical education has not been established [84-86]. Panelists argued that an AI definition must be properly explained to avoid confusion or misrepresentation. In relation to family medicine, a recent primer for AI in primary care was published, which provided the definition, "The field of AI is broad and rapidly expanding. The field is centred on how computers might be able to perform humanlike 'intelligent tasks,' such as summarizing large amounts of information or making inferences about a situation" [87]. The discussions regarding this framework highlight the necessity of a standardized AI definition for better development of teaching and learning content. This is especially true when specializing in different

XSL•FO
RenderX

fields of medical education, including family medicine and primary care.

There is a need to introduce AI education within family medicine; however, the low urgency and priority to integrate this type of education at the moment were noted throughout the discussions. This can be due to the lack of AI-enabled tools for family physicians currently being developed, tested, and implemented in practice [88,89]. Furthermore, some residency programs lack the appropriate AI tools or are in lower-resource settings. As a result of the minimal exposure family physicians have with AI, their motivation to learn about the topic can also be reduced. This reduced priority of AI education competes with the CFPC's 105 topics of family medicine curricula [4]. This is exacerbated by the fact that Canada is in a unique position, in which the length of residency training is only 2 years. In addition, the rapid advancement of AI introduces an extra layer of complexity. As new AI-based tools emerge and existing ones advance, educators and family physicians must frequently reassess and update their knowledge and skills. For example, the recent introduction of generative AI and generative AI tools such as ChatGPT has gained widespread popularity in medical and academic settings [90]. Thus, it is difficult to maintain a robust framework due to the inevitable rapid changes of AI in health care. Therefore, the eagerness to integrate this type of education within the curriculum should be met with caution to manage the expectations of both educators and learners.

A key observation made throughout the panel discussion was about the AI content and how much should a family physician know about AI. During the discussions, many of the participants voiced support on the application and appraisal of AI-enabled tools. This is especially challenging when residency is only 24 months, and there are no required AI educational programs presented in the Canadian undergraduate medical education system. Therefore, within the learning objectives, in regard to how much a family physician should know about AI remains undetermined. Further research must be conducted to investigate the level of AI education a family physician should be aware of. Overall, the aforementioned challenges must be addressed in order for this curriculum framework to be effectively implemented.

Future Studies

Following the analysis, design, development, implementation, and evaluation model process, researchers may move forward to the implementation and evaluation of the AIFM-ed framework. During the implementation step, an educational program such as a course or workshop can be developed with the main concepts originating from the curriculum framework. The training for family medicine is already packed; thus, the implementation of this framework will depend on several factors including the current use of AI-enabled tools in family medicine training, previous training in AI (eg, the undergraduate foundation of AI), and the capacity of experienced teachers. However, once implemented, certain success indicators will need to be evaluated to understand its impact as well as any areas for improvement. Future studies could explore indicators such as the perceived impact of the framework, degree of implementation, as well as knowledge and skill apprehensions. These indicators can be evaluated through the framework-derived educational training program according to the Kirkpatrick model [61].

Strengths and Limitations of This Study

This study had several strengths, including the formation of a national, multidisciplinary panel of family medicine educators. This diverse panel facilitated enriching discussions with varied expertise and insights, allowing for a comprehensive understanding of practical implications and current perspectives on AI education in family medicine postgraduate training. Additionally, by involving both educators and residents, the AIFM-ed curriculum framework ensures the representation from key stakeholders involved in the teaching and learning process of AI education. This co-design approach enhanced the relevance and applicability of the AIFM-ed curriculum framework. Regarding the overall development of this framework, a multi-method systematic approach was used, which includes a comprehensive systematic scoping review and multiple expert panel discussions. This approach allowed us to identify and build on existing AI curriculum topics and resources while also creating new ones. Furthermore, this structured and reproducible methodology ensures a robust foundation that can be used by other educators and researchers to develop training programs (eg, courses) following the established framework.

Despite the strengths, this study also had few limitations. First, the study was developed for programs in Canada, which limits its applicability to other countries due to the different medical education structures globally and their current relationships with AI. However, this could be a starting guide for other researchers to adapt it to their own context. Additionally, expert panel diversity was limiting, where the resident panel came from a single institution, which may further limit the generalizability of the framework. Furthermore, as the participants for the panel discussion were not randomized and were purposively recruited, the results may be subject to selection bias.

Conclusions

We co-developed an AIFM-ed framework for family medicine residency training that outlines its curricular purpose, learning objectives, AI curricular topics, delivery methods, and evaluation strategies to be used by medical institutions. The AIFM-ed curriculum framework ultimately aims to enhance the education of future family physicians, equipping them to effectively integrate AI-enabled tools into their practice and patient care. It is hoped that this framework will provide further advocacy, productivity, and gradual change within the area of curriculum development and AI medical education. Overall, medical institutions are encouraged to begin equipping future physicians with the knowledge, skills, and confidence to effectively use AI-enabled tools, as these technologies will continue to grow within the field of health care and family medicine.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank the panelists for their support, time, and contributions to this research. The authors would also like to acknowledge Dr Pierre Pluye who provided key guidance to the development of this project prior to his passing in August 2023. SAR is Canada Research Chair (Tier II) in artificial intelligence and Advanced Digital Primary Health Care, received salary support from a Research Scholar Junior 1 Career Development Award from the Fonds de Recherche du Québec-Santé during part of this project, and her research program is supported by the Natural Sciences Research Council Discovery (grant 2020-05246).

Authors' Contributions

Conceptualization was led by SAR and RT, who established the study's goals, design, and research questions and obtained the funding for the project. The methodology was developed by SAR and RT. The data collection was done by RT and SAR. Data curation was managed by RT. A formal analysis was conducted by RT. The original draft was written by RT. SAR, FH-E, and MY provided critical revisions. Reviewing and editing were a collaborative effort with all authors. Supervision and overall project leadership were provided by SAR.

Conflicts of Interest

None declared.

References

1. Hennen BK. Academic family medicine in Canada. CMAJ 1993;148(9):1559-1563 [FREE Full text] [Medline: 8477381]
2. Oandasan I, Working Group on Postgraduate Curriculum Review. Advancing Canada's family medicine curriculum: Triple C. Can Fam Physician 2011;57(6):739-40, e237 [FREE Full text] [Medline: 21673223]
3. Shaw E, Oandasan I, Fowler N. CanMEDS-FM 2017: a competency framework for family physicians across the continuum. The College of Family Physicians of Canada. 2017. URL: https://www.cfpc.ca/CFPC/media/Resources/Medical-Education/CanMEDS-Family-Medicine-2017-ENG.pdf [accessed 2025-03-15]
4. Crichton T, Schultz K, Lawrence K, Donoff M, Laughlin T, Brailovsky C, et al. Assessment objectives for certification in family medicine. College of Family Physicians of Canada. 2020. URL: https://www.cfpc.ca/CFPC/media/Resources/Examinations/Assessment-Objectives-for-Certification-in-FM-full-document.pdf [accessed 2025-03-15]
5. Wojda T, Hoffman C, Jackson J, Conti T, Maier J. AI in healthcare: implications for family medicine and primary care. In: Artificial Intelligence in Medicine and Surgery—An Exploration of Current Trends, Potential Opportunities, and Evolving Threats—Volume 1. London: IntechOpen; 2023.
6. Buja LM. Medical education today: all that glitters is not gold. BMC Med Educ 2019;19(1):110 [FREE Full text] [doi: 10.1186/s12909-019-1535-9] [Medline: 30991988]
7. Jamieson S. State of the science: quality improvement of medical curricula—how should we approach it? Med Educ 2023;57(1):49-56 [FREE Full text] [doi: 10.1111/medu.14912] [Medline: 35950304]
8. Mcleod P, Steinert Y. Twelve tips for curriculum renewal. Med Teach 2015;37(3):232-238. [doi: 10.3109/0142159X.2014.932898] [Medline: 25010218]
9. Jones R, Higgs R, de Angelis C, Prideaux D. Changing face of medical curricula. Lancet 2001;357(9257):699-703. [doi: 10.1016/S0140-6736(00)04134-9] [Medline: 11247568]
10. Sultan N, Torti J, Haddara W, Inayat A, Inayat H, Lingard L. Leadership development in postgraduate medical education: a systematic review of the literature. Acad Med 2019;94(3):440-449. [doi: 10.1097/ACM.0000000000002503] [Medline: 30379659]
11. Hunter KA, Thomson B. A scoping review of social determinants of health curricula in post-graduate medical education. Can Med Ed J 2019;10(3):e61-e71. [doi: 10.36834/cmej.61709]
12. Hong DZ, Goh JL, Ong ZY, Ting JJQ, Wong MK, Wu J, et al. Postgraduate ethics training programs: a systematic scoping review. BMC Med Educ 2021;21(1):338 [FREE Full text] [doi: 10.1186/s12909-021-02644-5] [Medline: 34107935]
13. Pritchard J, Alavian S, Soogoor A, Bartels S, Hall A. Global health competencies in postgraduate medical education: a scoping review and mapping to the CanMEDS physician competency framework. Can Med Educ J 2023;14(1):70-79 [FREE Full text] [doi: 10.36834/cmej.75275] [Medline: 36998501]
14. Gupta A, Talavlikar R, Ng V, Chorny Y, Chawla A, Farrugia M, et al. Global health curriculum in family medicine: resident perspective. Can Fam Physician 2012;58(2):143-146 [FREE Full text] [Medline: 22439168]
15. Drain PK, Primack A, Hunt DD, Fawzi WW, Holmes KK, Gardner P. Global health in medical education: a call for more training and opportunities. Acad Med 2007;82(3):226-230. [doi: 10.1097/ACM.0b013e3180305cf9] [Medline: 17327707]
16. Runyan C, Savageau JA, Potts S, Weinreb L. Impact of a family medicine resident wellness curriculum: a feasibility study. Med Educ Online 2016;21:30648 [FREE Full text] [doi: 10.3402/meo.v21.30648] [Medline: 27282276]
17. Penwell-Waines L, Runyan C, Kolobova I, Grace A, Brennan J, Buck K, et al. Making sense of family medicine resident wellness curricula: a Delphi study of content experts. Fam Med 2019;51(8):670-676. [doi: 10.22454/FamMed.2019.899425] [Medline: 31269221]

XSL•FO
RenderX

18. Eckleberry-Hunt J, Van Dyke A, Lick D, Tucciarone J. Changing the conversation from burnout to wellness: physician well-being in residency training programs. J Grad Med Educ 2009;1(2):225-230 [FREE Full text] [doi: 10.4300/JGME-D-09-00026.1] [Medline: 21975983]
19. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
20. Abbasgholizadeh Rahimi S, Légaré F, Sharma G, Archambault P, Zomahoun HTV, Chandavong S, et al. Application of artificial intelligence in community-based primary health care: systematic scoping review and critical appraisal. J Med Internet Res 2021;23(9):e29839 [FREE Full text] [doi: 10.2196/29839] [Medline: 34477556]
21. Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240029200 [accessed 2025-03-12]
22. Upshaw TL, Craig-Neil A, Macklin J, Gray CS, Chan TCY, Gibson J, et al. Priorities for artificial intelligence applications in primary care: a Canadian deliberative dialogue with patients, providers, and health system leaders. J Am Board Fam Med 2023;36(2):210-220 [FREE Full text] [doi: 10.3122/jabfm.2022.220171R1] [Medline: 36948537]
23. Fowler N, Oandasan I, Wyman R. Preparing our future family physicians: an educational prescription for strengthening health care in changing times. College of Family Physicians of Canada. 2022. URL: https://www.cfpc.ca/CFPC/media/Resources/Education/AFM-OTP-Report.pdf [accessed 2025-03-15]
24. Fowler N, Lemire F, Oandasan I, Wyman R. The evolution of residency training in family medicine: a Canadian perspective. Fam Med 2021;53(7):595-598 [FREE Full text] [doi: 10.22454/FamMed.2021.718541] [Medline: 34000054]
25. Reznick R, Harris K, Horsley T, Hassani M. Artificial intelligence (AI) and emerging digital technologies. The Royal College of Physicians and Surgeons of Canada. 2020. URL: https://www.royalcollege.ca/content/dam/document/membership-and-advocacy/2020-task-force-report-on-ai-and-emerging-digital-technologies-e.pdf [accessed 2022-06-18]
26. Topol E. The Topol review: preparing the health care workforce to deliver the digital future. National Health Service. 2019. URL: https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-2019.pdf [accessed 2023-04-25]
27. AMA passes first policy recommendations on augmented intelligence. American Medical Association. 2018. URL: https://www.ama-assn.org/press-center/press-releases/ama-passes-first-policy-recommendations-augmented-intelligence [accessed 2023-04-25]
28. AI for family medicine. College of Family Physicians of Canada. URL: https://cfpclearn.ca/ai-for-family-medicine [accessed 2025-03-12]
29. Austin Z. Competency and its many meanings. Pharmacy (Basel) 2019 Apr 22;7(2):e12402 [FREE Full text] [doi: 10.3390/pharmacy7020037] [Medline: 31013596]
30. Liaw W, Kueper JK, Lin S, Bazemore A, Kakadiaris I. Competencies for the use of artificial intelligence in primary care. Ann Fam Med 2022;20(6):559-563 [FREE Full text] [doi: 10.1370/afm.2887] [Medline: 36443071]
31. Stabback P. What makes a quality curriculum? UNESCO. 2016. URL: https://unesdoc.unesco.org/ark:/48223/pf0000243975 [accessed 2022-07-10]
32. Training tools for curriculum development: developing and implementing curriculum frameworks. UNESCO International Bureau of Education. 2017. URL: https://unesdoc.unesco.org/ark:/48223/pf0000250052 [accessed 2022-07-10]
33. Chan KS, Zary N. Applications and challenges of implementing artificial intelligence in medical education: integrative review. JMIR Med Educ 2019;5(1):e13930 [FREE Full text] [doi: 10.2196/13930] [Medline: 31199295]
34. Tolentino R, Baradaran A, Gore G, Pluye P, Abbasgholizadeh-Rahimi S. Curriculum frameworks and educational programs in AI for medical students, residents, and practicing physicians: scoping review. JMIR Med Educ 2024;10:e54793 [FREE Full text] [doi: 10.2196/54793] [Medline: 39023999]
35. Tolentino R, Baradaran A, Gore G, Pluye P, Abbasgholizadeh-Rahimi S. Curriculum frameworks and educational programs in artificial intelligence for medical students, residents, and practicing physicians: a scoping review protocol. JBI Evid Synth 2023;21(7):1477-1484. [doi: 10.11124/JBIES-22-00374] [Medline: 37434376]
36. Redwood-Campbell L, Pakes B, Rouleau K, MacDonald CJ, Arya N, Purkey E, et al. Developing a curriculum framework for global health in family medicine: emerging principles, competencies, and educational approaches. BMC Med Educ 2011;11:46 [FREE Full text] [doi: 10.1186/1472-6920-11-46] [Medline: 21781319]
37. Masters K. Artificial intelligence developments in medical education: a conceptual and practical framework. MedEdPublish (2016) 2020;9:239 [FREE Full text] [doi: 10.15694/mep.2020.000239.1] [Medline: 38058891]
38. Valikodath NG, Cole E, Ting DSW, Campbell JP, Pasquale LR, Chiang MF, et al. Impact of artificial intelligence on medical education in ophthalmology. Transl Vis Sci Technol 2021;10(7):14 [FREE Full text] [doi: 10.1167/tvst.10.7.14] [Medline: 34125146]
39. Obadeji A. Health professions education in the 21st century: a contextual curriculum framework for analysis and development. J Contemp Med Edu 2019;9(1):34. [doi: 10.5455/jcme.20181212085450]
40. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for Reporting Qualitative Research: a synthesis of recommendations. Acad Med 2014;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
41. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. Am J Public Health 1984;74(9):979-983. [doi: 10.2105/ajph.74.9.979] [Medline: 6380323]

XSL•FO
RenderX

42. Black N, Murphy M, Lamping D, McKee M, Sanderson C, Askham J, et al. Consensus development methods: a review of best practice in creating clinical guidelines. J Health Serv Res Policy 1999;4(4):236-248. [doi: 10.1177/135581969900400410] [Medline: 10623041]
43. Vankova D, Videnova J. Delphi technique for curriculum development. 2019 Presented at: 12th annual International Conference of Education, Research and Innovation; November 11-13, 2019; Seville, Spain p. 6167-6171 URL: https://library.iated.org/view/VANKOVA2019DEL
44. Lewthwaite S, Nind M. Teaching research methods in the social sciences: expert perspectives on pedagogy and practice. Br J Educ Stud 2016;64(4):413-430. [doi: 10.1080/00071005.2016.1197882]
45. Mangold K, Bartell T, Doobay-Persaud A, Adler M, Sheehan K. Expert consensus on inclusion of the social determinants of health in undergraduate medical education curricula. Acad Med 2019;94(9):1355-1360. [doi: 10.1097/ACM.0000000000002593] [Medline: 31460933]
46. Hart A, Romney D, Sarin R, Mechanic O, Hertelendy A, Larson D, et al. Developing telemedicine curriculum competencies for graduate medical education: outcomes of a modified Delphi process. Acad Med 2022;97(4):577-585. [doi: 10.1097/ACM.0000000000004463] [Medline: 34670239]
47. Hsu T, Kessler E, Parker I, Dale W, Gajra A, Holmes H, et al. Identifying geriatric oncology competencies for medical oncology trainees: a modified Delphi consensus study. Oncologist 2020;25(7):591-597 [FREE Full text] [doi: 10.1634/theoncologist.2019-0950] [Medline: 32237179]
48. Evans C. The use of consensus methods and expert panels in pharmacoeconomic studies. Practical applications and methodological shortcomings. Pharmacoeconomics 1997;12(2 Pt 1):121-129. [doi: 10.2165/00019053-199712020-00003] [Medline: 10169665]
49. Duncan EAS, Nicol MM, Ager A. Factors that constitute a good cognitive behavioural treatment manual: a delphi study. Behav Cognit Psychother 1999;32(2):199-213. [doi: 10.1017/s135246580400116x]
50. Mead D, Moseley L. The use of the Delphi as a research approach. Nurse Res 2001;8(4):4-23. [doi: 10.7748/nr2001.07.8.4.4.c6162]
51. Framework for research ethics. UK Research and Innovation—Economic and Social Research Council. URL: https://www.ukri.org/councils/esrc/guidance-for-applicants/research-ethics-guidance/framework-for-research-ethics/ [accessed 2025-03-25]
52. Tri-council policy statement: ethical conduct for research involving humans. Government of Canada. 2018. URL: https://ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2022.html [accessed 2025-04-07]
53. Khodyakov D, Hempel S, Rubenstein L, Shekelle P, Foy R, Salem-Schatz S, et al. Conducting online expert panels: a feasibility and experimental replicability study. BMC Med Res Methodol 2011;11:174 [FREE Full text] [doi: 10.1186/1471-2288-11-174] [Medline: 22196011]
54. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
55. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
56. Hamid AYS, Chandra YA, Putri AF, Wakhid A, Falahaini A, Yulianingsih Y. Sustainable disaster risk reduction training model for nurses: a descriptive qualitative approach. Nurse Educ Pract 2023;69:103616. [doi: 10.1016/j.nepr.2023.103616] [Medline: 36996553]
57. Itchhaporia D. The evolution of the quintuple aim: health equity, health outcomes, and the economy. J Am Coll Cardiol 2021;78(22):2262-2264 [FREE Full text] [doi: 10.1016/j.jacc.2021.10.018] [Medline: 34823665]
58. Nundy S, Cooper LA, Mate KS. The quintuple aim for health care improvement: a new imperative to advance health equity. JAMA 2022;327(6):521-522. [doi: 10.1001/jama.2021.25181] [Medline: 35061006]
59. Chatterjee D, Corral J. How to write well-defined learning objectives. J Educ Perioper Med 2017;19(4):E610 [FREE Full text] [Medline: 29766034]
60. Badyal D, Singh T. Learning theories: the basics to learn in medical education. Int J Appl Basic Med Res 2017;7(Suppl 1):S1-S3 [FREE Full text] [doi: 10.4103/ijabmr.IJABMR_385_17] [Medline: 29344448]
61. Kirkpatrick D, Kirkpatrick J. Evaluating Training Programs: The Four Levels. Oakland, CA: Berrett-Koehler Publishers; 2006.
62. Stogiannos N, Gillan C, Precht H, Reis CSD, Kumar A, O'Regan T, et al. A multidisciplinary team and multiagency approach for AI implementation: a commentary for medical imaging and radiotherapy key stakeholders. J Med Imaging Radiat Sci 2024;55(4):101717. [doi: 10.1016/j.jmir.2024.101717] [Medline: 39067309]
63. Knopp MI, Warm EJ, Weber D, Kelleher M, Kinnear B, Schumacher DJ, et al. AI-enabled medical education: threads of change, promising futures, and risky realities across four potential future worlds. JMIR Med Educ 2023;9:e50373 [FREE Full text] [doi: 10.2196/50373] [Medline: 38145471]
64. Kueper JK, Emu M, Banbury M, Bjerre LM, Choudhury S, Green M, et al. Artificial intelligence for family medicine research in Canada: current state and future directions: report of the CFPC AI Working Group. Can Fam Physician 2024;70(3):161-168 [FREE Full text] [doi: 10.46747/cfp.7003161] [Medline: 38499374]
65. Inventory of Artificial Intelligence Resources in Family Medicine Education. Rahimi's Lab. URL: https://rahimislab.ca/inventory

