

JMIR Medical Education

Impact Factor (2024): 3.2

Volume 12 (2026) ISSN 2369-3762 Editor-in-Chief: Blake J. Lesselroth, MD, MBI, FACP, FAMIA

Contents

Original Papers

Evaluating AI-Generated Geriatric Case Studies for Interprofessional Education: Systematic Analysis Across 5 Platforms (e83085)	4
Nicole Ruggiano, Sudikshya Sahoo, Ava Brashear, Uche Nwatu, Amie Brunson, Hyunjin Noh, Heather Cole, Robert McKinney, C Framil Suarez, Ellen Brown, Suzanne Prevost.	
Global Learner Feedback on Continuing Medical Education—Accredited e-Learning Modules in Pediatric Endocrinology and Diabetes: Cross-Sectional Study (e67332)	41
Yvonne van der Zwan, Conny van Wijngaard-de Vugt, Abdulsalam Abu-Libdeh, Evangelia Kalaitzoglou, Zacharoula Karabouta, Stenvert Drop, Annemiek Boot, Sze Ng, Jan Idkowiak.	
Impact of Community-Oriented Medical Education on Medical Students' Perceptions of Community Health Care: Qualitative Study (e84406)	54
Kiyoshi Shikino, Kazuyo Yamauchi, Nobuyuki Araki, Naoto Ozaki, Yu Kamata, Shinya Aoki, Yota Katsuyama, Daichi Sogai, Mai Miyamoto, Kensuke Yoshimura, Takeshi Oki, Shoichi Ito.	
Comparative Efficacy of Simulation-Based and Traditional Training in Ultrasound-Assisted Regional Anesthesia for Medical Students: Randomized Controlled Trial (e77702)	68
David Sánchez-Poveda, Juan Vicente-Mampel, Belén Curto, Vidal Moreno, Juan García-Estebar, Felipe Hernández-Zaballos, Pablo Alonso-Hernández.	
Influence of Admission Pathways on Learning Strategies, Assessment Engagement, and Academic Performance Among First-Year Medical Students: Mixed Methods Retrospective Observational and Cross-Sectional Survey Study (e68636)	79
Issarawan Keadkraichaiwat, Chantacha Sittcharoorn, Punyapat Maprapho, Nisa Jangboon, Nadda Wannarat.	
The Effect of a Traditional Chinese Medicine Course on Western Medicine Students' Attitudes Toward Traditional Chinese Medicine: Self-Controlled Pre-Post Questionnaire Study (e55972)	106
Haoyu He, Hanjin Cui, Li Guo, Yanhui Liao.	
Interactive, Image-Based Modules as a Complement to Prosection-Based Anatomy Laboratories: Multicohort Evaluation (e85028)	117
Caroline Sumner, Sami Case, Samuel Franklin, Kristen Platt.	
Disaster Medicine Training for Medical Students in Lebanon: Quasi-Experimental Comparison of e-Learning and Face-to-Face Modalities (e80409)	127
Ali Msheik, Ruben Peralta, Zeinab Al Mokdad, Christine Bartulec, Linda Abou-Abbas, Dima Anani, Hussein Al Mawla, Hassan Salame, Mohamad Moussa, Awsan Bahattab.	