XSL•FO
RenderX

66. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT—reshaping medical education and clinical management. Pak J Med Sci 2023;39(2):605-607 [FREE Full text] [doi: 10.12669/pjms.39.2.7653] [Medline: 36950398]
67. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023;11(6):887 [FREE Full text] [doi: 10.3390/healthcare11060887] [Medline: 36981544]
68. Ossa LA, Rost M, Lorenzini G, Shaw DM, Elger BS. A smarter perspective: learning with and from AI-cases. Artif Intell Med 2023;135:102458 [FREE Full text] [doi: 10.1016/j.artmed.2022.102458] [Medline: 36628794]
69. Sauder M, Tritsch T, Rajput V, Schwartz G, Shoja MM. Exploring generative artificial intelligence-assisted medical education: assessing case-based learning for medical students. Cureus 2024;16(1):e51961 [FREE Full text] [doi: 10.7759/cureus.51961] [Medline: 38333501]
70. Weidener L, Fischer M. Teaching AI ethics in medical education: a scoping review of current literature and practices. Perspect Med Educ 2023;12(1):399-410 [FREE Full text] [doi: 10.5334/pme.954] [Medline: 37868075]
71. Laupichler MC, Hadizadeh DR, Wintergerst MWM, von der Emde L, Paech D, Dick EA, et al. Effect of a flipped classroom course to foster medical students' AI literacy with a focus on medical imaging: a single group pre-and post-test study. BMC Med Educ 2022;22(1):803 [FREE Full text] [doi: 10.1186/s12909-022-03866-x] [Medline: 36397110]
72. Krive J, Isola M, Chang L, Patel T, Anderson M, Sreedhar R. Grounded in reality: artificial intelligence in medical education. JAMIA Open 2023;6(2):ooad037 [FREE Full text] [doi: 10.1093/jamiaopen/ooad037] [Medline: 37273962]
73. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023;183(6):589-596 [FREE Full text] [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
74. Jain A, Way D, Gupta V, Gao Y, de Oliveira Marinho G, Hartford J, et al. Development and assessment of an artificial intelligence-based tool for skin condition diagnosis by primary care physicians and nurse practitioners in teledermatology practices. JAMA Netw Open 2021;4(4):e217249 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.7249] [Medline: 33909055]
75. Lee C, Britto S, Diwan K. Evaluating the impact of artificial intelligence (AI) on clinical documentation efficiency and accuracy across clinical settings: a scoping review. Cureus 2024;16(11):e73994. [doi: 10.7759/cureus.73994] [Medline: 39703286]
76. Ahmed MI, Spooner B, Isherwood J, Lane M, Orrock E, Dennison A. A systematic review of the barriers to the implementation of artificial intelligence in healthcare. Cureus 2023;15(10):e46454 [FREE Full text] [doi: 10.7759/cureus.46454] [Medline: 37927664]
77. Hassan M, Kushniruk A, Borycki E. Barriers to and facilitators of artificial intelligence adoption in health care: scoping review. JMIR Hum Factors 2024;11:e48633 [FREE Full text] [doi: 10.2196/48633] [Medline: 39207831]
78. Nair M, Svedberg P, Larsson I, Nygren JM. A comprehensive overview of barriers and strategies for AI implementation in healthcare: mixed-method design. PLoS One 2024;19(8):e0305949 [FREE Full text] [doi: 10.1371/journal.pone.0305949] [Medline: 39121051]
79. Singh RP, Hom GL, Abramoff MD, Campbell JP, Chiang MF, AAO Task Force on Artificial Intelligence. Current challenges and barriers to real-world artificial intelligence adoption for the healthcare system, provider, and the patient. Transl Vis Sci Technol 2020;9(2):45 [FREE Full text] [doi: 10.1167/tvst.9.2.45] [Medline: 32879755]
80. Terry AL, Kueper JK, Beleno R, Brown JB, Cejic S, Dang J, et al. Is primary health care ready for artificial intelligence? What do primary health care stakeholders say? BMC Med Inform Decis Mak 2022;22(1):237 [FREE Full text] [doi: 10.1186/s12911-022-01984-6] [Medline: 36085203]
81. Perkins M, Pregowska A. The role of artificial intelligence in higher medical education and the ethical challenges of its implementation. AIH 2025;2(1):1-13 [FREE Full text] [doi: 10.36922/aih.3276]
82. Jiang L, Wu Z, Xu X, Zhan Y, Jin X, Wang L, et al. Opportunities and challenges of artificial intelligence in the medical field: current application, emerging problems, and problem-solving strategies. J Int Med Res 2021;49(3):3000605211000157 [FREE Full text] [doi: 10.1177/03000605211000157] [Medline: 33771068]
83. Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: systematic review. J Med Internet Res 2021;23(4):e25759 [FREE Full text] [doi: 10.2196/25759] [Medline: 33885365]
84. Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P. Introducing artificial intelligence training in medical education. JMIR Med Educ 2019;5(2):e16048 [FREE Full text] [doi: 10.2196/16048] [Medline: 31793895]
85. Grunhut J, Wyatt AT, Marques O. Educating future physicians in artificial intelligence (AI): an integrative review and proposed changes. J Med Educ Curric Dev 2021;8:23821205211036836 [FREE Full text] [doi: 10.1177/23821205211036836] [Medline: 34778562]
86. Lee J, Wu AS, Li D, Kulasegaram KM. Artificial intelligence in undergraduate medical education: a scoping review. Acad Med 2021;96(11S):S62-S70. [doi: 10.1097/ACM.0000000000004291] [Medline: 34348374]
87. Waljee AK, Higgins PDR. Machine learning in medicine: a primer for physicians. Am J Gastroenterol 2010;105(6):1224-1226. [doi: 10.1038/ajg.2010.173] [Medline: 20523307]

XSL•FO
RenderX

88. Blease C, Kaptchuk TJ, Bernstein MH, Mandl KD, Halamka JD, DesRoches CM. Artificial intelligence and the future of primary care: exploratory qualitative study of UK general practitioners' views. J Med Internet Res 2019;21(3):e12802 [FREE Full text] [doi: 10.2196/12802] [Medline: 30892270]
89. Irfan F. Artificial intelligence: help or hindrance for family physicians? Pak J Med Sci 2021;37(1):288-291 [FREE Full text] [doi: 10.12669/pjms.37.1.3351] [Medline: 33437293]
90. Mousavi M, Shafiee S, Harley JM, Cheung JCK, Abbasgholizadeh Rahimi S. Performance of generative pre-trained transformers (GPTs) in certification examination of the College of Family Physicians of Canada. Fam Med Com Health 2024 May 28;12(Suppl 1):e002626. [doi: 10.1136/fmch-2023-002626] [Medline: 38806403]

Abbreviations

AI: artificial intelligence
AIFM-ed: Artificial Intelligence Training in Postgraduate Family Medicine Education
CanMEDS: Canadian Medical Education Directives for Specialists
CFPC: College of Family Physicians of Canada
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
SRQR: Standards for Reporting Qualitative Research

Edited by B Lesselroth; submitted 24.09.24; peer-reviewed by K Thompson, G Evangelinos; comments to author 12.12.24; revised version received 04.02.25; accepted 25.02.25; published 25.04.25.

Please cite as:
Tolentino R, Hersson-Edery F, Yaffe M, Abbasgholizadeh-Rahimi S
AIFM-ed Curriculum Framework for Postgraduate Family Medicine Education on Artificial Intelligence: Mixed Methods Study
JMIR Med Educ 2025;11:e66828
URL: https://mededu.jmir.org/2025/1/e66828
doi:10.2196/66828
PMID:

©Raymond Tolentino, Fanny Hersson-Edery, Mark Yaffe, Samira Abbasgholizadeh-Rahimi. Originally published in JMIR Medical Education (https://mededu.jmir.org), 25.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Leveraging Datathons to Teach AI in Undergraduate Medical Education: Case Study

Michael Steven Yao[1,2,3], BS; Lawrence Huang[3,4*], BS; Emily Leventhal[3,5*], BA; Clara Sun[3,6], BS; Steve J Stephen[3,7,8], MBA; Lathan Liou[3,5], MPhil

[1]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
[2]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, United States
[3]MDplus, New York, NY, United States
[4]Warren Alpert Medical School, Brown University, Providence, RI, United States
[5]Icahn School of Medicine at Mount Sinai, New York, NY, United States
[6]School of Medicine, Case Western Reserve University, Cleveland, OH, United States
[7]School of Medicine and Dentistry, University of Rochester, Rochester, NY, United States
[8]Simon Business School, University of Rochester, Rochester, NY, United States
[*]these authors contributed equally

Corresponding Author:
Lathan Liou, MPhil
MDplus, New York, NY, United States

Abstract

Background: As artificial intelligence and machine learning become increasingly influential in clinical practice, it is critical for future physicians to understand how such novel technologies will impact the delivery of patient care.

Objective: We describe 2 trainee-led, multi-institutional datathons as an effective means of teaching key data science and machine learning skills to medical trainees. We offer key insights on the practical implementation of such datathons and analyze experiences gained and lessons learned for future datathon initiatives.

Methods: We detail 2 recent datathons organized by MDplus, a national trainee-led nonprofit organization. To assess the efficacy of the datathon as an educational experience, an opt-in postdatathon survey was sent to all registered participants. Survey responses were deidentified and anonymized before downstream analysis to assess the quality of datathon experiences and areas for future work.

Results: Our digital datathons between 2023 and 2024 were attended by approximately 200 medical trainees across the United States. A diverse array of medical specialty interests was represented among participants, with 43% (21/49) of survey participants expressing an interest in internal medicine, 35% (17/49) in surgery, and 22% (11/49) in radiology. Participant skills in leveraging Python for analyzing medical datasets improved after the datathon, and survey respondents enjoyed participating in the datathon.

Conclusions: The datathon proved to be an effective and cost-effective means of providing medical trainees the opportunity to collaborate on data-driven projects in health care. Participants agreed that datathons improved their ability to generate clinically meaningful insights from data. Our results suggest that datathons can serve as valuable and effective educational experiences for medical trainees to become better skilled in leveraging data science and artificial intelligence for patient care.

(JMIR Med Educ 2025;11:e63602) doi:10.2196/63602

KEYWORDS

data science education; datathon; machine learning; artificial intelligence; undergraduate medical education

Introduction

The exploration of machine learning (ML), artificial intelligence (AI), and other data science-driven technologies is becoming increasingly popular within clinical medicine [1-5]. Given the rapidly growing presence of ML in health care innovation, it is important for both current and future physicians to understand the fundamentals of ML technology and how they may help inform clinical decision-making.

However, data science and AI education in current medical school curricula are lacking. Despite recent efforts to integrate AI learning objectives into medical education [6-10], few US medical schools have formally integrated AI-based topics into their curricula. Pupic et al [11] and Civaner et al [12] report studies of small self-selected groups of medical students and

XSL•FO
RenderX

residents participating in both student- and faculty-led electives covering the fundamental theory behind AI applications for medicine. However, opportunities facilitating real-world experience remain limited [13,14].

One potential method for hands-on AI education popular across many fields of science and engineering is the "datathon," which is a short competition where teams of students work together to create new solutions to domain-specific challenges through leveraging real-world data and algorithms. Following Daneshvar et al [15], we also make the important distinction between datathons and hackathons. Traditionally, hackathons are product-orientated initiatives where team projects are primarily focused on programming novel products and applications. By contrast, the primary learning objectives for our datathons were to (1) teach student participants how to analyze complex datasets to support clinical insights, and (2) leverage ML models to derive these clinical insights from data. Oyetade et al [16] offer a scoping review of datathons and found that such events help students learn both technical and soft skills and argue that datathon-based pedagogies be incorporated in classroom environments. Silver et al [17] describe a hackathon event for current attendings in clinical practice and found that study participants were better equipped to accelerate specialty-focused innovation after the hackathon. However, similar events specifically designed for medical students and other undergraduate trainees are not well described in the literature.

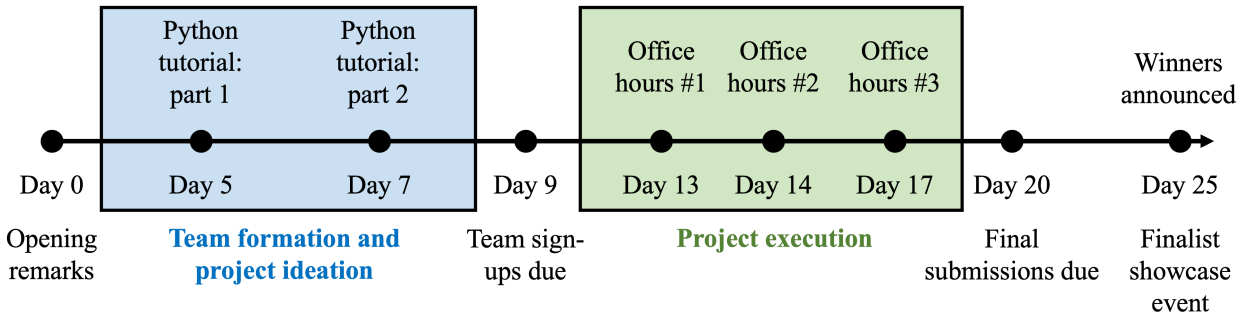
In this work, we hypothesize that datathons can be an effective training initiative to teach skills in AI to medical trainee participants. To evaluate this hypothesis, we describe 2 datathons hosted by MDplus, a 501(c)3 national student-run nonprofit whose mission is to support and empower future physician-innovators. We describe the structure of the events, present data on educational outcomes, and offer resources and recommendations for putting together similar events in the future. Our results suggest that datathons and similar events may be an effective means for AI education for medical students.

Methods

Overview

In this section, we detail the logistics of organizing and executing 2 trainee-led datathon events. A number of features distinguish our events from prior work. First, the datathons are trainee-led; all members of the organizing committee were undergraduate medical trainees at the time of the event. Second, the datathons were held digitally over the course of approximately 3 weeks (Figure 1). Finally, the target participant audience of our datathons included current undergraduate medical trainees at institutions granting doctoral degrees. These features of our datathons substantially differentiate them from prior work [15-17], and also affected our design and organization of the events that we detail below.

Figure 1. Overview of the datathon event. The MDplus datathon ran for approximately 4 weeks and was loosely divided into two parts: (1) Team formation and project ideation and (2) project execution.



Timeline and Participant Recruitment

We, the datathon organizing team, detail 2 datathon events organized by MDplus between 2023 and 2024, herein referred to as "the datathons." Each datathon ran for approximately 3 weeks (Figure 1), and was organized by the medical trainee-led executive team of MDplus, consisting of a core datathon planning team of 8 medical trainees. To accommodate the participation of medical trainees from across the United States, the entirety of both datathons was held digitally. The MDplus' Slack community, monthly newsletter, and social media pages (LinkedIn, Instagram, and Twitter) were used to advertise the datathon. Over the span of 3 months prior to the start of the datathon, 2 organizing team members were tasked with recruiting sponsors, mentors, knowledge experts, and judges through the MDplus and personal networks, while 3 organizing team members—all with prior experience working as software engineers prior to medical school—crafted and iterated the educational material and dataset curation for the event. One team member helped publicize the event on social media. Registration for the event was limited to current trainees (ie, medical students, residents, and graduate students) in the United States. To provide a fair learning environment for trainees, our organizing team opted to exclude attending physicians, industry professionals, and individuals with extensive technical backgrounds in software engineering from participation. Participants were asked to form their own teams of 3-5 individuals.

Datathon Theme and Dataset

Each of the datathons focused on a specific theme to help participants contextualize their projects within a specific application relevant to health care. The theme of the 2024 datathon was responsible generative AI and that of the 2023 datathon was value-based care (VBC). Generative AI is an area of ML that uses technologies such as large language models (LLMs) to create new content by learning patterns from existing human-generated examples [18-20]. While such technologies

XSL•FO
RenderX

have the potential to improve health care delivery, recent work has highlighted a growing need to better evaluate how clinicians can use these tools responsibly before real-world integration is possible [21-23]. Separately, VBC refers to a health care delivery model in which providers are held accountable for improving patient outcomes. In a VBC system, providers are often rewarded with incentivized payments based on quality of care, provider performance, and the patient experience [24].

To enable participants to explore projects related to each of these themes, a medical dataset was made available for participants to use in each of our datathons. All datasets were made available via Hugging Face (Hugging Face, Inc), a public repository to facilitate the sharing of ML data and models. In our 2023 VBC datathon, participants were required to use the Medical Information Mart for Intensive Care (MIMIC)-IV dataset [25], a single-site dataset of patient records and admission details. Briefly, the MIMIC-IV dataset contains anonymized patient data aggregated from over 500,000 patients at the Beth Israel Deaconess Medical Center between 2008 and 2019. Variables from this rich dataset include electrocardiograms, medical imaging studies, health records, and patient laboratory values and outcomes, among others. We chose to use this dataset specifically for the datathon because of the following factors:

Public Availability

In similar prior events organized by the authors, we found that procuring a real-world dataset of health care data can often be prohibitively expensive or constraining, especially for trainee-led initiatives with limited budgets. To circumvent this problem, we used the MIMIC-IV dataset, which is made publicly available by Johnson et al [25].

Real Patient Data and Outcomes

The primary learning objective of our datathons is to teach participants how to derive data-driven insights to affect and ultimately improve patient care. We therefore sought to provide real patient data for participants to explore and use for their projects in alignment with this goal.

Prevalence of Prior Work

The vast majority of our participants have minimal (if any) prior experience with programming and data analysis techniques. For this reason, the abundance of prior literature and publicly available coding resources for interacting with the MIMIC-IV dataset helped lower the barrier to participating in the datathon.

Multiple Modalities of Data

Many participants have individual academic and personal interests in medicine, and we sought to encourage participants to craft and work on projects that were interesting to them. The abundance of textual, image, biomedical signal, and laboratory data available in the MIMIC-IV dataset was important to make this possible.

All participants in the datathon were required to sign a data use agreement and complete responsible data handling training in order to gain access to the MIMIC-IV dataset. Participating teams were tasked with thinking critically about quantitative methods, conducting appropriate analyses (eg visualization, statistics, and other computational tools), and contextualizing clinical insights into actionable proposals that solve a problem related to VBC for relevant stakeholders.

While organizing for the 2024 generative AI datathon, we found that one limitation of the MIMIC-IV dataset was its size and complexity, making it unwieldy for some participants to work with for their projects. To overcome this challenge while simultaneously retaining the desirable features listed above, our 2024 datathon introduced the concept of datathon “tracks”: teams were able to choose to participate in 1 of 3 tracks within the broader theme of responsible generative AI. Each track was associated with its own dataset: (1) Clinical Documentation track participants used the MTS-Dialog dataset of patient-physician conversation transcripts from Abacha et al [26]; (2) Medical Education track participants used the MedQA dataset of practice medical board examination questions from Jin et al [27]; and (3) Mental Health track participants used the SuicideWatch and Mental Health Collection dataset of tagged social media posts from Ji et al [28]. Participants were allowed to participate in at most 1 of the 3 tracks.

Resources and Support

An official datathon page [29] was created for participants as a central hub with instructions, registration, and materials for the event. Links to the datathon’s Github Repository were provided with written tutorials and example code, including (1) downloading and overview of the datasets; (2) introduction to Python (Python Software Foundation; offered in both the 2023 and 2024 datathons; see Multimedia Appendix 1); and (3) an introduction to R (R Foundation; offered only in the 2023 datathon). Optional workshops and private Zoom (Zoom Communications, Inc) events with experienced data scientists were offered to participating trainees, including Python and R bootcamps, oral presentation workshops, and a prerecorded Zoom talk with physician experts. The scope of the projects was largely left up to the discretion of individual team members; participant teams were encouraged to leverage the optional workshop sessions and public discussion channels on Slack if they would benefit from discussing potential project ideas with others, although no explicit guidance on project ideation or constraints was given other than all teams had to (1) use the official datathon dataset and (2) work on a project under the broad datathon theme (ie, VBC in 2023 and responsible generative AI in 2024) and track. No tutorials or structured datathon programming were provided for teaching participants how to use GitHub, GitLab, Microsoft Excel, or other computing tools. Communication and announcements throughout the datathon were conducted through Slack.

Submission Requirements and Judging Criteria

In the 2023 VBC datathon, teams were asked to submit a written technical report of their work without restrictions on the word count and were asked to record a 5-minute-long oral presentation highlighting key contributions and findings. Participants were free to use any programming language or software to perform their analysis. In the 2024 generative AI datathon, teams were asked to submit a 1-page extended abstract with at most 1 figure and unlimited references and a written technical report without word count restrictions. Judging criteria in both datathons

XSL•FO
RenderX

included statistical rigor, relevance to the datathon theme (VBC), creativity of visualization and analysis, and team diversity (Multimedia Appendix 2).

Final Showcase Event

In the 2023 VBC datathon, an internal set of 4 blinded judges composed of members of the MDplus datathon organizing committee evaluated the initial anonymized submissions and selected 7 finalist teams to present at the finalist datathon showcase event. Each team played their recorded 5-minute oral presentations and were allotted 2 minutes immediately after for responding to judge questions. A panel of 5 judges—recruited for their diverse range of expertise in the VBC space—evaluated the finalists' submissions. In total, 3 of the judges are health care executives, 4 are practicing clinicians, and 1 is a product manager.

In the 2024 generative AI datathon, an internal set of 3 blinded judges composed of members of the organizing committee evaluated the 14 initial anonymized team submissions and selected 8 finalist teams to present at the final datathon showcase. Finalist teams were invited to a 2-hour finalist showcase event where they were each allotted 8 minutes for a live oral presentation followed by 2 minutes of question answering with the judges. We recruited a panel of 4 judges to evaluate the finalist submissions: 1 judge is a software engineer at a health care company, 1 judge is a postdoctoral fellow in a health care AI lab, and 2 judges are practicing physicians in the United States. In general, we found the live oral presentations to be better received by the judges and audience members than playing prerecorded presentations.

Postdatathon Survey

Upon the conclusion of each datathon, an anonymous 16-question open survey (Multimedia Appendix 3) was electronically sent to all registered participants that submitted a final project via both Slack and email; this survey study was exempted by the University of Pennsylvania Institutional Review Board (protocol #856530). The survey was created in close collaboration with an attending physician at a US academic medical institution with expertise in medical education and assessing educational outcomes and was piloted within the datathon organizing team prior to the public release of the survey. Participants were requested to complete the survey within the 2 weeks immediately following the conclusion of the respective datathon, and the survey remained open for 3 weeks. Participant emails were collected to ensure that no individual filled out the survey multiple times but were removed prior to analysis. The optional, opt-in survey asked respondents questions pertaining to team demographics, medical education status, medical specialty interest, familiarity with technical and computational tools, and subjective datathon quality. The questions were divided between 4 survey pages, each taking approximately 1 minute to complete; no partial survey responses were submitted. Participants were asked to rate their familiarity with quantitative tools before and after the datathon on a 4-point scale (1=no familiarity, 4=a lot of familiarity) and were also informed that the study results would be anonymized and deidentified prior to analysis. To assess the efficacy of the aforementioned technical Python and R tutorials for datathon participants, we compared them against participant subjective familiarity with quantitative tools—namely, GitHub and Microsoft Excel—that were not taught explicitly as a part of the datathon. Data were analyzed using the Fisher exact test in Python 3. To better characterize participant experiences during the datathon, survey respondents also rated their agreement with a set of 5 standardized statements regarding (1) overall enjoyment of the datathon, (2) VBC topic understanding, (3) ability to identify problems in health care, (4) ability to generate insights from data, and (5) likelihood of future datathon participation. Participant sentiment was quantified using a 5-point Likert scale (1=strongly disagree, 5=strongly agree) [30].

Ethical Considerations

This study was exempted by the University of Pennsylvania Institutional Review Board (protocol #856530). All opt-in participants provided informed consent prior to data collection and were not compensated for participating in our optional, opt-in survey as a part of our study. Confidentiality and privacy were maintained during data acquisition and analysis, and participants had the right to withdraw their data from the study at any time without any consequences.

Results

Datathon Logistics

In the 2023 datathon, 28 teams consisting of a total of 109 participants registered for the datathon, of which 13 of the initial registered teams submitted a final project, while in the 2024 datathon, 25 teams consisting of 110 participants registered for the datathon, of which 14 of the initial registered teams submitted a final project. Among the submitted projects, 7 and 8 were chosen as finalists to present at the synchronous digital showcase in 2023 and 2024, respectively. In the 2023 VBC datathon, the 7 projects addressed a variety of topics related to VBC, including chronic kidney disease underdiagnosis, the efficacy of social work referrals, and readmission rates for alcohol-related conditions, among others. Similarly, the 2024 responsible generative AI datathon featured 2 clinical documentation track teams, 3 medical education track teams, and 3 mental health track teams. We include brief descriptions of each of the finalist projects in Table 1. The final showcase was followed by the announcement of the 3 winning projects; we announced the winning teams at the end of the showcase in the 2023 datathon and 48 hours after the end of the showcase in the 2024 datathon.

XSL•FO
RenderX

Table . Sample datathon project descriptions. Descriptions of finalist datathon projects for the 2023 and 2024 MDplus datathons are shown to illustrate the diversity of project submissions from participating teams.

Theme or track	Project description
Value-based care	• Minimizing chronic kidney disease (CKD) underdiagnosis using machine learning • Significant association of social work referral and 30-day unplanned hospital readmission for patients with alcohol-related disorders using MIMIC[a]-IV data • Can we curb frequent emergency department (ED) visits due to alcohol-related conditions? • Automatic knowledge graph extraction from medical discharge notes for clinical decision support • Contrast overuse in patients with renal disease: a targeted analysis • Analyzing acuity as a tool for value-based care • Machine learning-driven forecasting and characterization of the intensive care unit (ICU)-admitted heart failure patient population in the MIMIC-IV (version 04) database
Responsible generative AI: clinical documentation	• Cost-benefit analysis of non-artificial intelligence (AI) and AI models implemented for predicting chief complaints • Bridging speech documentation and clinical support through LLM[b] automation • Automating trust in AI-generated clinical notes: developing a look-up tool for real-time verification
Responsible generative AI: medical education	• Using a LLM for USMLE[c] preparation via generative AI • Use of LLMs in assessing how age and gender affect model accuracy in clinical reasoning
Responsible generative AI: mental health	• Reassessing specialist models: risks in fine-tuning LLMs for mental health tasks • Robust text classification and grounded LLM integration for personalized mental health support • Characterizing suicidal ideation subtypes in social media posts via unsupervised contrastive feature identification

[a]MIMIC: Medical Information Mart for Intensive Care.

[b]LLM: large language model.

[c]USMLE: United States Medical Licensing Examination

The organization-accrued cost of organizing and running the datathons was US $28 per participant, averaged over the number of participants who individually registered for the datathon regardless of whether they ultimately submitted a final project. The majority of expenses supported prize money, computing resources for participants, technical skill-based workshops, and other resources that were provided during the datathon. In our experience, most of the costs accrued were for (1) the prize money of the datathon winners and (2) honorariums for the guest judges in the finalist showcase events. We primarily relied on sponsorships from industry partners to provide computing resources for participants, and MDplus community members readily volunteered to help lead technical skill-based workshops and offer pro-bono mentorship to participating teams.

Survey Results

Out of the 219 registered participants (summed over both datathons), 61 (28%) completed the postdatathon survey (Table 2). A majority who completed the survey identified as male (71%, 43/61) and were under the age of 25 years (61%, 37/61). Survey respondents self-reported as Asian (69%, 42/61), White (20%, 12/61), Middle Eastern or North African (3.3%, 2/61), Hispanic or Latinx (3.3%, 2/61), or Black or African American (1.6%, 1/61); and 3.3% (2/61) preferred not to say. In total, 49/61 (80%) of survey respondents were medical students (Table 3); there was a wide range of medical specialty interests amongst the medical trainee survey respondents, with internal medicine (21/49), surgery (17/49), and radiology (11/49) being the most popular specialties.

XSL•FO
RenderX

Table . Demographic information of participants who completed the postdatathon survey (N=61).

Characteristics	Value, n (%)
Age, years	
<25	37 (61)
25–30	22 (36)
30–35	1 (1.6)
≥35	1 (1.6)
Self-reported race and ethnicity	
Asian	42 (69)
Hispanic or Latinx	2 (3.3)
Middle Eastern or North African	2 (3.3)
White	12 (20)
Black or African American	1 (1.6)
Prefer not to say	2 (3.3)
Gender	
Male	43 (71)
Female	18 (29)
Sexual orientation	
Heterosexual or straight	57 (93)
Bisexual, gay, lesbian, or other	4 (6.6)
Disability status	
Does not identify as a person with a disability	54 (89)
Does identify as a person with a disability	5 (8.2)
Prefer not to answer	2 (3.3)
Current education status	
Medical student or resident physician	49 (80)
Other	12 (20)

XSL•FO
RenderX

Table . Datathon participant analysis. Current medical education status and medical specialty interest information for participants who completed the postdatathon survey (N=49) filtered by medical student and resident physician status. Note that respondents were allowed to select multiple medical specialties.

Characteristics	Value
Current medical education status, n (%)	
First-year medical student	14 (29)
Second-year medical student	19 (39)
Third-year medical student	6 (12)
Year-out medical student	4 (8.2)
Fourth-year medical student	5 (10)
Resident physician	1 (2.0)
Medical specialty interests, n (%)	
Anesthesia or critical care	9 (18)
Cardiology	1 (2.0)
Dermatology	5 (10)
Emergency medicine (EM)	4 (8.2)
Family medicine (FM)	1 (2.0)
Internal medicine	21 (43)
Mental health counseling and therapy	2 (4.1)
Neurology	9 (18)
Obstetrics and gynecology (OB/GYN)	3 (6.1)
Ophthalmology	5 (10)
Pediatrics	5 (10)
Physical medicine and rehabilitation (PM&R)	1 (2.0)
Plastic surgery	1 (2.0)
Psychiatry	8 (16)
Radiology	11 (22)
Surgery (general or unspecified)	17 (35)
Orthopedic surgery	2 (4.1)
Not currently exploring a medical specialty	1 (2.0)

Familiarity with quantitative tools, Python, R, Github/Gitlab, and Microsoft Excel before and after participating in the datathon was assessed (Figure 2). As a reminder, a core component of the programming of both our VBC and generative AI datathons was the educational workshops and tutorials on data analysis and ML skills using Python. Workshops on the programming language R were only offered in the 2023 VBC datathon. As our negative controls, we also asked participants to rate their skills with Github/Gitlab and Microsoft Excel; neither of these software were primary educational components of the datathons. As expected, participant familiarity with GitHub/Gitlab and Microsoft Excel did not significantly change before and after the datathon (GitHub/Gitlab: P=.92; Microsoft Excel: P=1.00; pairwise Fisher exact test). In contrast, subjective participant familiarity with Python significantly improved through participation in the datathons (P=.04; pairwise Fisher exact test); familiarity with R showed some evidence of improvement (P=.83; pairwise Fisher exact test), although it did not reach the traditional threshold for statistical significance likely due to the limited sample size of the study. Our reports support that targeted educational tutorials during the datathon event can empower participants with improved technical skills relevant to data science applications in medicine.

XSL•FO
RenderX

Figure 2. Bar plot visualizing participant self-assessment of technical skills before and after participating in the datathon for all 61 survey responses. Python was the only skill out of the 4 above that was an educational component in both the 2023 VBC and 2024 generative AI datathons. Participant scores correspond to the following: (1) no familiarity; (2) a little familiarity; (3) some familiarity; (4) a lot of familiarity. * Indicates a statistically significant difference in the distribution of scores before and after participating in the datathon (Python: *P*=.041; pairwise Fisher exact test). n.s. indicates no statistically significant difference in the distribution of scores. (R: *P*=.83; GitHub/Gitlab: *P*=.92; Microsoft Excel: *P*=1.00; pairwise Fisher exact test).

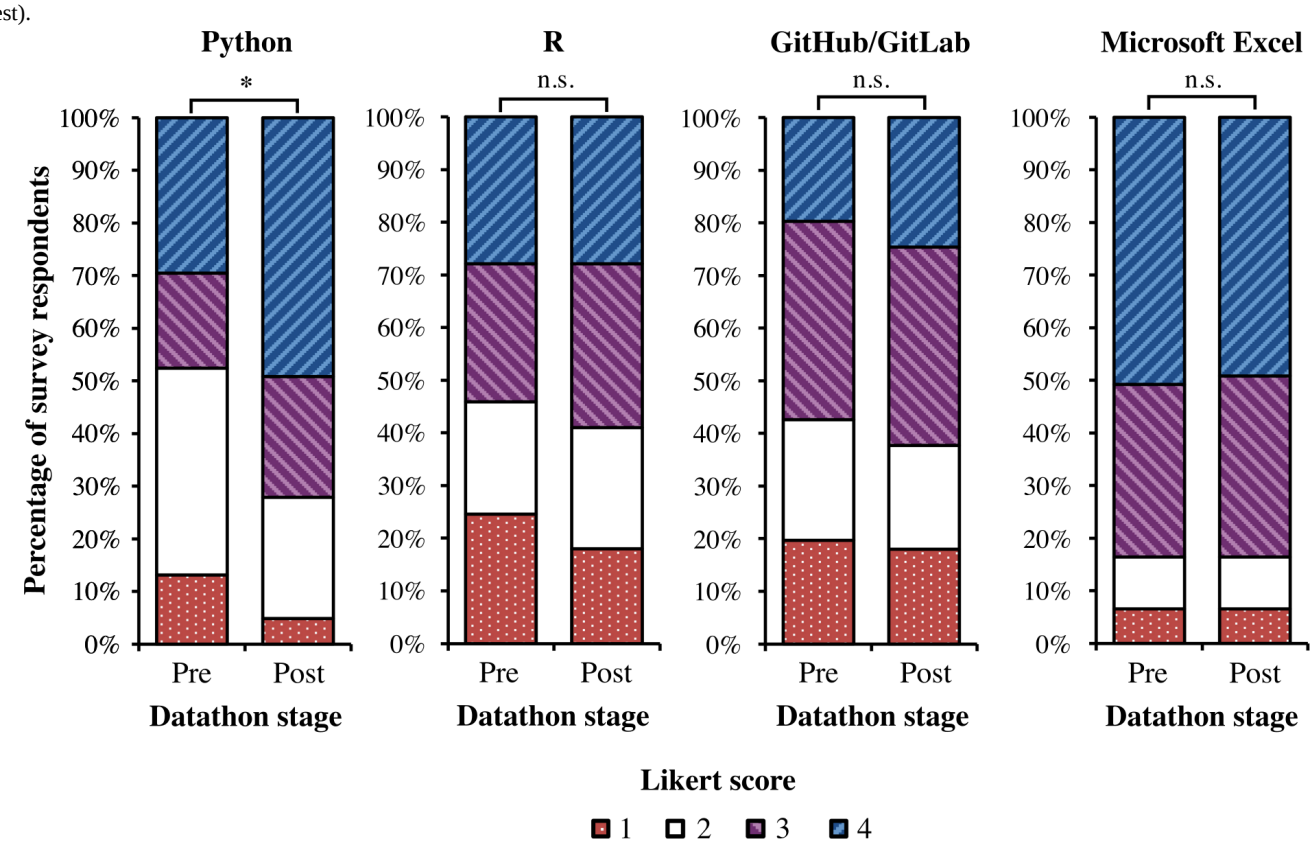
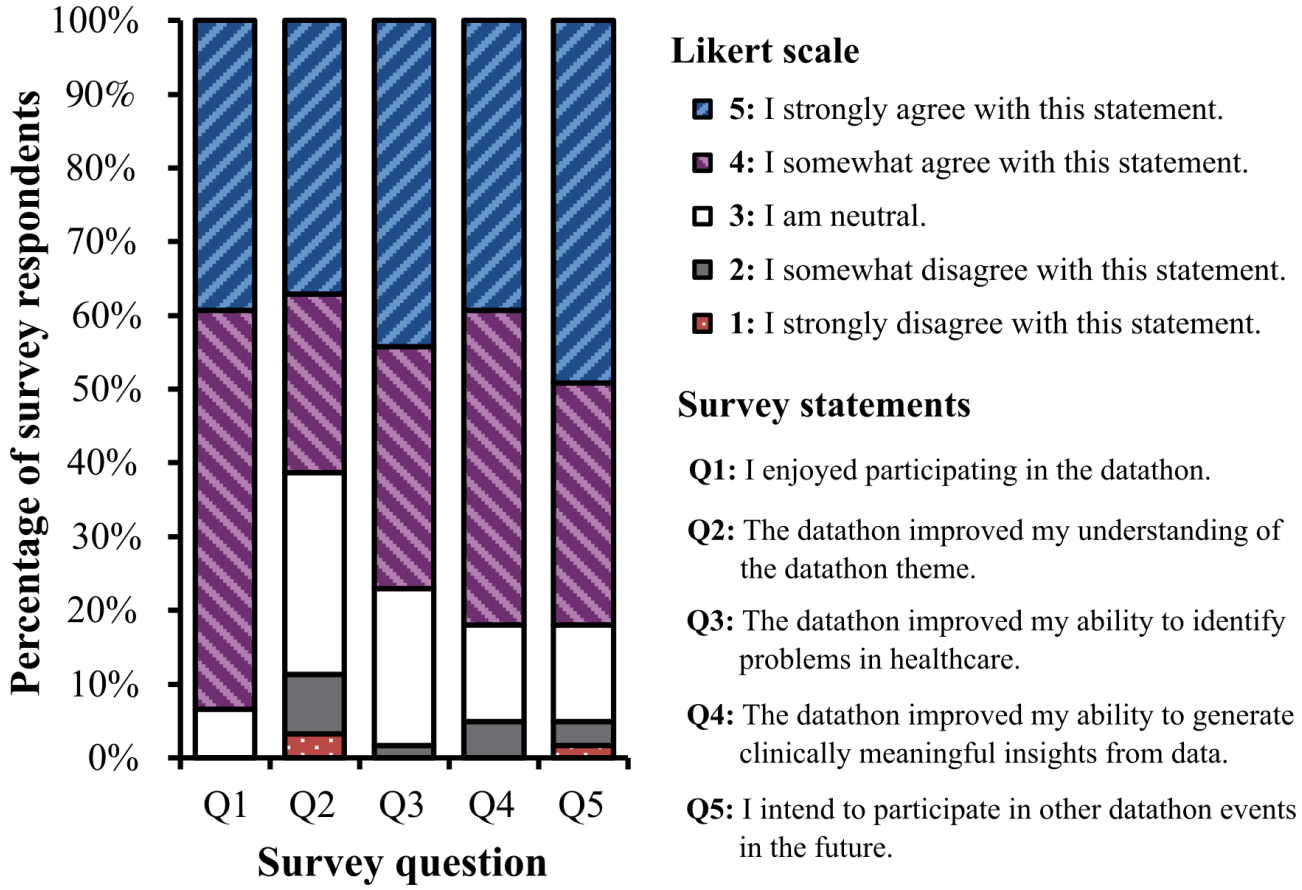


Figure 3 examines the participant experience quantified by participant agreement with a set of standardized statements. Overall, 57/61 (93%) survey respondents enjoyed participating in the datathon, and 38/61 (62%) respondents affirmed that the datathon improved their understanding of the VBC or responsible generative AI theme (ie, Likert score of 4 or 5). We also found that 47/61 (77%) respondents stated that their ability to identify problems in health care improved, and 50/61 (82%) respondents agreed that they were better equipped to generate meaningful insights from data. Of the 61 participants, 50 participants (82%) also expressed interest in participating in similar datathon events in the future. For each of these statements, an "agreeable sentiment" was determined by indicating a Likert scale value of either 4 ("I somewhat agree with the statement") or 5 ("I strongly agree with the statement") on a 5-point scale in the participant survey response.

XSL•FO
RenderX

Figure 3. Bar plot visualizing survey results assessing for subjective datathon quality. Participant scores correspond to the following: (1) Strongly disagree; (2) Disagree; (3) Neither agree nor disagree; (4) Agree; and (5) Strongly agree.



Qualitative Survey Results

The survey also included an open-ended response option for participants to provide any additional comments. There was a mix of short, positive comments and comments that offered suggestions for future events. Based on our qualitative analysis, key areas for improvement to consider for future datathon iterations include (1) ensuring a balanced distribution of technical skills between participating teams; (2) expediting the team creation process; and (3) offering additional technical workshops and tutorials to participants. Representative example unedited participant comments are shown below:

> *I think in the future, it'd be more effective to make sure each team at least has a "senior" tech lead (someone with 3-5+ years of tech experience) and a "junior" tech lead (1-2 years) to ensure there is great education for all parties involved, as well as greater quality of work. This is of course for folks seeking out teams and not those who already have a team formed that they are comfortable with.*
>
> *…I feel like the team creation process could've been a little faster and I was only able to join a team around halfway into the datathon which didn't give us enough time to work on our idea. But overall, I really appreciate the effort and time put in by everyone involved and I definitely hope to be involved in this again!*

Discussion

Principal Findings

In this work, we describe an instance of a trainee-led datathon to teach medical trainees how to effectively leverage modern computational tools to solve real-world problems in medicine. We show preliminary evidence that trainees become more familiar with foundational skills such as reading and writing computer programs in Python and R, are satisfied with their participation, and are eager to participate in similar initiatives in the future. To our knowledge, our national, trainee-led datathons were the first to bring together teams of medical students, residents, and graduate students to propose data-driven solutions within VBC. Our study ultimately supports that datathons can be effective platforms to teach medical trainees how to leverage AI to advance clinical medicine.