Enhancing Engagement With Endocrine Guidelines and Fostering Medical Student Interest Through Concise Medical Information Cines: Qualitative Co-Design Study (e83711)	143
Rahul Sagu, Sangamithra Ravi, Aspasia Manta, Punith Kempegowda.....	
Web-Based Virtual Environment Versus Face-To-Face Delivery for Team-Based Learning of Anesthesia Techniques Among Undergraduate Medical Students: Randomized Controlled Trial (e80097)	151
Darunee Sripadungkul, Suhattaya Boonmak, Monsicha Somjit, Narin Plailaharn, Wimonrat Sriraj, Polpun Boonmak.....	
Virtual Reality in Training and Assessment Among Clinical Students and Lecturers at a Nigerian University: A Phenomenological Study (e75021)	163
Abiola Komolafe, Omotade Ijarotimi, Oluseye Okunola, Olufemi Adetutu, Ayodeji Oluwatope, Olatunde Abiona, Ojo Agunbiade, Adeboye Ayinde, Stephen Aregbesola, Olawale Akinwale, Babatope Kolawole, Lanre Idowu, Alaba Adediwura, Olayinka Otuyemi.....	
Live Podcasting as an Educational Intervention in Dentomaxillofacial Radiology: Controlled Cohort Study (e77980)	175
Anna-Lena Hillebrecht, Daniel Fritzsche, Thamar Voss, Anne Kruse, Andreas Keßler, Kirstin Vach, Markus Altenburger, Rainer Schmelzeisen, Wiebke Semper-Hogg.....	
Development of Therapeutic Competencies in Health Care Students: Qualitative Focus Group Study Using 360-Degree Video and Virtual Reality Technology (e75776)	186
Camilla Lauritzen, Charlotte Reedtz, Kjersti Kristensen, Vår Mathisen, Eva Næss, Rigmor Furu, Hege Nermo, Rita Jentoft.....	
Implementation and Evaluation of a Cancer Immunotherapy Elective for Medical Students: Mixed Methods Descriptive Study (e71628)	196
Mark Raynor, Rivers Hock, Brandon Godinich, Satish Maharaj, Houriya Ayoubieh, Cynthia Perry, Jessica Chacon.....	
Effect of an Online Continuing Professional Development Course on Physicians' Intention to Approach a Colleague in Difficulty: Mixed Methods Convergent Study (e80199)	213
Florence Lizotte, Martin Tremblay, Caroline Biron, Éloi Lachance, Souleymane Gadio, Roberta de Carvalho Corôa, Claude Uwizeye, Sam Daniel, France Légaré.....	
Combining Problem-Based Learning Methods With the WeChat Platform in Teaching Ophthalmology: Randomized Controlled Trial (e65279)	234
Fang Fang, Bing Bu, Wenmin Jiang.....	
Ultrasound-Guided Regional Anesthesia in a Resource-Limited Hospital: Prospective Pilot Study of a Hybrid Training Program (e84181)	246
Jakob Gamboa, Inge Tamm-Daniels, Roland Flores, Nancy Sarat Diaz, Mario Villasenor, Mitchell Gist, Aidan Hoie, Christopher Kurinec, Colby Simmons.....	
Exploring the Icarus Paradox in Indonesia's Specialist Medical Education System Using the Public Perspective From Online Media: Convergent Mixed Methods Study (e60452)	260
Faisal Binsar, Mohammad Hamsal.....	
Investigating the Impact of a Virtual Reality Experience on Medical Student Empathy: Mixed Methods Study (e76504)	280
Allen Mundok, Vivian Ho, Lauren Fowler, Ann Kennedy, Shannon Stark-Taylor.....	
Blended Learning Compared With Face-to-Face Learning Among Family Medicine Residents: Randomized Controlled Trial (e86387)	292
Pierre-Yves Meunier, Sophie Schlatter, Juliette Macabrey, Frédéric Zorzi, Thomas Colleony, Rémy Boussageon, Hubert Maisonneuve, Marion Lamort-Bouché.....	