Logistical Insights and Best Practice Recommendations

In this section, we offer additional discussion on subjective design choices and lessons learned from the MDplus datathon organizing team. We hope that our experiences and takeaways can serve as a foundation for which future datathon educational initiatives can build upon.

Perhaps one of the most notable logistical details that distinguish our datathons from related hackathons that are traditionally organized by computer science students outside of medicine is that our datathons each spanned the course of multiple weeks asynchronously, whereas hackathons are often held over the

XSL•FO
RenderX

course of a few days in a single physical location. While we recognize that there are likely untapped benefits with this alternative strategy, we chose to run an extended digital datathon due to two primary reasons: (1) to support participation from MDplus members spanning multiple countries and timezones; and (2) to minimize potential time conflicts with concurrent medical school curricula for participants. In our work, these 2 constraints together necessarily precluded an in-person datathon; in situations where either one or both constraints are not limiting, future work may warrant exploring similar datathon initiatives spanning a few days hosted in a single physical location.

Separately, we also emphasize the importance of carefully choosing the datasets used in the datathon. In our 2023 VBC datathon using the MIMIC-IV dataset, we retrospectively observed that some participants initially struggled with the technical implementation details of working with the MIMIC-IV dataset due to the sheer volume of data available and the preprocessing steps before any ML modeling could be done. This consideration was especially important as the majority of participants registered in the datathon with little or no prior experience with computer programming (Figure 2). At the same time, participants also voiced enthusiasm for the diversity of data available in the MIMIC-IV dataset—making multiple modalities of data available, such as medical imaging, textual clinical documentation, biometric signals, and tabular data, allowed for participating teams to design and execute projects tailored to their specific interests. In our 2024 generative AI datathon, we found that the introduction of datathon "tracks" enabled us to offer 3 diverse dataset options while simultaneously removing the extra data processing steps outside the scope of the datathon learning objectives.

We also evaluated the utility of unstructured "office-hour" sessions where participating teams could ask experienced members of the community for assistance with their projects. Despite holding multiple office-hour sessions at different times of the day throughout the datathon, we found that only 1 team attended any of the office-hour sessions in the 2023 VBC datathon. Because of this low attendance, we opted to remove synchronous office hours from the 2024 datathon programming and instead implemented a custom anonymous discussion forum via the datathon Slack communication channel where participants could ask questions anonymously that could be viewed and answered by anyone. Subjectively, we found that this asynchronous mode of communication made it easier for participants to seek help with their projects and observed greater engagement in public discussions after this feature was implemented. Future work is warranted to more rigorously evaluate the utility of such interventions.

Finally, we acknowledge that disciplines such as medicine and computer science have historically seen disproportionate participation from trainees of certain racial, socioeconomic, and gender backgrounds. These systemic trends well described in prior work [31,32] are reproduced in our datathons as well (Table 2); as topics such as data science, VBC, and generative AI become increasingly important components of modern health care, it is crucial that all future clinicians from all backgrounds can interact meaningfully with these concepts and their applications. We hope that future work will explore how to reduce barriers to participation for historically marginalized groups of trainees.

Related Work

The majority of prior work published in related literature details short datathons lasting a few days at a single physical location with a different target participant group. Hochheiser et al [33] describe a 2-day datathon consisting of 5 participating teams of clinicians and informaticians working on elucidating potential sources of bias within health care ML models. While their synchronous datathon model may be suitable for participants at a single physical site, such a model was intractable for our purposes as participating trainees were distributed across multiple institutions and time zones. Sobel et al [15] detail a similar datathon at a single physical location, but their study was primarily conducted with undergraduate and graduate students with pre-existing computational backgrounds, as opposed to undergraduate medical trainees from institutions granting postdoctoral fellowships as in our case. Anecdotally, we found evidence of similar initiatives held at the institutional level, such as the Digital Critical Care Datathon [34], the New York University Health Tech Datathon [35], and the Society of Critical Care Medicine Datathon [36]; each of these were single-institution initiatives with different datathon design constraints. To our knowledge, we are the first to describe a trainee-led, multi-institutional, asynchronous datathon effort and demonstrate preliminary evidence of its efficacy and potential role in the future of medical education.

Limitations

There are also limitations associated of our study. Firstly, our datathon was coordinated digitally with participants joining from across the United States. While we acknowledge there are both benefits and drawbacks to a datathon (as opposed to their in-person counterparts), we leave a rigorous comparison between their utilities in modern medical education paradigms for future work. Furthermore, both participating in the datathon and completing the postparticipation survey were opt-in processes, and so it is unclear how our findings would translate to undergraduate medical trainees who might have systematically chosen to not participate in the datathons—for example, potential participants who were more hesitant in learning about AI and data science practices in medicine and those whose medical school coursework made concurrent participation in the datathon unfeasible. Our survey results also exclude individuals who initially signed up to express interest in participating in the datathon but ultimately decided not to submit a final project. Given the opt-in design of our survey study, we were unable to assess the efficacy of our datathons for these individuals. Future work might evaluate how similar initiatives could scale across more diverse participant profiles and foster participation from student trainees of all backgrounds and perspectives. Finally, our postparticipation survey makes use of retrospective questions that ask participants to subjectively reflect on their skill development, rather than an objective evaluation of participant skills through a standardized programming examination. We chose this study design for two primary reasons: (1) because of the diverse array of participant projects, the skill sets that they developed through their

XSL•FO
RenderX

participation in the datathon are likely equally diverse, making a single standardized examination challenging to construct; and (2) in our initial efforts in organizing the datathon, we hypothesized that the survey response rate would be too low to adequately power our study if we asked participants to complete opt-in programming examinations. We leave the exploration of using more standardized assessments of programming skill competencies attained through datathon initiatives as future work.

Conclusions

Ultimately, the goal of this datathon was to provide opportunities for trainees—especially medical students—to improve their data skills and to identify data-driven solutions to problems in health care. Participants practiced using hands-on data science and artificial intelligence to explore meaningful clinical problems and voiced a collective interest in continuing to participate in similar initiatives in the future. Overall, our results and collective experiences suggest that datathons can be valuable within undergraduate medical education.

Acknowledgments

The authors thank the teams at Hugging Face, Doximity, Merck, Conduce Health, and AvoMD for their generous support and sponsorship of the 2023 and 2024 MDplus datathon events, and the datathon judges (listed in no particular order) Reza Alavi, MD, MHS, MBA; Caroline Berchuck, MD, MPH; Amit Phull, MD; Sid Salvi; Kathryn Teng, MD, MBA, FACP; Andrea Green; David Dupee, MD; Marc Triola, MD; and Jannes Jegminat, PhD for lending their time and expertise in helping run a successful datathon event. The authors also thank Eric Shan, Katie Link, Julia Bondar, Vamsi Chodisetty, and other members of the MDplus community and executive team for their help in organizing and running the datathon event. MSY was supported by NIH F30 MD020264, and ELL was supported by NIH T32 training grant GM146636; the content in this manuscript is solely the responsibility of the authors and does not necessarily represent the official view of the NIH.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Introduction to Python Datathon Tutorial.
[PDF File, 150 KB - mededu_v11i1e63602_app1.pdf]

Multimedia Appendix 2
Judging Rubric for Datathon Finalist Showcase Event.
[XLSX File, 13 KB - mededu_v11i1e63602_app2.xlsx]

Multimedia Appendix 3
Post-Datathon Survey.
[DOCX File, 18 KB - mededu_v11i1e63602_app3.docx]

References

1. Hege I, Kononowicz AA, Adler M. A clinical reasoning tool for virtual patients: design-based research study. JMIR Med Educ 2017 Nov 2;3(2):e21. [doi: 10.2196/mededu.8100] [Medline: 29097355]
2. Pongdee T, Larson NB, Divekar R, Bielinski SJ, Liu H, Moon S. Automated identification of aspirin-exacerbated respiratory disease using natural language processing and machine learning: algorithm development and evaluation study. JMIR AI 2023 Jun 12;2:e44191. [doi: 10.2196/44191] [Medline: 39105270]
3. Chae A, Yao MS, Sagreiya H, et al. Strategies for implementing machine learning algorithms in the clinical practice of radiology. Radiology 2024 Jan;310(1):e223170. [doi: 10.1148/radiol.223170] [Medline: 38259208]
4. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. N Engl J Med 2023 Mar 30;388(13):1201-1208. [doi: 10.1056/NEJMra2302038] [Medline: 36988595]
5. Kendale S, Bishara A, Burns M, Solomon S, Corriere M, Mathis M. Machine learning for the prediction of procedural case durations developed using a large multicenter database: algorithm development and validation study. JMIR AI 2023 Sep 8;2:e44909. [doi: 10.2196/44909] [Medline: 38875567]
6. Seth P, Hueppchen N, Miller SD, et al. Data science as a core competency in undergraduate medical education in the age of artificial intelligence in health care. JMIR Med Educ 2023 Jul 11;9:e46344. [doi: 10.2196/46344] [Medline: 37432728]
7. Arango-Ibanez JP, Posso-Nuñez JA, Díaz-Solórzano JP, Cruz-Suárez G. Evidence-based learning strategies in medicine using AI. JMIR Med Educ 2024 May 24;10:e54507. [doi: 10.2196/54507] [Medline: 38801706]
8. Shimizu I, Kasai H, Shikino K, et al. Developing medical education curriculum reform strategies to address the impact of generative AI: qualitative study. JMIR Med Educ 2023 Nov 30;9:e53466. [doi: 10.2196/53466] [Medline: 38032695]

XSL•FO
RenderX

9. Furlan R, Gatti M, Mene R, et al. Learning analytics applied to clinical diagnostic reasoning using a natural language processing-based virtual patient simulator: case study. JMIR Med Educ 2022 Mar 3;8(1):e24372. [doi: 10.2196/24372] [Medline: 35238786]
10. Sun L, Yin C, Xu Q, Zhao W. Artificial intelligence for healthcare and medical education: A systematic review. Am J Transl Res 2023;15(7):4820-4828. [Medline: 37560249]
11. Pupic N, Ghaffari-zadeh A, Hu R, et al. An evidence-based approach to artificial intelligence education for medical students: A systematic review. PLOS Digit Health 2023;2(11):e0000255. [doi: 10.1371/journal.pdig.0000255]
12. Civaner MM, Uncu Y, Bulut F, Chalil EG, Tatli A. Artificial intelligence in medical education: A cross-sectional needs assessment. BMC Med Educ 2022 Nov 9;22(1):772. [doi: 10.1186/s12909-022-03852-3] [Medline: 36352431]
13. Gray K, Slavotinek J, Dimaguila GL, Choo D. Artificial intelligence education for the health workforce: expert survey of approaches and needs. JMIR Med Educ 2022 Apr 4;8(2):e35223. [doi: 10.2196/35223] [Medline: 35249885]
14. Chan KS, Zary N. Applications and Challenges of Implementing Artificial Intelligence in Medical Education: Integrative Review. JMIR Med Educ 2019 Jun 15;5(1):e13930. [doi: 10.2196/13930] [Medline: 31199295]
15. Sobel J, Almog R, Celi L, Yablowitz M, Eytan D, Behar J. How to organise a datathon for bridging between data science and healthcare? Insights from the Technion-Rambam machine learning in healthcare datathon event. BMJ Health Care Inform 2023 Sep;30(1):e100736. [doi: 10.1136/bmjhci-2023-100736] [Medline: 37696642]
16. Oyetade K, Zuva T, Harmse A. Educational benefits of hackathon: a systematic literature review. WJET 2022;14(6):1668-1684 [FREE Full text] [doi: 10.18844/wjet.v14i6.7131]
17. Silver JK, Binder DS, Zubcevik N, Zafonte RD. Healthcare hackathons provide educational and innovation opportunities: a case study and best practice recommendations. J Med Syst 2016 Jul;40(7):177. [doi: 10.1007/s10916-016-0532-3] [Medline: 27277278]
18. Barabucci G, Shia V, Chu E, Harack B, Laskowski K, Fu N. Combining multiple large language models improves diagnostic accuracy. NEJM AI 2024 Oct 24;1(11). [doi: 10.1056/AIcs2400502]
19. Goh E, Gallo R, Hom J, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. JAMA Netw Open 2024 Oct 1;7(10):e2440969. [doi: 10.1001/jamanetworkopen.2024.40969] [Medline: 39466245]
20. Savage T, Nayak A, Gallo R, Rangan E, Chen JH. Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine. NPJ Digit Med 2024 Jan 24;7(1):20. [doi: 10.1038/s41746-024-01010-1] [Medline: 38267608]
21. Ong JCL, Chang SYH, William W, et al. Ethical and regulatory challenges of large language models in medicine. Lancet Digit Health 2024 Jun;6(6):e428-e432. [doi: 10.1016/S2589-7500(24)00061-X] [Medline: 38658283]
22. Omiye JA, Lester JC, Spichak S, Rotemberg V, Daneshjou R. Large language models propagate race-based medicine. NPJ Digit Med 2024;6(1):195. [doi: 10.1038/s41746-023-00939-z]
23. Hager P, Jungmann F, Holland R, et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nat Med 2024 Sep;30(9):2613-2622. [doi: 10.1038/s41591-024-03097-1] [Medline: 38965432]
24. Teisberg E, Wallace S, O'Hara S. Defining and implementing value-based health care: a strategic framework. Acad Med 2020 May;95(5):682-685. [doi: 10.1097/ACM.0000000000003122] [Medline: 31833857]
25. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 2023 Jan 3;10(1):1. [doi: 10.1038/s41597-022-01899-x] [Medline: 36596836]
26. Ben Abacha A, Yim WW, Fan Y, Lin T. An empirical study of clinical note generation from doctor-patient encounters. Presented at: Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics; 2023; Dubrovnik, Croatia p. 2291-2302. [doi: 10.18653/v1/2023.eacl-main.168]
27. Jin D, Pan E, Oufattole N, Weng WH, Fang H, Szolovits P. What disease does this patient have? A large-scale open domain question answering dataset from medical exams. Appl Sci (Basel) 2021;11(14):6421. [doi: 10.3390/app11146421]
28. Ji S, Li X, Huang Z, Cambria E. Suicidal ideation and mental disorder detection with attentive relation networks. Neural Comput & Applic 2022 Jul;34(13):10309-10319. [doi: 10.1007/s00521-021-06208-y] [Medline: 33746365]
29. MDplus. MDplus DS & AI - datathon. URL: http://ai.mdplus.community/datathon/2023 [accessed 2024-01-18]
30. Likert R. A technique for the measurement of attitudes. Arch Psychol 1932;140:1-55.
31. Hendricks-Sturrup R, Simmons M, Anders S, et al. Developing ethics and equity principles, terms, and engagement tools to advance health equity and researcher diversity in AI and machine learning: modified Delphi approach. JMIR AI 2023 Dec 6;2:e52888. [doi: 10.2196/52888] [Medline: 38875540]
32. Aggarwal R, Farag S, Martin G, Ashrafian H, Darzi A. Patient perceptions on data sharing and applying artificial intelligence to health care data: cross-sectional survey. J Med Internet Res 2021 Aug 26;23(8):e26162. [doi: 10.2196/26162] [Medline: 34236994]
33. Hochheiser H, Klug J, Mathie T, et al. Raising awareness of potential biases in medical machine learning: Experience from a Datathon. medRxiv 2024 Nov 2:2024.10.21.24315543. [doi: 10.1101/2024.10.21.24315543] [Medline: 39502657]
34. Elbers P, Thoral P, Bos LDJ, Greco M, Wendel-Garcia PD, Ercole A. The ESICM datathon and the ESICM and ICMx data science strategy. Intensive Care Med Exp 2024 Mar 12;12(1):29. [doi: 10.1186/s40635-024-00615-w] [Medline: 38472595]
35. Health Tech Datathon. NYU grossman school of medicine. URL: https://med.nyu.edu/our-community/health-technology/events/health-tech-datathon [accessed 2025-04-07]

XSL•FO
RenderX

36. Datathon. Society of Critical Care Medicine (SCCM): The Intensive Care Professionals. URL: https://sccm.org/research/discovery-research-network/datascience/datathon [accessed 2024-12-01]

Abbreviations

LLM: large language model
MIMIC: Medical Information Mart for Intensive Care
ML: machine learning
VBC: value-based care

Edited by B Lesselroth; submitted 24.06.24; peer-reviewed by M Knopp, PM Naliyatthaliyazchayil, S Purkayastha; revised version received 02.12.24; accepted 25.02.25; published 16.04.25.

Please cite as:
Yao MS, Huang L, Leventhal E, Sun C, Stephen SJ, Liou L
Leveraging Datathons to Teach AI in Undergraduate Medical Education: Case Study
JMIR Med Educ 2025;11:e63602
URL: https://mededu.jmir.org/2025/1/e63602
doi:10.2196/63602

© Michael Steven Yao, Lawrence Huang, Emily Leventhal, Clara Sun, Steve J Stephen, Lathan Liou. Originally published in JMIR Medical Education (https://mededu.jmir.org), 16.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effect of Immersive Virtual Reality Teamwork Training on Safety Behaviors During Surgical Cases: Nonrandomized Intervention Versus Controlled Pilot Study

Lukasz Mazur[1,2], PhD; Logan Butler[1], MD; Cody Mitchell[1], BS; Shaian Lashani[1], BS; Shawna Buchanan[1], BSN; Christi Fenison[1], MA; Karthik Adapa[1], MD, PhD; Xianming Tan[3], PhD; Selina An[4], MD; Jin Ra[4], MD

[1]Department of Radiation Oncology, Division of Healthcare Engineering, School of Medicine, University of North Carolina, Campus Box 7512, Chapel Hill, NC, United States

[2]School of Information and Library Science, University of North Carolina, Chapel Hil, NC, United States

[3]Gillings School of Global Public Health, University of North Carolina, Biostatistics, Chapel Hill, NC, United States

[4]Department of Surgery, School of Medicine, University of North Carolina, Chapel Hill, NC, United States

Corresponding Author:
Lukasz Mazur, PhD
Department of Radiation Oncology, Division of Healthcare Engineering, School of Medicine, University of North Carolina, Campus Box 7512, Chapel Hill, NC, United States

Abstract

Background: Approximately 4000 preventable surgical errors occur per year in the US operating rooms, many due to suboptimal teamwork and safety behaviors. Such errors can result in temporary or permanent harm to patients, including physical injury, emotional distress, or even death, and can also adversely affect care providers, often referred to as the “second victim.”

Objective: Given the persistence of adverse events in the operating rooms, the objective of this study was to quantify the effect of an innovative and immersive virtual reality (VR)–based educational intervention on (1) safety behaviors of surgeons in the operating rooms and (2) sense-making regarding the overall training experience.

Methods: This mixed methods pre- versus postintervention pilot study was conducted in a large academic medical center with 55 operating rooms. Safety behaviors were observed and quantified using validated Teamwork Evaluation of Non-Technical Skills instrument during surgical cases at baseline (101 observations; 83 surgeons) and postimmersive VR based intervention (postintervention: 24 observations within each group; intervention group [with VR training; 10 surgeons] and control [no VR training; 10 surgeons]). VR intervention included a 45-minute immersive VR-based training incorporating a pre- and postdebriefing based on Team Strategies and Tools to Enhance Performance and Patient Safety (TeamSTEPPS) principles to improve safety behaviors. A 2-tailed, 2-sample *t*-test with adjustments for multiplicity of the tests was used to test for significance in observable safety behaviors between the groupings. The debriefing data underwent analysis through the phenomenological analysis method to gain insights into how participants interpreted the training.

Results: Preintervention, all safety behaviors averaged slightly above “acceptable” scores, with an overall average of 2.2 (range 2‑2.3; 0‑3 scale). The 10 surgeons that underwent our intervention showed statistically significant ($P<.05$) improvements in 90% (18/20) of safety behaviors when compared to the 10 surgeons that did not receive the intervention (overall average 2.5, range 2.3‑2.7 vs overall average 2.1, range 1.9‑2.2). Our qualitative analysis based on 492 quotes from participants suggests that the observed behavioral changes are a result of an immersive experience and sense-making of key TeamSTEPPS training concepts.

Conclusions: VR-based immersive training intervention focused on TeamSTEPPS principles seems effective in improving safety behaviors in the operating rooms as quantified via observations using the Teamwork Evaluation of Non-Technical Skills instrument. Further research with larger, more diverse sample sizes is needed to confirm the generalizability of these findings.

International Registered Report Identifier (IRRID): RR2-10.2196/40445.