Fostering Inductive and Deductive Learning in Oral Microbiology and Immunology With a Dual-Role Duel Card Game: Explanatory Sequential Mixed Methods Study (e80048)	305
Kawin Sipiyaruk, Ratchapin Srisatjaluk, Ponpentip Chaisiri, Kanokwan Jantawatchai, Jidapha Srihatai, Chutikarn Sonyaem, Thanaphat Traikorakot, Suphasuta Pookboonmee.....	
Data Science Education for Residents, Researchers, and Students in Psychiatry and Psychology: Program Development and Evaluation Study (e75125)	321
Hayoung Donnelly, David Mandell, Sy Hwang, Emily Schriver, Ugurcan Vurgun, Graydon Neill, Esha Patel, Megan Reilly, Michael Steinberg, Amber Calloway, Robert Gallop, Maria Oquendo, Gregory Brown, Danielle Mowery.....	
Effectiveness of Informed AI Use on Clinical Competence of General Practitioners and Internists: Pre-Post Intervention Study (e75534)	329
Eyad Qunaibi, Ayman Al-Qaaneh, Baraa Ismail, Hussam Muhidat, Farhia Rageh, Najwa Musallam, Alaa Fawzy.....	
Comparing AI-Assisted Problem-Solving Ability With Internet Search Engine and e-Books in Medical Students With Variable Prior Subject Knowledge: Cross-Sectional Study (e81264)	343
Ajiith Xavier, Syed Naeem, Waseem Rizwi, Hiramani Rabha.....	
AI Literacy Among Chinese Medical Students: Cross-Sectional Examination of Individual and Environmental Factors (e80604)	351
Chunqing Li, Sian Tsuei, Hongbin Wu.....	
GPT-4o and OpenAI o1 Performance on the 2024 Spanish Competitive Medical Specialty Access Examination: Cross-Sectional Quantitative Evaluation Study (e75452)	365
Pau Benito, Mikel Isla-Jover, Pablo González-Castro, Pedro Fernández Esparcia, Manuel Carpio, Iván Blay-Simón, Pablo Gutiérrez-Bedia, María Lapastora, Beatriz Carratalá, Carlos Carazo-Casas.....	
Adaptation of the Japanese Version of the 12-Item Attitudes Towards Artificial Intelligence Scale for Medical Trainees: Multicenter Development and Validation Study (e81986)	377
Hirohisa Fujikawa, Hirotake Mori, Kayo Kondo, Yuji Nishizaki, Yuichiro Yano, Toshio Naito.....	
Evaluation of a Problem-Based Learning Program's Effect on Artificial Intelligence Ethics Among Japanese Medical Students: Mixed Methods Study (e84535)	387
Yuma Ota, Yoshikazu Asada, Saori Kubo, Takeshi Kanno, Machiko Yagi, Yasushi Matsuyama.....	
AI-Driven Objective Structured Clinical Examination Generation in Digital Health Education: Comparative Analysis of Three GPT-4o Configurations (e82116)	402
Zineb Zouakia, Emmanuel Logak, Alan Szymczak, Jean-Philippe Jais, Anita Burgun, Rosy Tsopra.....	

Viewpoints

Digital Choice Architecture in Medical Education: Applying Behavioral Economics to Online Learning Environments (e86497)	18
Victoria Ekstrom.....	
Implementing Large Language Models to Support Misconception-Based Collaborative Learning in Health Care Education (e81875)	28
Brandon Cheah, Shefaly Shorey, Jun Ch'ng, Chee Tan.....	

Original Paper

Evaluating AI-Generated Geriatric Case Studies for Interprofessional Education: Systematic Analysis Across 5 Platforms

Nicole Ruggiano¹, MSW, PhD; Sudikshya Sahoo¹, MSW; Ava Brashear²; Uche Nwatu¹, MSSW; Amie Brunson¹, LMSW, DSW; Hyunjin Noh¹, MSW, PhD; Heather Cole³, RN, PhD; Robert McKinney⁴, MSW, PhD; C Victoria Framil Suarez⁵, RN, DNP; Ellen L Brown⁵, RN, MS, EdD; Suzanne Prevost³, RN, PhD

¹School of Social Work, University of Alabama, Tuscaloosa, AL, United States

²Department of Kinesiology, College of Education, University of Alabama, Tuscaloosa, AL, United States

³Capstone College of Nursing, University of Alabama, Tuscaloosa, AL, United States

⁴College of Community Health Sciences, University of Alabama, Tuscaloosa, AL, United States

⁵Nicole Wertheim College of Nursing and Health Sciences, Florida International University, Miami, AL, United States

Corresponding Author:

Nicole Ruggiano, MSW, PhD

School of Social Work

University of Alabama

801 University Blvd.

129 Farrah Hall

Tuscaloosa, AL, 35401

United States

Phone: 1 2053484654

Email: nruggiano@ua.edu

Abstract

Background: Simulation-based learning (SBL) has become standard practice in educating health care professionals to apply their knowledge and skills in patient care. While SBL has demonstrated its value in education, many educators find the process of developing new, unique scenarios to be time-intensive, creating limits to the variety of issues students may experience within educational settings. Generative artificial intelligence (AI) platforms, such as ChatGPT (OpenAI), have emerged as a potential tool for developing simulation case studies more efficiently, though little is known about the performance of AI in generating high-quality case studies for interprofessional education.

Objective: This study aimed to generate geriatric case scenarios across 5 AI platforms by a transdisciplinary team and systematically evaluate them for quality, accuracy, and bias.

Methods: Ten geriatric case studies were generated using the same prompt from 5 different generative AI platforms (N=50): ChatGPT, Claude (Anthropic AI), Copilot (Microsoft), Gemini (Google), and Grok (xAI). An evaluation tool was developed to collect evaluative data to assess the content and quality of each case, sociodemographic data of the featured patient, the appropriateness of each case for interprofessional education, and potential bias. Case quality was evaluated using the Simulation Scenario Evaluation Tool (SSET). Each case was evaluated by 3 team members who had experience in SBL education. Assessment scores were averaged, and qualitative responses were extracted to triangulate patterns found in the quantitative data.

Results: While each AI platform was able to generate 10 unique case studies, the quality of studies varied within and across platforms. Generally, evaluators felt that the content in the cases was accurate, though some cases were not realistic. Some patient populations and common conditions among older adults were underrepresented or absent across the cases. All cases were set within traditional health care settings (eg, hospitals and routine medical visits). No cases featured home-based care. Based on the average SSET scores, reviewers assessed ChatGPT to be the highest overall performer (mean 3.27, SD 0.45, 95% CI 2.95-3.59) while Grok received the lowest scores (mean 1.61, SD 1.26, 95% CI 0.71-2.51). Platforms performed best at generating learning objectives (mean 3.35, SD 1.08, 95% CI 3.04-3.65) and lowest on their ability to describe supplies and materials that may be available in hypothetical scenarios (mean 1.27, SD 0.84, 95% CI 1.03-1.51).

Conclusions: This study is the first to systematically evaluate and compare multiple generative AI platforms for case study generation using a validated assessment tool (SSET) and provides evidence-based guidance on selecting and using AI tools effectively. The findings offer practical direction for educators navigating available generative AI tools to enhance training for health care professionals, including specific strategies for prompt engineering that can improve the quality of SBL resources in interprofessional education. These insights enable educators to leverage AI capabilities while maintaining pedagogical rigor.

(*JMIR Med Educ* 2026;12:e83085) doi:10.2196/83085

KEYWORDS

chatbots; generative artificial intelligence; geriatric nursing; interprofessional education; patient simulations

Introduction

Simulation-Based Learning in Health Care Professional Education

Simulation-based learning (SBL) is an integral component of health care professional education, offering a safe and controlled environment for students to develop clinical skills without risking patient safety [1,2]. SBL is recognized as an effective educational tool that has grown within higher education in many disciplines and can be implemented using a number of approaches, including role-playing, games, computer-assisted virtual instruction, and specially designed skills laboratories [3]. SBL often requires educators to develop highly detailed case studies to guide student learning activities, though these are often challenging to design. Recently, there has been increasing interest in the use of artificial intelligence (AI) platforms to support health care education [4]. While generative AI tools have the potential to generate case studies for use in SBL, little is known about the potential for AI platforms to generate high-quality, effective cases suitable for interprofessional health education. To address this, an interdisciplinary team systematically generated and evaluated geriatric case studies from 5 commercially available generative AI platforms and evaluated their content. The findings have implications for how AI platforms may be appropriately and ethically integrated into SBL.