(JMIR Med Educ 2025;11:e66186) doi:10.2196/66186

KEYWORDS

Teamwork Evaluation of Non-Technical Skills; TENTS; Team Strategies and Tools to Enhance Performance and Patient Safety; TeamSTEPPS; immersive virtual reality; virtual reality; VR; safety behavior; surgical error; operating room; OR; training intervention; training; pilot study; nontechnical skills; surgery; surgical; patient safety; medical training; medical education

XSL•FO
RenderX

Introduction

High-quality health care necessitates ongoing efforts to reduce the occurrence of medical errors [1]. Surgical patients face heightened risks of adverse outcomes related to errors due to the invasive nature of surgical procedures [2]. It is estimated that more than 4000 preventable surgical errors occur annually on a national scale [1,2]. Such errors can result in temporary or permanent harm to patients, including physical injury, emotional distress, or even death, and can also adversely affect care providers, often referred to as the "second victim." A notable example is the unintended retention of foreign objects, which is believed to happen at least once in every 5500 surgeries [3]. This can lead to the need for reoperation, extended hospital stays, and complications such as sepsis. Furthermore, the average additional cost associated with each incident of unintended retention is estimated to exceed US $200,000 [4]. Common underlying causes of surgical errors identified by the Joint Commission include the lack of established policies and procedures, issues related to hierarchy and intimidation, ineffective communication among care team, and the failure of staff to relay pertinent patient information [5]. Additionally, factors such as excessive workload, time constraints, and burnout are linked to increased error rates [5]. Addressing these root causes has proven challenging, as complex health care delivery systems tend to evolve over time, leading to the emergence of new failure sources and pathways. Teamwork skills are often essential for preventing such errors that could lead to patient harm [6]. [7], To address these issues, the Team Strategies and Tools to Enhance Performance and Patient Safety (TeamSTEPPS) framework was specifically designed as a resource to help health care providers improve patient safety behaviors through effective communication, leadership, situation monitoring, and mutual support [8]. [9,10] By using the TeamSTEPPS framework, it is possible to assess the use and quality of patient safety education and behaviors among operating room staff and establish a baseline for improvement.

Virtual reality (VR) is a digital technology that enables a virtual manifestation of the real world [11]. VR provides a more captivating experience compared to viewing a conventional video, as it fully envelopes the viewer within the narrative [12]. In VR, the audience becomes an integral part of the story rather than merely an onlooker. The advantages of immersive VR primarily include: (1) viewers are placed within a 360-degree environment, where each movement of the head unveils new dimensions of the scene. Conversely, traditional video confines viewers to a fixed perspective on a flat screen. (2) Participants in VR actively engage with the narrative, rather than remaining passive spectators. (3) A profound experiential learning opportunity, allowing participants to engage with the contextual realities of a surgical error in the surgical environment.

While live mock simulations with standardized patients are used for certain health care scenarios, the logistical demands of accurately recreating a surgical environment are substantial. These include the coordination of a full surgical team, time spent in an operating room, and the creation of a realistic setting, along with patient representation and various special effects that must align with the error, necessitating cleanup and reset after each simulation. The extensive resources required make this approach neither cost-effective nor scalable for providing multiple realistic experiences for individual learners. Furthermore, a live mock simulation often fails to address the comprehensive needs of learners, as it does not provide insights into the broader contributing factors and repercussions that extend beyond the confines of the operating room. While this disconnect may not significantly affect certain types of learning, such as factual recall, the deeper cognitive processing of theoretical scenarios, particularly in the context of complex situations, could be enhanced through more experiential learning methods such as VR [13]. Experiential learning fosters a personalized and cognitive interaction with educational content, highlighting the relationship between learning and its practical application in the real world. If health care workers are to engage in sense-making regarding the intricate realities associated with adverse events, VR may provide distinct advantages over 2D content and live mock simulations by offering an immersive perspective of events as they would unfold in real life.

In recent years, VR has seen increasing use for safety training across industries [14-17]. VR safety training has great potential, as it allows trainees to experience complex, challenging situations that are difficult to replicate in the real world due to ethical, cost, and time constraints. However, there is a lack of research examining the efficacy of immersive VR-based interventions focused on safety education and behaviors as proposed by the TeamSTEPPS framework. Thus, given the persistence of adverse events in the operating rooms, the objective of this pilot study was to quantify the effect of an innovative and immersive VR-based educational intervention focused on TeamSTEPPS on (1) safety behaviors of surgeons in the operating rooms and (2) sense-making regarding the overall training experience and contributing factors associated with the surgical error.

Methods

Ethical Considerations

This study obtained ethics approval via the Institutional Review Board of the University of North Carolina at Chapel Hill (22 - 1150). Participants were provided with and signed an informed consent form before engaging with activities related to this study. Participants in the intervention group were provided with a small token of appreciation ($25) for compleating the VR intervention process. Patients or the public were not involved in the design, conduct, reporting, or dissemination plans of our research. All results are reported in an aggregate manner to ensure the privacy and confidentiality of participants. The protocol for the full mixed methods pre-versus postintervention study design was published in *JMIR Research Protocols* [18].

Recruitment

A scripted email with a flyer was sent through a listserv to inform prospective participants about this study. Our research team was on standby to answer any questions from prospective participants, and the principal investigators' contact information was available on the flyer. First, for the baseline measurement, a volunteer sample of 83 surgeons (35 attendings, 41 residents,

XSL•FO
RenderX

and 7 fellows; 36 female and 47 male) was enrolled and participated. Surgeons volunteered by providing verbal consent to be observed and scored for safety behaviors before each surgical case. One surgeon refused to participate (no reasons stated). For the pilot study, 10 out of 83 (12%) surgeons (7 attendings and 3 residents; 5 female and 5 male) volunteered to undergo the immersive VR-based educational intervention and to undergo the follow-up observations and scoring. An additional 10 of 83 (12%) surgeons (10 attendings and 0 residents; 3 female and 7 male) volunteered to undergo follow-up observation and scoring without being exposed to the intervention.

Study Design

This mixed methods pre- versus postintervention study with a baseline and intervention or control groups was conducted at a large academic medical center with 55 operating rooms. For the baseline measurement, safety behaviors were quantified while observing 101 surgical cases from October 31, 2022, to February 21, 2023, with 83 surgeons. For the pilot study, 24 observations within each group were conducted from April 17, 2023, to November 2, 2023, with 10 surgeons in the intervention and control group, respectively. Data was collected using the Teamwork Evaluation of Non-Technical Skills (TENTS) instrument. We opted for the TENTS instrument instead of alternatives such as the TeamSTEPPS Team Performance Observation Tool due to our requirement for a more focused evaluation of nontechnical teamwork aspects. TENTS is specifically designed to assess nontechnical dimensions of teamwork, encompassing team member interactions, task delegation, and information management, which makes it particularly relevant in situations where technical skills are not the primary concern. Additionally, TENTS provides a more detailed observational framework, allowing for a comprehensive evaluation of specific teamwork behaviors, in contrast to the broader categories covered by the TeamSTEPPS Team Performance Observation Tool. In our study, we used TENTS items 1A to 4D to assess individual behaviors of surgeons in independent settings (except for the 3 residents who were part of the intervention group), while items 5 and 6 were used to evaluate the overall team functioning and leadership. To train the observers for scoring items 5 and 6, we assembled a team of 17 medical student volunteers who were tasked with conducting observations using the TENTS instrument. They used a simplified scoring scale ranging from 0 to 3, where 0 indicates expected behavior not observed, 1 signifies observed behavior that was poorly executed or counterproductive, 2 denotes acceptable performance, and 3 represents excellent performance. The students received 1.5 hours of instruction on TeamSTEPPS patient safety behaviors and were trained to apply the validated TENTS instrument consistently. Following this, they participated in facility tours and were guided through the observation protocol, which included a practical demonstration of the TENTS scoring process. The students were assigned to surgical cases by aligning their weekly schedules with the relevant cases, while being blinded to the treatment status of the surgeons during postintervention observations. TENTS scores were recorded during real-time observations of surgical procedures using paper forms, and students were required to provide written justifications for behaviors rated as 1 (poor) or 3 (excellent) to offer context for their evaluations.

Interventions or Exposures

We used state-of-the-art filming equipment to capture a 360° view of the event and the perspective of those involved in the error event and contributing events. We built the scripts for the scenes, recruited actors (attendings, residents, students, and administrators) with lived experiences in health care to help with the filming, identified filming locations, rehearsed all the scenes, and filmed our scenes. This training was delivered to the participants using a VR head-mounted display to ensure an immersive environment (see Multimedia Appendix 1 for a 1-minute summary clip of the training). Specifically, we used the Pico Neo 3 Pro Eye headset (PICO Technology Co Ltd) with 6DoF VR hardware or software to administer the training.

Participants were exposed to a 45-minute immersive VR-based training based on TeamSTEPPS principles to improve safety behaviors. Overall, our overarching training story is focused on a human error that occurs at the operating table but is caused by many factors, as explained by the James Reason Swiss Cheese Model [19] and Human Factors Analysis and Classification System (HFACS) [20]. Training involved a standardized pre- and postbriefing aimed at comprehending and identifying potential behavioral enhancements within the training narrative, grounded in TeamSTEPPS principles. Specifically, we sought to collect participants' perspectives on the overall training experience and their interpretation of the TENTS and HFACS-related factors that contributed to the patient safety incident illustrated in the VR training. The debriefing data underwent analysis through the phenomenological analysis method [21] to gain insights into how participants interpreted the training and the patient safety incidents they encountered, thereby refining their understanding, behavior, and commitment to TeamSTEPPS practices. This qualitative analysis was performed by 2 researchers, a senior surgical attending and a fourth-year surgical resident, using the data frame theory of sense-making [22]. The primary objective of this qualitative effort was to enhance our understanding of the main outcomes and measures outlined below.

Main Outcomes and Measures

The primary outcome measures were the pooled average and 95% CI of observed TeamSTEPPS related behaviors quantified using the validated TENTS instrument (including 20 types of safety behaviors across 4 domains [communication, leadership, situation monitoring, and mutual support] scored from 0=expected but not observed, 1=observed but poorly performed or counterproductive, 2=observed and acceptable, and 3=observed and excellent).

Statistical Analysis

A 2-tailed, ANOVA was used to test for significance between the groupings. For given comparators (eg, [1] vs [2] or [2] vs [3]), we used a false discovery rate (FDR) control (the Benjamini-Hochberg procedure) to adjust the resulting *P* values (from the 2-sample *t*-tests) to account for the multiplicity of the tests. We claimed a result to be statistically significant if the adjusted *P* value is less than .05. All analyses were performed

XSL•FO
RenderX

with R statistical software. Data analysis was conducted from August 10 to December 1, 2023.

Results

Preintervention, all safety behaviors averaged slightly above 2 ("acceptable"), with an overall average of 2.2 (range 2 - 2.3). There was no significant difference between the intervention and the control group at the preintervention stage. The results indicated that the 10 surgeons that underwent our intervention showed statistically significant ($P<.05$) improvements in 90% (18/20) of safety behaviors when compared to the 10 surgeons that did not receive the intervention (overall average 2.5, range 2.3 - 2.7 vs overall average 2.1, range 1.9 - 2.2; Table 1). Our qualitative analysis revealed 492 individual quotes. The results suggest that the observed behavioral changes are a result of a sense-making emerging from 5 specific themes as discussed below.

XSL•FO
RenderX

Table . Summary of results (baseline vs intervention).

	TENTS[a] behavior	Baseline, mean (SD)	Intervention, mean (SD)	*P* value
1a	Communicates and receives information appropriately	2.227 (0.53)	2.708 (0.455)	.001
1b	Comfortable speaking up and asking questions	2.237 (0.452)	2.625 (0.484)	.001
1c	Responses to feedback between team members	2.135 (0.504)	2.458 (0.498)	.02
1d	Communicates and receives information to or from patient	2.097 (0.433)	2.261 (0.439)	.14
1e	Uses language in urgent situations appropriately	2.133 (0.505)	2.444 (0.497)	.04
1f	Uses teamwork tools	2.26 (0.464)	2.542 (0.498)	.03
1g	Learns together, focuses on improvement following a problem	2.222 (0.451)	2.55 (0.497)	.02
2a	Leaders effectively manage team during their roles	2.274 (0.493)	2.708 (0.455)	.001
2b	Verbalizes plan: intentions, recommendations, or timeframes	2.247 (0.501)	2.583 (0.571)	.02
2c	Delegates tasks appropriately	2.104 (0.369)	2.458 (0.498)	.007
2d	Instructs as appropriate to the situation	2.281 (0.475)	2.542 (0.498)	.03
3a	Pays attention to surroundings or environment	2.11 (0.526)	2.5 (0.5)	.005
3b	Aware of each other, contributions, strengths, and weaknesses	2.208 (0.433)	2.5 (0.5)	.02
3c	Verbalizes adjustments in plan as changes occur	2.236 (0.428)	2.524 (0.499)	.03
4a	Willingness to ask for help or additional resources	2.26 (0.464)	2.708 (0.455)	.001
4b	Willingness to support others across different roles	2.253 (0.437)	2.417 (0.493)	.15
4c	Accomplishes and prioritizes tasks appropriately	2.103 (0.338)	2.333 (0.471)	.04
4d	Employs conflict resolution	2.038 (0.341)	2.316 (0.465)	0.03
5	Rating of how well the team functioned as a whole	2.258 (0.44)	2.708 (0.455)	0.001
6	Rate how well leaders functioned and how the team responded	2.247 (0.434)	2.708 (0.455)	0.001

[a]TENTS: Teamwork Evaluation of Non-Technical Skills.

Discussion

Principal Results

Overview

The results show that participants exposed to our intervention displayed improved levels of safety behaviors, as quantified by the TENTS instrument, in 90% (18/20) of the safety behaviors measured. Specifically, quantitative data suggests that surgical teams were more effective in developing and maintaining a dynamic awareness of the situation in the operating room. This was achieved by assembling and understanding data from various sources (eg, patient, team, time, displays, and equipment), and using strong communication and leadership skills to think ahead and provide clear direction while being

XSL•FO
RenderX

considerate of individual team members' needs. Importantly, these improvements were observed not only at the individual level, as shown by the TENTS instrument, but also in the overall team functioning and leadership, as indicated by aggregate measures of "how well the team functioned as a whole," "how well the leaders functioned," and "how the team responded."

The 2 behaviors that did not reach statistical significance, despite trending positively, were 1d (communicating and receiving information with patients), and 4b (willingness to support others across different roles) (Table 1). For the communication behavior, very few such interactions were observed during the surgical cases, limiting our assessment of this behavior. Regarding the willingness to support others across roles, this was the highest-scoring behavior at baseline, suggesting it may have been challenging to improve further. This implies that in the dynamic and complex surgical environment, team members may struggle to step outside their designated roles to provide support to one another, as they focus intently on delivering excellent patient care and ensuring safety within their specific responsibilities.

Our qualitative analysis suggests that these behavioral improvements materialized from the enhanced understanding of skills needed in the operating room by reinforcing critical behaviors related to the sense-making of themes presented below.

Need for Effective Teamwork and Communication

The VR training modeled effective versus ineffective team communication, demonstrating how dismissing or ignoring concerns can lead to errors and decrease team trust. Participants may have practiced assertive communication, such as how a resident can escalate a concern when an anesthesiologist is distracted or how surgical technology can improve instrument handling. VR also emphasized calling out errors in a constructive way, rather than scolding or ignoring them.

The most consistent topics addressed by the subjects during the poststudy interview were teamwork and communication, and how these traits are essential for a well-functioning operating room. Participants noted how the VR training vividly illustrated both the consequences of poor communication and the benefits of a cohesive team. The training emphasized that errors often arise not solely from technical mistakes but from an inability to effectively relay and escalate concerns. The lack of teamwork and communication in the ineffective VR scenario was particularly alarming to participants. One individual highlighted the disconnect among team members: "There was clearly not a deep relationship between the surgeon, the resident, the scrub, the circulator, and the anesthesiologist; they just seemed completely disconnected [and] in their own worlds." Another key issue was the absence of assertive communication when concerns arose. One participant recalled: "The resident did pick up on the change in the heart rate tone, questioned the anesthesiologist about it, who blew her off." The absence of closed-loop communication was another prevalent theme. Participants remarked on how essential feedback loops were often missing in the ineffective scenarios, which contributed to preventable errors: "No closed-loop communication. Just a lot of people suggesting things but the other person either wasn't listening or just kind of ignored it and moved along. Even when it was something that could have prevented safety issues." Overall, the poststudy interviews reinforced that teamwork and communication are foundational to operating room effectiveness. The VR training provided a powerful demonstration of how dismissing concerns, failing to engage in open dialogue, and neglecting structured communication can significantly hinder patient safety. By recognizing these pitfalls and emphasizing assertive, closed-loop communication, participants reported a newfound appreciation for fostering a more cohesive and communicative operating room team.

Emphasis on Empathetic Workplace Culture and Psychological Safety

VR heightened participants' awareness of how fatigue, dismissive communication, and team dynamics impact psychological safety in the operating room. By simulating real-world scenarios, it demonstrated how exhaustion and distractions compromise decision-making, how seemingly minor quips or sarcasm can erode team cohesion, and how validating trainee concerns fosters a culture of safety and respect. This reinforced the importance of clear, professional communication and proactive leadership in creating a supportive and effective surgical intervention.

The VR training underscored how the psychological safe environment of the operating room directly impacts team effectiveness, patient safety, and overall workplace culture. Participants became more empathetic and attuned to how fatigue, dismissive behavior, and team dynamics can create a toxic versus supportive surgical setting. By immersing participants in scenarios where exhaustion led to oversight, sarcasm eroded trust, and concerns were either validated or dismissed, the training highlighted the importance of maintaining a psychologically safe and healthy work environment. One of the most striking realizations was how fatigue and distraction—often seen as inevitable in surgical practice—could significantly impair decision-making and communication. As one participant noted, "Exhaustion, lack of sleep, and lack of focus on the attending's part... distraction from the anesthesiologist... these are indicators that their wellness score is probably not stellar." Additionally, the training revealed how subtle, seemingly harmless behaviors can undermine psychological safety. Participants observed how sarcastic remarks or casual quips, even when meant humorously, created an environment where individuals felt less comfortable speaking up. "There were a lot of quips...I don't think they contributed too much. And they can be detrimental." Another crucial takeaway was the need to legitimize concerns when raised, rather than allowing the pressures of a high caseload to override safety. One participant reflected, "I think we can always do better legitimizing people raising concerns...the pressure to feel rushed and move quickly ... can obviously be counterproductive." Perhaps most telling was the recognition that team dynamics set the tone for the entire operating room. When interpersonal relationships are strained, it affects everyone, from the attending surgeon to the anesthesia technician. One participant encapsulated this sentiment: "Because I think we've all been in rooms where [if] the staff doesn't get along …it makes it miserable for everyone." The VR training effectively demonstrated how the psychological

XSL•FO
RenderX

environment of the operating room shapes both patient safety and team dynamics, making participants more aware of the subtle but powerful ways fatigue, communication styles, and validation of concerns impact surgical outcomes. By immersing participants in realistic scenarios, the training seemed to inspire an appreciation for operating rooms, where they do have a psychologically healthy environment.

Need for Leadership With Personal Responsibility and Accountability

VR simulation highlighted the ripple effect of individual actions—how fatigue, inattention, or lack of engagement from a team member impacts the whole operating room. It also reinforced that leaders (attendings, residents, or nurses) set the tone for safety culture, whether by ensuring protocols are followed or by fostering an environment where concerns can be raised. The VR training may have helped participants recognize their personal accountability in maintaining operating room safety and identifying behaviors that contribute to or undermine team effectiveness.

The training reinforced the critical role that leadership plays in shaping operating room culture, particularly in fostering accountability and ensuring adherence to safety protocols. Participants observed how leadership—or the lack thereof—had a cascading effect on communication, decision-making, and overall team cohesion. By placing participants in scenarios where leadership failures led to errors or unsafe practices, the training emphasized the responsibility of every operating room member to contribute to a culture of accountability. A key takeaway was the attending surgeon's responsibility in setting the tone for the operating room environment. As 1 participant remarked, "The attending surgeon sort of sets the tone in many ways, and by not looking into concerns...[and] cutting corners in order to be able to increase throughput...—that's just not good leadership." Beyond the attending, participants recognized how personal accountability extends to every team member. The training exposed moments where concerns were voiced quietly but never formally escalated, leading to missed opportunities to address potential safety issues. One participant reflected, "...everyone was making little comments, but no one was, again, like really saying them. They were all kind of talking to themselves...." The VR also made participants more aware of how the pressure to move quickly can lead to cutting corners, potentially compromising patient outcomes. One participant acknowledged, "We all get caught up in rush, rush, rush…and I've done that before, where you look through the labs, like, it's probably fine. It usually is. But what if it's not?" Ultimately, the training reinforced the idea that each case demands responsibility and accountability. The VR experience demonstrated how a disengaged or inattentive leader could undermine these traits by dismissing concerns or prioritizing efficiency over protocol.

Need for Stability

The VR training elicited participants' desires and personal experiences to have surgical teams that are consistent and connected. Understandably, those who are working in a high-stakes, high-stress environment would want to mitigate other factors that could lead to negative patient outcomes or contribute to workplace burnout and distress. The VR training showed participants what can happen with a more discordant or unstable working environment, which can mimic reality, and many participants were quick to point out how deleterious that can be to the dynamics of an operating room.