History of Simulation-Based Learning in Interprofessional Education

The use of simulation in health care dates back decades, with early examples involving standardized patients, who were individuals trained to portray real patients [5]. This innovation allowed medical trainees to practice patient interactions and diagnostic skills in realistic scenarios. Advancements in technology have enhanced the fidelity, effectiveness, and types of simulation tools, leading to widespread adoption across various health care disciplines [6]. For example, the introduction of high-fidelity mannequins and virtual reality platforms has enabled the replication of complex clinical scenarios, thereby enhancing the realism and effectiveness of training sessions [7]. These technological innovations have expanded the scope of SBL, allowing for the practice of rare or high-risk procedures in a risk-free setting, which is crucial for developing competence in various medical specialties. Recent literature underscores the efficacy of SBL in improving clinical competencies among health care professionals. A systematic review by Sawaya et al [8] highlighted that SBL not only enhances immediate

knowledge and skill acquisition but also contributes to long-term retention of clinical competencies. Furthermore, the integration of SBL into medical curricula has been associated with improved patient outcomes and a reduction in medical errors, emphasizing its critical role in contemporary health care professional education.

The Benefits and Drawbacks of Simulation-Based Education

One of the primary advantages of SBL is its ability to create a safe learning environment where learners can engage in high-risk scenarios or practice clinical skills without the risks associated with working with actual patients [9,10]. This immersive learning experience is further enhanced using high-fidelity simulators that mimic real patient responses, thereby engaging students emotionally and cognitively and improving the transfer of skills to clinical practice [9,11]. SBL has also been shown to improve learners' self-confidence and satisfaction. For example, previous studies have found that students who participated in SBL in emergency medicine reported higher satisfaction and performed better on assessments compared to those who received traditional instruction [12-14]. This suggests that SBL not only enhances knowledge acquisition but also fosters a positive learning experience. Interprofessional SBL may also emphasize the importance of teamwork in geriatric care, which promotes better collaboration and communication among health care professionals while providing a protected environment where trainees can practice handling complex clinical situations and learn from mistakes in a supportive setting [15-17]. Group-based SBL also promotes active engagement with equipment, simulated patients, and peers, leading to a deeper understanding and increased motivation to learn [18]. This interactive environment stimulates intellectual curiosity, encouraging critical thinking, questioning, and exploration of new knowledge, which aligns with curricula goals to foster curiosity and critical thinking in students.

Despite its numerous benefits, SBL is not without its challenges. One significant drawback is that SBL can be time- and resource-intensive (when using a lab setting) and may also pose challenges when integrated into an already crowded curriculum [9]. Lin et al [19] have asserted that ignoring the implementation and sustainability of SBL can result in higher costs, wasted resources, and the potential failure of educational interventions that are otherwise effective in achieving learning outcomes. Finally, educators may have difficulty in developing scenarios that are realistic and tailored to student learning needs. For example, certain physical findings, such as skin color variations, cannot be adequately represented in certain SBL approaches,

which may hinder the comprehensive training of students [9]. Instructors may wish to tailor scenarios to the varying abilities of students, but the standardized nature of many simulation exercises can hinder this individualized approach. This limitation may result in some learners not receiving the optimal level of challenge or support needed for their development.

Use of AI in Health Professions Education

Over the past decade, there has been increasing interest in identifying potential uses of large language model (LLM) AI tools in education for health care professionals. While the literature has identified opportunities for integrating AI into SBL activities, other potential applications have also been identified [20]. For example, Glauberman et al [20] posit that AI may enhance activities where students explore the impact of social determinants of health on patient care, support tutoring, and provide students with real-time feedback on assignments. However, they also identified potential challenges, such as AI providing hallucinations or biased content, and the risk that students may become over-reliant on AI, which could inhibit critical thinking skills [20].