The training highlighted the critical role that stability plays in fostering an effective and supportive surgical environment. Participants emphasized the need for consistency in team composition, resource availability, and leadership presence—elements that are often taken for granted but can significantly impact patient safety and staff well-being. In the high-stakes, high-stress environment of the operating room, an unstable or discordant team structure can lead to inefficiencies, miscommunication, and increased burnout, all of which were vividly demonstrated in the VR scenarios. Many participants noted how instability, whether due to staffing shortages, systemic pressures, or administrative constraints, can disrupt operating room dynamics. One participant reflected on the broader hospital structure, stating, "It didn't seem like there was support from a majority of people...No one person can do it all." A major recurring theme was the strain placed on attendings who were expected to be in multiple places at once, highlighting the impact of systemic pressures on operating room stability. As 1 participant observed, "I mean, there were clearly systemic pressures for the attending to be in multiple places at one time, I would say primarily pressure for throughput, short staffing." This speaks to the broader challenges of balancing efficiency with quality care. This particularly resonated with our interviewees as it is something that a vast majority of health care workers can relate to. By experiencing the challenges of an unstable operating room environment with the constant, relatable pressure to do more, participants gained a deeper appreciation for the structures and policies needed to foster a more reliable and effective surgical setting.

Emphasis on Outcome and Attention to Detail

This VR module showed participants what can happen when means do not at all justify ends. In a system where outcomes, whether it be several cases completed or the speed of the operation, are prioritized over how those end points are achieved, it can lead to consequential errors. While operating rooms need to maintain a high level of efficiency, the consequences of doing so can come at the expense of the patient. It can be challenging for surgeons to balance their commitment to good patient care with intense pressures for increased efficiency and decreased case turnaround time.

The VR training underscored the risks of prioritizing efficiency and case volume over patient safety and procedural integrity. Participants recognized how a results-driven culture, where speed and throughput are emphasized over safe surgical practice, can lead to critical errors and compromise team effectiveness. While efficiency is a necessary component of modern surgical workflows, the VR scenarios illustrated the consequences of allowing productivity pressures to override fundamental principles of patient care. One participant stated, "many other factors—fatigue [and]the pressure from hospital administration for revenue and productivity," when reflecting on the systemic pressures driving an outcome-based mindset. The VR module

XSL•FO
RenderX

demonstrated how this tunnel vision manifests at different levels of the operating room team. One participant observed, "The resident's main mission was 'I gotta close so I can go to the other room.' The surgeon was like 'I gotta get these 12 cases done because that's what administration says.' The circulator was like 'We gotta get these cases done so we can all go home.'" Participants also reflected on the human cost of this approach, not just for the patients but for the surgical teams themselves. One particularly striking insight was, " ...it's fine to be efficient. It's not fine to be in a hurry. And I think that it's really, really important for all of us to think about...." The training allowed participants to experience the tension between efficiency and quality care, a concept that is ubiquitous among health care settings. By highlighting the potential dangers of an outcome-driven approach, the VR module reinforced the need for deliberate, methodical teamwork, where safety is never sacrificed in the name of speed or the bottom line.

The VR training also reinforced the critical importance of attention to detail in the operating room, highlighting how even small lapses in accuracy can lead to significant consequences. By immersing participants in real-world scenarios, the module demonstrated how factors such as fatigue, communication breakdowns, and time pressures can contribute to oversights. It emphasized that attention to detail is not just an individual responsibility but a collective effort, where every team member plays a role in maintaining surgical precision.

The training reinforced the critical role of accuracy in the operating room, emphasizing how seemingly minor lapses in attention to detail can have significant consequences for patient outcomes. Participants recognized that surgical safety is not solely dependent on technical skill but also on the thoroughness of preoperative preparation, intraoperative vigilance, and collective situational awareness. Several interviewees pointed out the dangers of neglecting critical details in the operating room. For example, many participants highlighted that the failure to properly assess a patient's anticoagulation status before surgery leads to cancellation of cases and delays in care. Another key theme was the lack of a shared mental model among the surgical team, leading to fragmented awareness of the patient's condition. As one participant described, "I'm not sure that there was a complete understanding of the entire situation... that global shared mental model of where everybody was, the implications of the decisions, and how things were happening was kind of missing." The scenarios also illustrated that once the ball was set in motion, no one had the care or will to change it. One participant noted, "It seemed like both the surgeon and the resident didn't have a good sense of who the patient was, what the case was, if they were on anticoagulation. And then even when the resident realized and brought it up to the attending, the attending was like, 'It's fine, it's too late, moving on.'" Ultimately, the VR experience reinforced the necessity of meticulous preparation, comprehensive team awareness, and an environment where concerns are acknowledged rather than dismissed. It demonstrated that attention to detail is not merely an individual responsibility but a collective effort, where every member of the surgical team plays a vital role in ensuring safe and effective patient care.

Comparison With Prior Work

This pilot study is the first to quantify the effects of an immersive VR-based educational intervention focused on improving TeamSTEPPS-related behaviors among surgeons in the operating room. Overall, the findings align with previous non–VR-based research highlighting the importance of teamwork training for enhancing soft skills critical to patient safety [23-27]. Our findings also align with the conclusions of the work by Abelson et al [11] that supports the motion that VR is a feasible solution for team-based training, and Gasteiger et al [12] that postulate that technical and nontechnical skills training programs using VR for health care staff may trigger perceptions of realism and deep immersion and enable easier visualization, interactivity, enhanced skills, and repeated practice in a safe environment, which in turn may improve skills and increase learning, knowledge, and learner satisfaction. Notably, prior work using VR to teach TeamSTEPPS for cesarean section surgery showed that the VR-based content improved teamwork competencies in interprofessional surgical teams [28]. By addressing the need for teamwork training, while using the TeamSTEPPS framework, and incorporating innovative educational technologies such as VR, this study demonstrated how collaboration among surgical team members can be enhanced [28]. Finally, in a randomized trial, Liaw et al [13] showed that learning outcomes did not show an inferiority of team training using VR when compared with live simulations, which supports the potential use of VR to substitute conventional simulations for communication team training.

However, many of these initiatives concentrate on skills pertinent to the immediate context of errors, such as communication and teamwork in the operating room, as well as technical competencies, while often neglecting training related to systemic flaws, the culture of patient safety, and the unreliable thought processes and behaviors that can lead to mistakes or hinder their prevention. We propose that trainees' comprehension of the factors leading to patient safety incidents as highlighted by the HFACS framework, their sense-making of safety behaviors as outlined by the TENTS tool, and their understanding of TeamSTEPPS principles in the context of the surgical error can lead to improvements in patient safety.

Limitations

While the findings of this study offer valuable insights, there are several important limitations to consider. First, the results are based on a single experiment with a relatively small sample size from 1 academic medical center. To address the small sample size, we used a 2-tailed, ANOVA for significance between the groupings using the FDR to adjust the resulting P values (from the 2-sample t-tests) to account for the multiplicity of the tests. For a 2-tailed, ANOVA for significance between the groupings, without the FDR control, the analysis would require approximately 68 participants to obtain a medium and to large effect size, the power level of 0.8, and an α of .05. Larger-scale studies could also take into account various confounding factors, such as training levels, gender, and race. Additionally, the possibility that more motivated individuals enrolled in the intervention group may have skewed the results. Second, the participants' awareness of being observed may have

XSL•FO
RenderX

influenced their performance, potentially biasing the results toward better patient safety practices. To mitigate this effect, all participants were allowed to withdraw from the study at any time and were assured that their individual results would remain confidential. Third, the TENTS instrument and scoring could be imperfect. We address this by conducting robust training and practice with the TENTS tool and by blinding students to the treatment status of the surgeons during postintervention observations. Despite the limitations associated with this approach, the involvement of medical student volunteers was essential for executing this extensive observational study without incurring financial costs. Future iterations of this research could leverage such a program, benefiting both the institution through enhanced understanding of behaviors in operating rooms and the students through valuable operating room exposure and hospital experience. In summary, while the findings offer valuable insights, the limitations of this single-site study with a small sample size and potential participant bias must be considered. Larger, more robust studies will be needed to validate and expand upon these preliminary results. There is also a need for longitudinal studies to assess effects over time.

Conclusions

A VR-based immersive training program focused on TeamSTEPPS principles appears effective at improving safety behaviors, as measured by the TENTS tool. Our qualitative analysis based on 492 quotes from participants, suggest that the observed behavioral changes are a result of an immersive experience and sense-making of key training concepts where participants could see the consequences of suboptimal teamwork (eg, poor leadership and communication, lack of attention to detail, failure to take responsibility, low psychological safety to speak up, etc). Given the persistent patient safety issues in operating rooms nationwide, such innovative and immersive patient safety education programs could provide a scalable intervention to help reduce patient harm in the long run. However, further research with larger, more diverse sample sizes is needed to confirm the generalizability of these findings.

Acknowledgments

The authors want to acknowledge the contribution of the following medical student volunteers in collecting the data used for this study: Kwadwo Ansah, Cameron Kurz, Julian Mobley, SL, Ricardo Crespo, Mary Kaufman, Kainat Aslam, Dakota Perez, Grace Fuller, Grace Tabor, Lily Bender, Alexander Requarth, Hannah Black, Brenderia Cameron, Annie Bright, Zhanè Washington, Sierra Parkinson, Collin Shick, and Mandi. We also want to send special thanks to the following medical students for their contribution to the qualitative data analysis: Anu Chaparala, Nathan Adams, and SL.

Conflicts of Interest

LM is an advisor and equity holder in Communify.us LLC and the founder of MaiaZura LLC. SB, JR, and CF are equity holders and co-owners of MaiaZura LLC.

Multimedia Appendix 1
One-minute overview clip of our VR training used in this study. VR: virtual reality.
[MP4 File, 8549 KB - mededu_v11i1e66186_app1.mp4]

References

1. Rodziewicz TL, Houseman B, Vaqar S, Hipskind JE. Medical error reduction and prevention. In: StatPearls: StatPearls Publishing; 2024. URL: https://www.ncbi.nlm.nih.gov/books/NBK499956/ [accessed 2024-02-12]
2. Panagioti M, Khan K, Keers RN, et al. Prevalence, severity, and nature of preventable patient harm across medical care settings: systematic review and meta-analysis. BMJ 2019 Jul 17;366:l4185. [doi: 10.1136/bmj.l4185] [Medline: 31315828]
3. Cima RR, Kollengode A, Garnatz J, Storsveen AS, Weisbrod CA, Deschamps C. Incidence and characteristics of potential and actual retained foreign object events in surgical patients. J Am Coll Surg 2008 Jul;207(1):80-87. [doi: 10.1016/j.jamcollsurg.2007.12.047] [Medline: 18589366]
4. Williams TL, Tung DK, Steelman VM, Chang PK, Szekendi MK. Retained surgical sponges: findings from incident reports and a cost-benefit analysis of radiofrequency technology. J Am Coll Surg 2014 Sep;219(3):354-364. [doi: 10.1016/j.jamcollsurg.2014.03.052] [Medline: 25081938]
5. Sentinel Event Alert 51: preventing unintended retained foreign objects. The Joint Commission. 2022 May 1. URL: https://tinyurl.com/3k8tn89r [accessed 2025-04-23]
6. Keebler JR, Dietz AS, Lazzara EH, et al. Validation of a teamwork perceptions measure to increase patient safety. BMJ Qual Saf 2014 Sep;23(9):718-726. [doi: 10.1136/bmjqs-2013-001942]
7. Costar DM, Hall KK. Improving team performance and patient safety on the job through team training and performance support tools: a systematic review. J Patient Saf 2020 Sep;16(3S Suppl 1):S48-S56. [doi: 10.1097/PTS.0000000000000746] [Medline: 32810001]
8. TeamSTEPPS 30. Agency for Healthcare Research and Quality. URL: https://www.ahrq.gov/teamstepps-program/index.html [accessed 2025-04-23]

XSL•FO
RenderX

9. Skaret MM, Weaver TD, Humes RJ, Carbone TV, Grasso IA, Kumar H. Automation of the I-PASS tool to improve transitions of care. J Healthc Qual 2019;41(5):274-280. [doi: 10.1097/JHQ.0000000000000174] [Medline: 31483392]
10. Lin WT, Mayer C, Lee BO. Validity and reliability of the teamwork evaluation of non-technical skills tool. Aust J Adv Nurs 2019;36(3):67-74 [FREE Full text] [doi: 10.37464/2019.363.1460]
11. Abelson JS, Silverman E, Banfelder J, Naides A, Costa R, Dakin G. Virtual operating room for team training in surgery. Am J Surg 2015 Sep;210(3):585-590. [doi: 10.1016/j.amjsurg.2015.01.024] [Medline: 26054660]
12. Gasteiger N, van der Veer SN, Wilson P, Dowding D. How, for whom, and in which contexts or conditions augmented and virtual reality training works in upskilling health care workers: realist synthesis. JMIR Serious Games 2022 Feb 14;10(1):e31644. [doi: 10.2196/31644] [Medline: 35156931]
13. Liaw SY, Ooi SW, Rusli KDB, Lau TC, Tam WWS, Chua WL. Nurse-physician communication team training in virtual reality versus live simulations: randomized controlled trial on team communication and teamwork attitudes. J Med Internet Res 2020 Apr 8;22(4):e17279. [doi: 10.2196/17279] [Medline: 32267235]
14. Zhao J, Xu X, Jiang H, Ding Y. The effectiveness of virtual reality-based technology on anatomy teaching: a meta-analysis of randomized controlled studies. BMC Med Educ 2020 Apr 25;20(1):127. [doi: 10.1186/s12909-020-1994-z] [Medline: 32334594]
15. Kyaw BM, Saxena N, Posadzki P, et al. Virtual reality for health professions education: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959. [doi: 10.2196/12959] [Medline: 30668519]
16. Wang P, Wu P, Wang J, Chi HL, Wang X. A critical review of the use of virtual reality in construction engineering education and training. Int J Environ Res Public Health 2018 Jun 8;15(6):1204. [doi: 10.3390/ijerph15061204] [Medline: 29890627]
17. Makransky G, Klingenberg S. Virtual reality enhances safety training in the maritime industry: an organizational training experiment with a non‐WEIRD sample. Computer Assisted Learning 2022 Aug;38(4):1127-1140 [FREE Full text] [doi: 10.1111/jcal.12670]
18. Mazur LM, Khasawneh A, Fenison C, et al. A novel theory-based virtual reality training to improve patient safety culture in the department of surgery of a large academic medical center: protocol for a mixed methods study. JMIR Res Protoc 2022 Aug 24;11(8):e40445. [doi: 10.2196/40445] [Medline: 36001370]
19. Reason J. Human error: models and management. BMJ 2000 Mar 18;320(7237):768-770. [doi: 10.1136/bmj.320.7237.768] [Medline: 10720363]
20. Wiegmann DA, Shappell SA. A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System, 1st edition: Routledge; 2017:45-71. [doi: 10.4324/9781315263878]
21. Alase A. The Interpretative Phenomenological Analysis (IPA): a guide to a good qualitative research approach. IJELS 2017;5(2):9. [doi: 10.7575/aiac.ijels.v.5n.2p.9]
22. Klein G, Phillips JK, Rall EL, Peluso DA. A data-frame theory of sensemaking. In: Hoffman RR, editor. Expertise Out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making: Routledge; 2007:45-71. [doi: 10.4324/9780203810088]
23. McEwan D, Ruissen GR, Eys MA, Zumbo BD, Beauchamp MR. The effectiveness of teamwork training on teamwork behaviors and team performance: a systematic review and meta-analysis of controlled interventions. PLoS ONE 2017;12(1):e0169604. [doi: 10.1371/journal.pone.0169604] [Medline: 28085922]
24. Todsen T, Melchiors J, Wennerwaldt K. Use of virtual reality to teach teamwork and patient safety in surgical education. 2018 Presented at: 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR); Mar 18-22, 2018; Reutlingen, Germany. [doi: 10.1109/VR.2018.8446469]
25. Aaberg OR, Hall-Lord ML, Husebø SIE, Ballangrud R. A human factors intervention in a hospital - evaluating the outcome of a TeamSTEPPS program in a surgical ward. BMC Health Serv Res 2021 Feb 3;21(1):114. [doi: 10.1186/s12913-021-06071-6] [Medline: 33536014]
26. Shi R, Marin-Nevarez P, Hasty B, et al. Operating room in situ interprofessional simulation for improving communication and teamwork. J Surg Res 2021 Apr;260:237-244. [doi: 10.1016/j.jss.2020.11.051] [Medline: 33360307]
27. Kaplan HJ, Spiera ZC, Feldman DL, et al. Risk reduction strategy to decrease incidence of retained surgical items. J Am Coll Surg 2022 Sep 1;235(3):494-499. [doi: 10.1097/XCS.0000000000000264] [Medline: 35972170]
28. Khoshnoodifar M, Emadi N, Mosalanejad L, Maghsoodzadeh S, Shokrpour N. A new practical approach using TeamSTEPPS strategies and tools: - an educational design. BMC Med Educ 2024 Jan 4;24(1):22. [doi: 10.1186/s12909-023-04803-2] [Medline: 38178071]

Abbreviations

FDR: false discovery rate
HFACS: Human Factors Analysis and Classification System
TeamSTEPPS: Team Strategies and Tools to Enhance Performance and Patient Safety
TENTS: Teamwork Evaluation of Non-Technical Skills
VR: virtual reality

XSL•FO
RenderX

Edited by B Lesselroth; submitted 05.09.24; peer-reviewed by KP Wong, P Greilich; revised version received 04.03.25; accepted 06.04.25; published 01.05.25.

Please cite as:
Mazur L, Butler L, Mitchell C, Lashani S, Buchanan S, Fenison C, Adapa K, Tan X, An S, Ra J
Effect of Immersive Virtual Reality Teamwork Training on Safety Behaviors During Surgical Cases: Nonrandomized Intervention Versus Controlled Pilot Study
JMIR Med Educ 2025;11:e66186
URL: https://mededu.jmir.org/2025/1/e66186
doi:10.2196/66186

© Lukasz Mazur, Logan Butler, Cody Mitchell, Shaian Lashani, Shawna Buchanan, Christi Fenison, Karthik Adapa, Xianming Tan, Selina An, Jin Ra. Originally published in JMIR Medical Education (https://mededu.jmir.org), 1.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of Clinical Decision Support Systems on Medical Students' Case-Solving Performance: Comparison Study with a Focus Group

Marco Montagna[1*], MD; Filippo Chiabrando[1*], MD; Rebecca De Lorenzo[1], MD; Patrizia Rovere Querini[1,2], MD, PhD; Medical Students[3]

[1]School of Medicine, Vita-Salute San Raffaele University, Via Olgettina 58, Milan, Italy

[2]Unit of Medical Specialties and Healthcare Continuity, IRCCS San Raffaele Scientific Institute, Milan, Italy

[3]

[*]these authors contributed equally

Corresponding Author:
Marco Montagna, MD
School of Medicine, Vita-Salute San Raffaele University, Via Olgettina 58, Milan, Italy

Abstract

Background: Health care practitioners use clinical decision support systems (CDSS) as an aid in the crucial task of clinical reasoning and decision-making. Traditional CDSS are online repositories (ORs) and clinical practice guidelines (CPG). Recently, large language models (LLMs) such as ChatGPT have emerged as potential alternatives. They have proven to be powerful, innovative tools, yet they are not devoid of worrisome risks.

Objective: This study aims to explore how medical students perform in an evaluated clinical case through the use of different CDSS tools.

Methods: The authors randomly divided medical students into 3 groups, CPG, n=6 (38%); OR, n=5 (31%); and ChatGPT, n=5 (31%); and assigned each group a different type of CDSS for guidance in answering prespecified questions, assessing how students' speed and ability at resolving the same clinical case varied accordingly. External reviewers evaluated all answers based on accuracy and completeness metrics (score: 1 - 5). The authors analyzed and categorized group scores according to the skill investigated: differential diagnosis, diagnostic workup, and clinical decision-making.

Results: Answering time showed a trend for the ChatGPT group to be the fastest. The mean scores for completeness were as follows: CPG 4.0, OR 3.7, and ChatGPT 3.8 (P=.49). The mean scores for accuracy were as follows: CPG 4.0, OR 3.3, and ChatGPT 3.7 (P=.02). Aggregating scores according to the 3 students' skill domains, trends in differences among the groups emerge more clearly, with the CPG group that performed best in nearly all domains and maintained almost perfect alignment between its completeness and accuracy.

Conclusions: This hands-on session provided valuable insights into the potential perks and associated pitfalls of LLMs in medical education and practice. It suggested the critical need to include teachings in medical degree courses on how to properly take advantage of LLMs, as the potential for misuse is evident and real.

(JMIR Med Educ 2025;11:e55709) doi:10.2196/55709

KEYWORDS

chatGPT; chatbot; machine learning; ML; artificial intelligence; AI; algorithm; predictive model; predictive analytics; predictive system; practical model; deep learning; large language models; LLMs; medical education; medical teaching; teaching environment; clinical decision support systems; CDSS; decision support; decision support tool; clinical decision-making; innovative teaching

Introduction

Clinical reasoning and decision-making are at the core of the medical workflow. If they are accurate and grounded on solid and updated evidence, they help ensure the best health outcomes for patients. Clinical decision support systems (CDSS) have been implemented to aid practitioners in this duty [1-3]. Clinical practice guidelines (CPG) serve as the prototype for CDSS. They are published and updated at varying frequencies by scientific societies and policy makers, covering virtually every medical field or disorder. Over time, the number and complexity of CPG have increased, resulting in more detailed and robust recommendations. However, this has also led to reduced immediacy and ease of access and comprehension for medical professionals. Additionally, there may be multiple CPG for a single pathological condition, sometimes with conflicting recommendations. As a potential solution, emerging technologies on the internet have given rise to new CDSS options known as online repositories (ORs). These repositories, like encyclopedias, consolidate and synthesize knowledge

XSL•FO
RenderX

related to various medical disorders. They draw from current practices, available CPG, and the latest published evidence, making this information easily accessible to physicians. Typically provided by publishing groups, ORs often require subscription-based access. Two of the most popular options are "UpToDate" (by Wolters Kluwer [4]) and "BMJ Best Practice" (by BMJ Publishing Group [5]), both available as websites and mobile apps. The recent introduction of large language models (LLMs) for public use has generated both excitement and debate. Their adoption has rapidly grown across various human activities [6]. Many foresee the immense potential benefits of applying such technology to medical practice, while others harbor concerns about the dangers it might pose if left unregulated and misaligned [7-12].

Without a doubt, LLMs like ChatGPT represent a new generation of CDSS with unparalleled assistance capabilities. They can engage in active interactions with users and directly interpret medical information, extending far beyond simple guideline consultation. They can suggest possible diagnostic workups (DWs) or treatment algorithms [6]. In such cases, physicians would no longer need to navigate extensive datasets of clinical information, distill practical advice from lengthy text pages, or grapple with uncertainty about consulting the correct or sufficient sources. On the flip side, it is evident that LLMs also carry the potential for misuse, which could lead to significant harm to patients [7-9,11,13]. There's a risk of guiding clinicians down erroneous thought processes, potentially resulting in wasted time and the unintentional complication of cases. When the alternatives being evaluated are either incorrect or become excessively numerous, the complexity of a case may inevitably worsen. As a result, there is legitimate concern regarding how the indiscriminate use of LLMs might inadvertently drive-up health care costs. This underscores the importance of integrating LLMs into clinical decision-making (CDM) processes with caution and judiciousness.

However, although the adoption of innovative CDSS tools is steadily rising, the lack of dedicated training in their proper utilization undermines their full potential as valuable aids [9,10].

The aim of the present study was to investigate how senior medical students employ CDSS in the resolution of a clinical case with the ultimate intention of designing specific educational programs. Specifically, we conducted a hands-on session to compare CPG, ORs, and an LLM (ChatGPT) in terms of speed and accuracy of the clinical decisions proposed after consultancy with the CDSS.

Methods

Study Design

The present is a report of a hands-on practical session taking place during the Course of Internal Medicine at our university. The subject of the analysis was the quality of students' answers to a number of open-ended questions related to clinical reasoning and problem-solving, as a proxy for their capacity to employ different CDSS. A fictional clinical case was designed by the authors to control for complexity. Additionally, ChatGPT (version 3.5) generative capabilities were used to fabricate vital parameters, physical examination, and laboratory results for the fictional patient. ChatGPT was asked to include confounding factors in the answers provided. The authors revised generated elements to make sure they met the study requirements. The complete clinical case, open-ended questions and conversation with ChatGPT are available in Multimedia Appendix 1 and Multimedia Appendix 2.

Participants, Recruitment Strategies, and Sampling Method

Students attending the last lesson in the academic year 2022/23 Course of Internal Medicine of the International MD Program degree at Vita-Salute San Raffaele University, Milan (IT), were all included in the study, with convenience sampling. No exclusion criteria were applied.

Experimental Groups, Randomization, and Blinding

In total, 3 groups with comparable numbers of students were defined at the beginning of the lesson (Multimedia Appendix 3). Starting from the first seating rows, students were randomly assigned a number from 1 to 3 and consequently formed the 3 groups. Each group nominated a delegate who randomly picked an envelope containing the indication of the type of CDSS to be used by his or her group, either (i) CPG, (ii) OR, or (iii) ChatGPT. For each group, only the delegate was allowed web-based access to the CDSS. Group assignments were open label. The group assigned to CPG was allowed to use the internet to search for and consult CPG deemed useful to solve the case. The group assigned to OR was allowed to use the internet to access UpToDate and look for articles and algorithms or tables deemed useful to solve the case. The group assigned to ChatGPT was allowed to log into the LLM and use it to ask and gather information deemed useful to solve the case.

The delegate was also in charge of sending his or her group's clinical decision to the researchers via a mobile phone SMS text message. An inspector of the research staff was assigned to each group to guarantee that only the assigned CDSS was used. The questions were shown along with the presentation of the clinical case in a Microsoft PowerPoint (Microsoft Corporation) slideshow. For each question, a countdown timer was shown on the projector screen, and the start time was recorded by the researchers. Time required by each group for each answer was calculated by subtracting the start time from the mobile message arrival time.

Outcomes/Assessment

In total, 1 junior resident in internal medicine, 1 senior resident in internal medicine, and 2 internal medicine junior consultants were asked to perform blind external assessment of the answers. They were provided with a form containing the same clinical case and questions shown to the students together with the answers given by each group, with no details on CDSS used. Answers were graded from 1 to 5 in terms, respectively, of completeness and accuracy. Completeness was described as: "Is the answer complete or does it miss anything?". Accuracy was described as: "Is the answer precise and adherent to clinical practice, or too vague, too wide, too superficial?"

XSL•FO
RenderX

Sample Size

Sample size was not defined a priori. Sample size was determined by the number of students that attended the lesson on that day.

Statistical Analysis

Scores are reported as mean. For further analysis, the 8 questions were grouped into 3 domains according to the students' skill investigated: (1) differential diagnosis (DD, Q 1-5-6), (2) DW (Q 2 - 7), and (3) CDM (Q 3-4-8). The Kruskal-Wallis test was performed to establish whether there was a significant difference among the 3 groups in the times and scores overall and in each student skill domain. Microsoft Excel (Microsoft Corporation) and GraphPad Prism (version 9.0; GraphPad Software) were used as software tools for the analysis.

Ethical Considerations

This study does not require ethical approval as this is a report of data collected for monitoring and reporting purposes of innovative teaching activities taking place at Vita-Salute San Raffaele University, according to the Self-assessment-Evaluation-Accreditation system (AVA) of ANVUR, to which our institution is subject [14]. The data were generated during a hands-on session taking place during a usual lesson and in a teaching, non-experimental environment, with no risks for the participants. Data represent the output of each student group; they therefore collect aggregated information, with no individual identity linked to them. No confidential data were collected. No compensation was provided to participants, and they were able to opt out at any time during the lesson. AVA aims to improve the quality of teaching and research carried out in the Italian universities through the application of a quality assurance model based on internal procedures for planning, management, self-assessment, and improvement of training and scientific activities and on an external verification carried out in a clear and transparent manner. The requirements of the new AVA3 model underline the importance for the universities to promote, support, and monitor the participation of teachers in training and teaching refresher initiatives in the various disciplines, including those relating to the use of innovative teaching methodologies, also through the use of online tools and the provision of multimedia teaching materials [14]. The presented data were collected in this context, and, accordingly, no ethics approval was applied for (Page 3, Art.5, Clause 2 of [15]).

Results

A total of 16 students were included: 6 allocated to the CPG group (*F*=5, 83%), 5 to the OR group (*F*=2, 40%), and 5 to the ChatGPT group (*F*=3, 60%).

During the presentation of the clinical case, all 3 groups were presented with questions, and students were required to provide their responses as quickly as possible, within predefined time limits. The time taken to answer each question was recorded for all groups. Except for one response, all answers were given within the allocated time (see Table 1). Of the 49 total allocated minutes, the CPG group took 41 minutes to complete the clinical case, the OR group 45 minutes, and the ChatGPT group 38 minutes. The total time taken to answer, expressed as a percentage of the allocated time, was not significantly different among groups (*P*=.69).

XSL•FO
RenderX

Table . Time (min) required for answers and mean score received at the external assessment in terms of completeness and accuracy for each answer given by each group. Overall time for completion and median score across all answers for each group are also reported. Allotted time was exceeded only in Q4 by OR group.

	Clinical practice guidelines			Online repositories			ChatGPT		
	Time (min)	Quality		Time (min)	Quality		Time (min)	Quality	
		Completeness, mean (SD)	Accuracy, mean (SD)		Completeness, mean (SD)	Accuracy, mean (SD)		Completeness, mean (SD)	Accuracy, mean (SD)
Q1. Rank the possible differential diagnoses in terms of probability. 8 min[a]	8	4.0 (0.8)	3.8 (1.3)	8	3.5 (0.6)	3.0 (0.8)	8	4.0 (0.8)	3.8 (0.5)
Q2. Based on the previous list, which diagnostic workup would you set up? 8 min[b]	7	4.0 (0.8)	3.8 (0.5)	7	4.8 (0.8)	2.5 (1.0)	6	3.8 (0.5)	3.8 (1.3)
Q3. Which values are altered? 5 min[c]	3	4.0 (0.8)	4.3 (1.0)	5	3.8 (1.0)	4.0 (1.2)	2	4.0 (0.8)	2.3(1.3)
Q4. Which treatment do you start? 5 min[c]	5	4.3 (1.5)	3.8 (1.5)	6	4.0 (0.8)	4.3 (1.0)	4	4.0 (0.8)	4.3 (1.0)
Q5. Which are the possible causes of hypercalcemia? 8 min[a]	5	3.8 (1.0)	3.8 (1.0)	4	3.8 (0.5)	3.8 (1.0)	6	3.0 (0)	3.3 (1.3)
Q6. Can you narrow down the previous list based on these findings? 5 min[a]	3	4.0 (1.4)	4.3 (1.5)	5	3.0 (0.8)	2.0 (0.8)	5	4.3 (1.0)	4.0 (0.8)
Q7. Which are the primary diagnostic tests that you order? 5 min[b]	5	4.3 (1.0)	4.5 (1.0)	5	3.8 (1.5)	3.4 (0.7)	2	4.0 (0.8)	4.3 (0.9)
Q8. Which therapeutic choice do you offer to the patient? 5 min[c]	5	3.8 (0.5)	4.3 (1.0)	5	3.5 (1.3)	3.4 (1.3)	5	3.8 (1.0)	4.0 (1.1)

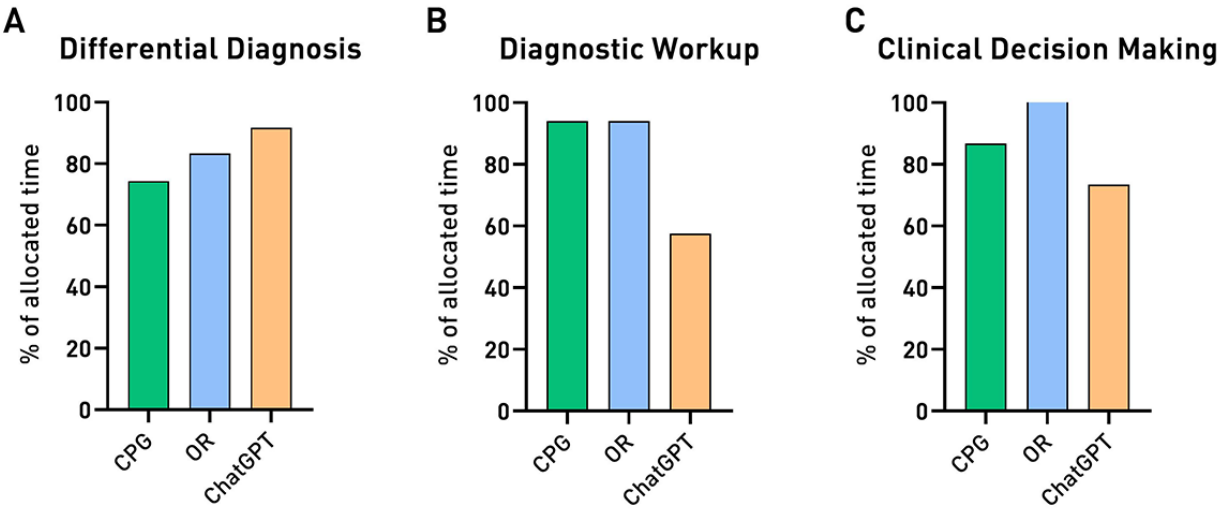
XSL•FO
RenderX

	Clinical practice guidelines			Online repositories			ChatGPT		
	Time (min)	Quality		Time (min)	Quality		Time (min)	Quality	
		Completeness, mean (SD)	Accuracy, mean (SD)		Completeness, mean (SD)	Accuracy, mean (SD)		Completeness, mean (SD)	Accuracy, mean (SD)
TOT/mean[a]	41	4.0 (0.2)	4.0[d] (0.3)	45	3.8 (0.5)	3.3[d] (0.8)	38	3.8 (0.4)	3.7[d] (0.7)

[a]Differential diagnosis
[b]Diagnostic workup
[c]Clinical decision-making
[d]P=.02.

The questions were then categorized into 3 major domains: DD (Q 1-5-6), DW (Q 2 - 7), and CDM (Q 3-4-8). The time taken to answer, as a percentage of the allocated time, by each group of students according to the provided domains is shown in Figure 1(A, B and C). While no statistically significant differences were observed, it is worth noting that the ChatGPT group tended to respond more quickly to questions related to DW and CDM.

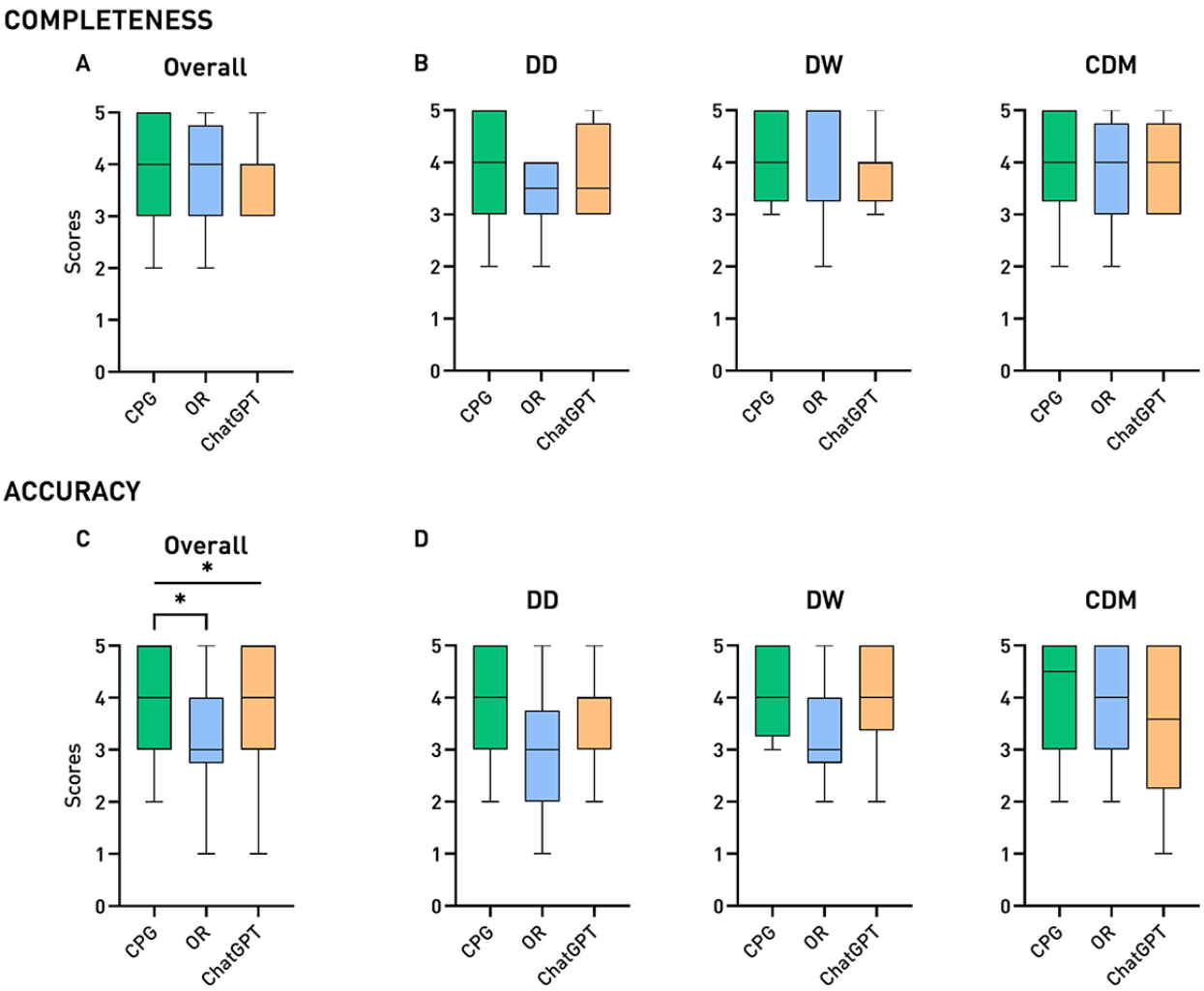
Figure 1. Sum of the time taken by the 3 groups of students to answer questions in the 3 domains. Results are shown as the percentage of the total allocated time for that domain. CPG: clinical practice guidelines; OR: online repositories.



Answers were assessed for completeness, which considered the depth of information provided, and for accuracy, which assessed adherence to clinical practice versus an excess or superficiality of information. These evaluations were conducted by 4 external reviewers who were blinded to the group assignment. Scores are reported in Table 1 and Figure 2.

XSL•FO
RenderX

Figure 2. Box plots of scores obtained by the 3 groups of students for (A)overall completeness, (C)overall accuracy, (B)completeness in the 3 domains, (D)accuracy in the 3 domains. DD: differential diagnosis; DW: diagnostic workup; CDM: clinical decision-making (*P<.05).



Overall, the CPG group performed best, reaching the highest mean scores for completeness (4.0) and accuracy (4.0) (Figure 2A and C). The ChatGPT group comes in second place, with equal completeness (3.8 vs 3.8) compared to the OR group but higher accuracy (3.7 vs 3.3). The Kruskal-Wallis test showed non-significant difference among groups for overall completeness (P=.49), and a significant difference in overall accuracy between the 3 groups (P=.02), particularly—at post hoc analysis—between the CPG and the OR group (P=.02).

Aggregating scores according to the 3 students' skill domains, trends in differences among the groups emerge more clearly (Figure 2B and C). When it comes to generating differential diagnoses, the CPG group was the most complete (3.9) and accurate (3.9) among the 3, whereas the OR group has the worst scores for both completeness and accuracy categories, with mean scores of 3.4 and 2.9, respectively, among the lowest registered. On the contrary, whenever students were asked to provide a DW for the patient, the OR group appeared as the most complete (4.3), although the least accurate (3.0), considering the source of the competition. In this domain, the other 2 groups come quite close (4.1 and 4.1 for CPG; 3.9 and 4.0 for ChatGPT) and maintain coherent scores in between their own completeness and accuracy. Lastly, when knowledge had to be applied to drive clinical decisions, as tested in CDM questions, the 3 groups obtained similar scores in terms of completeness, while the CPG group succeeded as the most accurate (4.1), followed by the OR group (3.9) and the ChatGPT group (3.5).

Of note, it appears evident how the CPG group performed the best in nearly all domains and maintained almost perfect alignment between its completeness and accuracy. The ChatGPT group maintained an overall mediocre yet stable performance, without ever achieving the best scores. On the other hand, the OR group showed mixed features, with notable peaks of performance–with scores equal to or higher than the ChatGPT contender–but some other dramatic drops in answer quality in terms of accuracy in more than one skill domain.

Discussion

In recent years, the availability of tools to support clinicians in the diagnostic and therapeutic processes has grown considerably. Although the use of CDSS is widespread, individuals often use them without specific education and pay little attention to their

XSL•FO
RenderX

inherent limitations, especially in the case of their newest potential counterparts, such as ChatGPT [1].

To assess how final-year medical school students make use of the available CDSS and to begin considering an instructional approach for the use of such tools, we designed the experiment outlined in this study. Observing the students' interaction with the assigned CDSS during the resolution of a clinical case and analyzing their answers, we recorded specific criticalities regarding the use of each CDSS.

In terms of rapidity of use, ChatGPT seems to represent a significant breakthrough in the world of CDSS. If traditional encyclopedic or textbook-like written resources call the reader to go through the entire material to properly understand a topic and retrieve correct clinical answers or guidance, an instantaneous chat environment allows for both a quick overview of disciplines and—at the same time—deep vertical dives into specific details. Students using ChatGPT arrived at the required answer almost always faster than their colleagues aided by CPG or OR, especially in questions regarding DW and CDM. It seems that the role of ChatGPT in the chat dialogue more closely resembles the attitude of a human counterpart (for example, a senior teaching physician asked for guidance), whose responses are fast, direct, and usually finely targeted. Such answers follow the students' line of reasoning and indirectly encourage them to choose a unique path of solution out of many possible scenarios. This dynamic emerges as brilliantly effective whenever the students embark on the correct clinical thought process but can lead to disastrous consequences whenever students feed cognitive biases or overt errors into their chat conversation. On the contrary, CPG and OR offer vast amounts of information, such as long lists of items and in-depth descriptions, whose digestion is neither easy nor fast. Therefore, whenever consulted correctly and for enough time, CPG and OR—especially the former, our results seem to suggest—generally help students give answers of higher quality, both in terms of completeness and accuracy. No real-time interaction is present; therefore, they appear almost immune to reader-introduced bias and misinterpretation.