Given the rising interest, there has been an increase in studies about integrating AI into health education settings. It should be noted that although there are currently multiple commercially available AI LLM platforms, much of the literature on using AI in health education has focused on ChatGPT (OpenAI), due to its rapid adoption among the public, reaching 1 million users in its first 5 days alone [21,22]. In a systematic review of the health education literature, Sallam [23] found that ChatGPT had the potential for creating tailored educational content and communication skill development for students that provides immediate feedback. However, challenges cited in the review included concerns regarding the potential bias and accuracy of information in educational material generated through LLMs.

Rogers et al [24] explored ChatGPT's capacity for generating simulation scenarios for health care education by evaluating 2 patient simulation cases it generated. Rather than prompting the platform to generate a complete case, in their study, they used a series of prompts to create individual components of a larger case (ie, develop a goal statement, create a corresponding scenario, and add specific detail based on existing information). There were several strengths the reviewers identified in the scenarios, including providing clear learning objectives, creating real-world scenarios, highlighting medication dosing guidelines, and describing participants' roles for each case [24]. They also identified several weaknesses in the cases, including inaccurate medication dosing, not providing treatment guidelines, missing details on equipment and supplies that would be needed for the case, objectives that would be too advanced for the learner audience, and inaccurate references [24]. Overall, reviewers reported that the main strengths were in the debriefing content,

while inaccuracy posed the greatest problem, citing the need for users to fact-check the output of scenarios generated through ChatGPT.

Rationale for the Study

Geriatric education (or geriatric care) is an increasingly important area of health care professional training, given the growing population of older adults with complex and chronic health needs. As health care systems adapt to meet these challenges, equipping students with the knowledge and skills to provide competent, person-centered care to older adults is essential across disciplines. Generative AI tools have the potential to make some educational tasks more efficient, such as developing scenarios for interprofessional health education simulation exercises. However, the appropriateness of using AI platforms for this task remains unclear. The study by Rodgers et al [24] work evaluated the performance of ChatGPT in creating scenarios; it only evaluated 2 cases and did not compare ChatGPT with other AI platforms. To better understand how generative AI may be used by health educators, this study aimed to evaluate the performance of 5 generative AI platforms in creating high-quality case studies that could be used for interprofessional health education.

Methods

Overview

This study involved a systematic evaluation of geriatric case studies that were generated through AI platforms. The methods and results described below reflect the reporting standards for the Generative Artificial Intelligence Tools in Medical Research (GAMER) statement [25], which can be found through the Enhancing the Quality and Transparency of Health Research Network [26].

The cases were generated in December 2024 from 5 different AI platforms: ChatGPT, Claude (Anthropic AI), Gemini (Google), Grok (xAI), and Copilot (Microsoft). Copilot was chosen because it was the platform recommended by the research team's institution for use in education and research. The other 4 platforms were selected based on their rankings on the LMArena leaderboard, an open-access database of AI platforms developed at the University of California, Berkeley, where users rank various AI platforms on their performance [27]. Before evaluating cases, the team generated 2 cases on each platform's free version. However, it was discovered that 1 platform (Grok) did not have a free version available, and all the platforms varied in what was offered through their free versions (eg, accessing internet data in real time). Therefore, the decision was made to evaluate cases generated through the paid subscription versions of each platform to minimize bias. Table 1 provides a description and comparison of the platforms.

Table 1. Details of the 5 artificial intelligence (AI) platforms selected and their distinguishing qualities.

Platform	Developer	Launch date	Description and distinguishing features
ChatGPT [28]	OpenAI	November 2022 (Public launch)	Built on a generative pretrained transformer (GPT) architecture, the LLM ^a is trained on conversational data to generate human-like responses to queries.
Claude [29,30]	Anthropic AI	March 2023	LLM incorporating “Constitutional AI” principles based on the UN Declaration of Human Rights and AI ^b research ethics; emphasizes safety and neutrality through self-critiquing mechanisms.
Gemini [31,32]	Google	2023 (originally as Bard)	Multimodal LLM that processes words, audio, and pictures; designed for productivity and tasks including image generation.
Grok [33]	xAI (Elon Musk)	November 2023	LLM trained using systems such as Kubernetes, JAX, and Rust, designed to create a more efficient AI platform than those previously developed; has access to information in real time through the platform X (formerly Twitter); programmed to have a witty personality.
Copilot [34]	Microsoft	March 2023	Integrates with Microsoft 365 applications; can access users’ personal data to provide tailored responses within a larger productivity suite.

^aLLM: large language model.

^bAI: artificial intelligence.

Generation of Case Studies

To assess the performance of AI platforms in generating geriatric case studies, each platform was asked to generate 10 unique case studies, for a total of 50 cases. The decision was made to submit 10 individual requests for cases on each platform rather than a single request for all 10 after discovering that the platforms created more detailed and robust cases when asked to create them one at a time rather than all at once. The prompts used to generate cases were as follows:

- Initial prompt: generate a geriatric case study that can be used in simulation learning for students from various health care professional disciplines, such as medical, nursing, and social work students.
- Subsequent prompts: generate another unique geriatric case study that can be used in simulation learning for students from various health care professional disciplines, such as medical, nursing, and social work students.

Team members uploaded the cases to a secure cloud storage platform. Two team members (NR and Ava Brashear) blinded the cases so that reviewers would not be able to link any individual case to the platform that generated it.

Case Study Evaluation Tool

The team developed an evaluation questionnaire for case study evaluators (ie, social work and nursing faculty) to use in assessing the case studies. The following describes the variables assessed.

Case Quality

Although there are several tools available to evaluate students’ performance on SBL activities and the quality of debriefing, little work has been done to create assessment tools to evaluate the quality of simulation case content. The research team used the Simulation Scenario Evaluation Tool (SSET), developed by Hernandez et al [35], to evaluate the quality of case studies. The SSET is a standardized assessment tool that was developed using a modified Delphi approach, a structured, iterative process designed to build consensus among experts. The final version of the SSET includes 20 items that are organized into 5 elements, described in [Table 2](#). The SSET also allows the reviewer to skip elements that are not available in the case study description. For each item, evaluators assess cases using a 5-point Likert scale with 3 anchor points. The anchor points are individualized for each item.

Table 2. Description of the assessment elements included in the Simulation Scenario Evaluation Tool (SSET).

Element	Element description	Number of items
Learning objectives	If learning objectives are included, the extent to which they are: a good fit for students' skill and knowledge levels, specific, measurable, action-oriented, relevant, and reflect different types of knowledge or skills.	7
Clinical context and scenario overview	The amount and quality of information provided about the case facilitate learning outcomes.	2
Critical actions	If the case study describes actions or decisions that the student should be able to demonstrate after reading the case study, the extent to which the actions are observable, support the learning objectives, and are attainable for the students' skill level.	3
Patient states	The details and cues provided to learners can help them carry out the critical actions, including the appropriateness of the case study progression, the ability of learners to take multiple pathways to addressing the case, and how they may help facilitate learning objectives, and require students to take critical actions about the case.	4
Scenario materials and resources	The identification of equipment, supplies, and resources that would normally be available and needed to demonstrate outcomes in clinical skills.	2
Debriefing plan	The extent to which the case guides a postsimulation discussion where learners can receive feedback on their performance.	2

Sociodemographic Data

Sociodemographic information (eg, race, gender, and socioeconomic status) from each of the cases was extracted to evaluate the diversity of patient populations presented across cases.

Appropriate for Interprofessional Education

Reviewers were asked to rate the extent to which cases were clinically accurate using a Likert scale of 1 (Completely Inaccurate) to 4 