In the modern era, velocity is a precious commodity, especially in the fast-paced clinical context, where less and less time is available for extensive reference consultation. This might influence current and future generations of medical school students to prefer chatbot-based guidance over preformed texts as routine help throughout their study [1,8].

As said, in terms of highest answer quality, there seems to be no rival to CPG. Old-fashioned guidelines might be slower to consult but grant far greater quality information to students, helping them to be complete and accurate in key tasks, such as generating differentials and deploying clinical decisions. Possibly, CPG might also enhance the student's comprehension of the analyzed topic, given the broader context and deeper description always provided. Nonetheless, in a continuously evolving context of increasing number and complexity of CPG, it would be presumptuous to expect students to rely on their use only [4,5,9].

ORs offered unexpected and ambiguous results, as the performance of these students was not stable across questions, and their answers could reach both extremes of the score spectrum. ORs are intrinsically designed and thought of as an evolution of CPG, more accessible and applicable to practice. This aspect may emerge in the excellent quality of answers given by this group in the DW domain, where students were able to follow the detailed workup algorithms offered by OR, which can graphically synthesize even complex clinical scenarios. An interesting experience across teaching hospitals in Japan evidenced a somewhat significant positive correlation between the use of OR and score performance in a national general medicine test, in a numerous population of over 3000 residents. According to the authors, frequent logging into and consultation of UpToDate might have contributed to improved clinical reasoning skills, specifically in the tested domains of DD generation, and clinical decision deployment [16].

Specific Observations Around the Students' Use of Each Type of CDSS

CPG Group

For this group's ability to reach a correct solution, the crucial step seemed to be selecting the proper guideline to consult. Once correctly identified, suitable CPG contain virtually everything a physician should know about a disease. During the initial process of elaboration of a clinical scenario, though, it is not immediately clear which disease entity, often more than one, is going to be selected as a candidate diagnosis for the patient. Choosing the right CPG can therefore be quite challenging. Additionally, CPG deliver their content in the form of plain text and interspersed summary tables and charts. The students had some difficulty in focusing on the right chart.

OR Group

Likely due to their lack of experience with the tool, the students failed to search for symptoms within the query. One hypothesis could be that they relied on their prior knowledge to make decisions rather than on the results obtained from each query, preventing them from breaking free from their preconceptions.

ChatGPT Group

Our results revealed students to lack substantial background on how to properly approach an LLM chatbot. *Masterprompting*, referred to as the assignment of the role and behavior expected from the chatbot, was not provided by the students.

Using an LLM as a CDSS inevitably introduces a "prompt bias," for which the human subjective way of reasoning and choice of questions to be asked directly influence how the chatbot perceives the information and how it transforms its responses accordingly. Along such a general trend, the students' prompts were not properly designed and translated into confused and misleading answers by ChatGPT. For example, clinical details were provided without any structure or further clinical context, triggering diverging suggestions by the chatbot. Accordingly, hints were given by the instructor on asking the LLM directly how to convey information for it to understand and elaborate on such information at its best, but consistent results did not follow.

On other instances, it was ChatGPT itself that derailed students' reasoning. For example, a chatbot answer about laboratory

XSL•FO
RenderX

values, lacking a relative unit of measurement and normal reference ranges, leads to misinterpretation of hypercalcemia as hypocalcemia, a critical mistake.

ChatGPT does not provide any literature citation and guideline reference to support its own line of reasoning, as other CDSS do in the form of easily accessible links to further explanations and deeper dives (eg differential diagnoses, varying DWs, tables of available first-line therapies, etc) [6]. Such an absence seemed to be associated with a significantly lower propensity of students to question either their own knowledge or the answers formulated by ChatGPT, blindly accepting the information provided. Whenever such indulgence was noted and pointed out by the instructors, students confessed how certain answers provided by ChatGPT were not completely clear and understandable; nonetheless, they willingly accepted them as valid.

Limitations

Some limitations of this report must be underlined. First, the restricted number of students who took part in the experiment. This may have led to uneven distribution of differently ranked students in the groups, despite the random strategy used for group definition. Second, our methodological constraints lead to insufficient statistical power to draw sound conclusions. Lastly, the experiment has not yet been repeated, and the results have been further confirmed or discarded with other clinical cases.

Conclusions

As in many other disciplines, the adoption of LLMs in medical practice and in the medical school curriculum is inevitable [10]. Our hands-on session suggests the critical need to include in medical degree courses teachings on how to properly take advantage of LLMs such as ChatGPT, as we verified that the potential for misuse is evident and real.

Our experience suggests the need for medical students to be acquainted with LLMs in their learning process and future profession. ChatGPT does not provide nor teach a reasoning method to approach a medical case resolution, a relevant issue for it to be recognized as part of the armamentarium in formal medical education. CPG and OR, on the contrary, most often provide step-by-step guidance on how to behave in each clinical scenario, how to approach diagnosis, and how to address treatment of diseases. References and recommendation strength form the cornerstone of these tools and help the student get progressively acquainted with ever-updating medical knowledge [4,5].

In conclusion, regarding the upcoming future, we suggest medical educators to:

- start to increasingly incorporate and refer to LLMs in their teachings, also by building tailored case studies [17-20];
- favor practice-based learning by using LLMs as a help to navigate guidelines and repositories with more ease and speed;
- exploit the very limitations of LLMs—such as the lack of an explicit reasoning method or unsure reliance on the latest published literature—to prompt students to consciously provide them themselves to the chatbot, turning the CDSS consultation process into a bidirectional teaching environment, possibly uncovering biases and misconceptions on both sides;
- help students in focusing on their accountability: they should be pushed to continuously look for evidence and validation of their own clinical reasoning, avoiding relying completely on that of LLMs.

Acknowledgments

The authors thank the Medical Students as Collaborators attending the hands-on session: ChatGPT group - Mariam Datukishvili, Roberto Leone, Mariagiulia Giugliano, Marifrancesca Forquet, Ino De Martino; UpToDate group - Lisa Marie Pereira, Pietro Felisatti, Francesco Paleari, Carlotta Dell'Anna Misurale, Angelo Manfredi; Guidelines Group - Giulia Pacini, Francesca Crippa, Ivan Shashkin, Carola Alberoni, Mazvita Mungwadzi, Sara Coacci. The authors thank Prof. Angelo Andrea Maria Manfredi for his valuable insights on the matter.

Authors' Contributions

FC and MM involved in conceptualization, methodology, investigation, data curation, formal analysis, and writing: original draft; RDL contributed to methodology, investigation, data curation, and writing: review & editing; MD students contributed to investigation and resources; PRQ involved in conceptualization, methodology, supervision, and writing: review & editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
ChatGPT conversation history for clinical case generation.
[DOCX File, 35 KB - mededu_v11i1e55709_app1.docx]

Multimedia Appendix 2
Clinical Case and questions for students.
[PDF File, 56 KB - mededu_v11i1e55709_app2.pdf]

XSL•FO
RenderX

Multimedia Appendix 3
Workflow Diagram.
[PDF File, 20 KB - mededu_v11i1e55709_app3.pdf]

References

1. Goodman KE, Rodman AM, Morgan DJ. Preparing physicians for the clinical algorithm era. N Engl J Med 2023 Aug 10;389(6):483-487. [doi: 10.1056/NEJMp2304839] [Medline: 37548320]
2. Huang S, Liang Y, Li J, Li X. Applications of clinical decision support systems in diabetes care: scoping review. J Med Internet Res 2023 Dec 8;25(1):e51024. [doi: 10.2196/51024] [Medline: 38064249]
3. Wiwatkunupakarn N, Aramrat C, Pliannuom S, et al. The integration of clinical decision support systems into telemedicine for patients with multimorbidity in primary care settings: scoping review. J Med Internet Res 2023 Jun 28;25(1):e45944. [doi: 10.2196/45944] [Medline: 37379066]
4. Evidence-Based Clinical Decision Support System| UpToDate. : Wolters Kluwer URL: https://www.wolterskluwer.com/en/solutions/uptodate [accessed 2023-07-11]
5. BMJ best practice. URL: https://bestpractice.bmj.com/ [accessed 2023-07-11]
6. Introducing ChatGPT. URL: https://openai.com/blog/chatgpt [accessed 2023-07-19]
7. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med 2023 Jul 6;6(1):120. [doi: 10.1038/s41746-023-00873-0] [Medline: 37414860]
8. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst 2023 Mar 4;47(1):33. [doi: 10.1007/s10916-023-01925-4] [Medline: 36869927]
9. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell 2023;6:1169595. [doi: 10.3389/frai.2023.1169595] [Medline: 37215063]
10. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023 Mar 19;11(6):887. [doi: 10.3390/healthcare11060887] [Medline: 36981544]
11. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med 2023 Aug;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
12. Eysenbach G. The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. JMIR Med Educ 2023 Mar 6;9:e46885. [doi: 10.2196/46885] [Medline: 36863937]
13. The Lancet Digital Health. ChatGPT: friend or foe? Lancet Digit Health 2023 Mar;5(3):e102. [doi: 10.1016/S2589-7500(23)00023-7] [Medline: 36754723]
14. II cycle (AVA3) – ANVUR – agenzia nazionale di valutazione del sistema universitario e della ricerca. URL: https://www.anvur.it/en/activities/ava/periodic-accreditation/ii-cycle-ava3/ [accessed 2024-10-01]
15. President of Italian Republic. Presidential decree february 1st, 2010, n. 76. rome. 2010 URL: https://www.normattiva.it/uri-res/N2Ls?urn:nir:presidente.repubblica:decreto:2010-02-01;76~art12-com4-letd [accessed 2024-10-25]
16. Kataoka K, Nishizaki Y, Shimizu T, et al. Hospital se of a web-based clinical knowledge support system and in-training examination performance among postgraduate resident physicians in Japan: nationwide observational study. JMIR Med Educ 2024 May 30;10(1):e52207. [doi: 10.2196/52207]
17. Lee J, Wu AS, Li D, Kulasegaram KM. Artificial intelligence in undergraduate medical education: a scoping review. Acad Med 2021 Nov 1;96(11S):S62-S70. [doi: 10.1097/ACM.0000000000004291] [Medline: 34348374]
18. Tsopra R, Peiffer-Smadja N, Charlier C, et al. Putting undergraduate medical students in AI-CDSS designers' shoes: an innovative teaching method to develop digital health critical thinking. Int J Med Inform 2023 Mar;171:104980. [doi: 10.1016/j.ijmedinf.2022.104980]
19. Feng S, Shen Y. ChatGPT and the future of medical education. Acad Med 2023 Aug 1;98(8):867-868. [doi: 10.1097/ACM.0000000000005242] [Medline: 37162219]
20. Jacobs SM, Lundy NN, Issenberg SB, Chandran L. Reimagining core entrustable professional activities for undergraduate medical education in the era of artificial intelligence. JMIR Med Educ 2023 Dec 19;9(1):e50903. [doi: 10.2196/50903] [Medline: 38052721]

Abbreviations

CDM: clinical decision-making
CDSS: clinical decision support systems
CPG: clinical practice guidelines
DD: differential diagnosis
DW: diagnostic workup
LLM: large language model
OR: online repositories

XSL•FO
RenderX

Edited by TDA Cardoso, T Leung; submitted 21.12.23; peer-reviewed by E Ogut, JJ Ríos Blanco; revised version received 25.10.24; accepted 26.10.24; published 18.03.25.

Please cite as:
Montagna M, Chiabrando F, De Lorenzo R, Rovere Querini P, Medical Students
Impact of Clinical Decision Support Systems on Medical Students' Case-Solving Performance: Comparison Study with a Focus Group
JMIR Med Educ 2025;11:e55709
URL: https://mededu.jmir.org/2025/1/e55709
doi:10.2196/55709

© Marco Montagna, Filippo Chiabrando, Rebecca De Lorenzo, Patrizia Rovere Querini, Medical Students. Originally published in JMIR Medical Education (https://mededu.jmir.org), 18.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Transforming Medical Education to Make Patient Safety Part of the Genome of a Modern Health Care Worker

Peter Lachman[1*], BA, MBBCH, MMed, MPH, MD; John Fitzsimons[1,2*], BMedSci, MBBCh, MSc

[1]Quality Improvement Department, Royal College of Physicians of Ireland, 19 South Frederick Street, Dublin, Ireland

[2]Children's Health Ireland at Temple Street, Dublin, Ireland

[*]all authors contributed equally

Corresponding Author:
Peter Lachman, BA, MBBCH, MMed, MPH, MD
Quality Improvement Department, Royal College of Physicians of Ireland, 19 South Frederick Street, Dublin, Ireland

Related Article:

https://mededu.jmir.org/2024/1/e64125

Abstract

Medical education has not traditionally recognized patient safety as a core subject. To foster a culture of patient safety and enhance psychological safety, it is essential to address the barriers and facilitators that currently impact the development and delivery of medical education curricula. The aim of including patient safety and psychological safety competencies in education curricula is to insert these into the genome of the modern health care worker.

(JMIR Med Educ 2025;11:e68046) doi:10.2196/68046

KEYWORDS

patient safety; psychological safety; medical curriculum; professional competence; clinical competence

It has been over 25 years since the beginning of the active development of the patient safety movement, during which the theories and methods of patient safety science have evolved. We now understand the key drivers and practices for safer care [1], and while there have been many achievements and successes in implementation, transfer to different contexts, reliability, and sustainability remain challenging.

One of the underlying problems is that the health care workforce has limited training in the theories and methods of patient safety and insufficient training in improvement or implementation science. Developing sustainable changes in the way we approach patient safety requires radically rethinking how we educate the health care workforce of the future, so that patient safety becomes integrated into the way they work, that is, part of the genome of a modern health care worker.

In the past, patient safety was presumed to be synonymous with being a professional, so it was implicit that on completing medical or nursing education, one would be safe. This clearly is not the case. The paper by Carillo et al [2] approaches this challenge from the perspective of the critical practice of psychological safety in the workforce as the foundation for safer care. They first assessed the current status of training programs for medical students and trainees across Europe, with regard to the acquisition of knowledge, skills, and attitudes about patient safety. This was followed by the development of a suggested set of competencies for psychological safety that should be acquired during training programs. The curriculum is a valuable addition to our understanding of foundations for safer care. The focus on psychological safety as a key competency of patient safety training, rather than a mere focus on knowledge, is a novel approach to the curriculum. Additionally, it can form the basis of a better response to the second victim following an adverse event. Several themes arose from the paper that should be considered.

First, the focus on psychological safety changes the focus of education, shifting away from concentrating only on knowledge acquisition, as in other curricula such as the World Health Organization (WHO) [3] patient safety curriculum, which is under review. A modern patient safety curriculum needs to specify the learner outcomes of knowing what to do and how to generate feelings of being psychologically safe when applying safety science in the workplace to create safer clinical teams. This will be essential for the delivery of the WHO Patient Safety Global Action Plan [4].

Second, a patient safety curriculum cannot stand alone outside the wider concepts of quality in health care. This is an ongoing debate, but many other domains of quality impact the safe delivery of care. Therefore, a patient safety curriculum should be part of a comprehensive set of competencies that facilitate the implementation of patient safety improvement initiatives. Knowledge and skills of improvement methodology and implementation science are essential, and there are examples

XSL•FO
RenderX

of frameworks that achieve this goal for comprehensive quality in health care [5]. Equally, psychological safety influences the success of improvement efforts, implementation efforts, and innovation, all of which depend on being able to speak openly and share new ideas without fear.

Third, it should be determined whether one can engender psychological safety via a training program alone and whether a program is the sole foundation on which a safe system can be built. Organizational culture is fundamental for psychological safety. Psychological safety can only thrive within a team or organization that has a culture of safety that includes a focus on communication, feedback, respect, and trust [6]. Although this is part of the program suggested, there is often a disconnect between the theoretical classroom and trainees' lived experiences. Psychological safety requires positive team and organizational relationships that facilitate team members being safe [7]. The proposed framework for psychological safety includes structural, interpersonal, and individual factors that extend beyond education and depend heavily on leadership. Applying this in practice is challenging [8].

Fourth, to create a strategy for safer care delivery, we need to consider the reasons why medical education has not made patient safety an integral part of the curriculum, despite growing evidence of interventions that decrease harm and create a safer health care environment. Most academic institutions remain hierarchical and are steeped in the traditional medical model of teaching. Reasons for the reluctance to incorporate patient safety include lack of awareness of the emerging science, lack of leadership prioritization, curriculum overload, and competition with other emerging sciences [9]. For the proposed curriculum to succeed, these challenges need to be addressed head-on, and a radical rethink of medical education is required.

Finally, we need to consider the efficacy of patient safety training to make a difference. Two systematic reviews indicate the heterogeneity of papers that assess the effectiveness of patient safety education programs. The link between education and improved clinical outcomes is not strong [10]. There appears to be a disconnect between undergraduate patient safety training and what happens in the clinical setting [11]. This indicates the need for training programs to be integrated into postgraduate and undergraduate programs. It also suggests the need for early evaluation of any new program to ensure that what is imagined is being achieved.

In conclusion, Carillo et al [2] have shown a way forward for patient safety training. The challenge will be implementation within traditional medical education curricula. Perhaps the solution to this could be coproduced by educators, trainees, and patients rather than created by experts and mentors alone. Even though the goal must be transformation in how patient safety is considered within medical education, we can start to create the conditions for these competencies to thrive at our next classroom meeting, simulation session, team huddle, or handover. Imagine the power of a senior clinician openly sharing their vulnerability of not being able to know everything that is required to be safe, inviting respectful dissent, and graciously embracing difficult news. That is the improvement way!

Conflicts of Interest

None declared.

References

1. Lachman P, Runnacles J, Jayadev A, Brennan J, Fitzsimons J, editors. Oxford Professional Practice: Handbook of Patient Safety: Oxford University Press; 2022. [doi: 10.1093/med/9780192846877.001.0001]
2. Carrillo I, Skoumalová I, Bruus I, et al. Psychological safety competency training during the clinical internship from the perspective of health care trainee mentors in 11 pan-european countries: mixed methods observational study. JMIR Med Educ 2024 Oct 7;10:e64125. [doi: 10.2196/64125] [Medline: 39374073]
3. Patient safety curriculum guide: multi-professional edition. World Health Organization. 2011 Jul 6. URL: https://www.who.int/publications/i/item/9789241501958 [accessed 2025-01-06]
4. Global patient safety action plan 2021-2030. World Health Organization. 2021. URL: https://www.who.int/publications/b/57613 [accessed 2025-01-06]
5. Schrimmer K, Williams N, Mercado S, Pitts J, Polancich S. Workforce competencies for healthcare quality professionals: leading quality-driven healthcare. J Healthc Qual 2019;41(4):259-265. [doi: 10.1097/JHQ.0000000000000212] [Medline: 31283704]
6. Hallam KT, Popovic N, Karimi L. Identifying the key elements of psychologically safe workplaces in healthcare settings. Brain Sci 2023 Oct 11;13(10):1450. [doi: 10.3390/brainsci13101450] [Medline: 37891818]
7. Edmondson AC. The Fearless Organization: Creating Psychological Safety in the Workplace for Learning, Innovation, and Growth: John Wiley & Sons, Inc; 2019.
8. Ito A, Sato K, Yumoto Y, Sasaki M, Ogata Y. A concept analysis of psychological safety: further understanding for application to health care. Nurs Open 2022 Jan;9(1):467-489. [doi: 10.1002/nop2.1086] [Medline: 34651454]
9. Wu AW, Busch IM. Patient safety: a new basic science for professional education. GMS J Med Educ 2019 Mar 15;36(2):Doc21. [doi: 10.3205/zma001229] [Medline: 30993179]
10. Kirkman MA, Sevdalis N, Arora S, Baker P, Vincent C, Ahmed M. The outcomes of recent patient safety education interventions for trainee physicians and medical students: a systematic review. BMJ Open 2015 May 20;5(5):e007705. [doi: 10.1136/bmjopen-2015-007705] [Medline: 25995240]

XSL•FO
RenderX

11. Sheehan P, Joy A, Fleming A, Vosper H, McCarthy S. Human factors and patient safety in undergraduate healthcare education: a systematic review. Hum Factors Health 2022 Dec;2(2772-5014):100019. [doi: 10.1016/j.hfh.2022.100019]

Abbreviations

WHO: World Health Organization

Edited by B Lesselroth; submitted 27.10.24; this is a non–peer-reviewed article;accepted 07.12.24; published 17.01.25.

Please cite as:
Lachman P, Fitzsimons J
Transforming Medical Education to Make Patient Safety Part of the Genome of a Modern Health Care Worker
JMIR Med Educ 2025;11:e68046
URL: https://mededu.jmir.org/2025/1/e68046
doi:10.2196/68046

© Peter Lachman, John Fitzsimons. Originally published in JMIR Medical Education (https://mededu.jmir.org), 17.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on https://mededu.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX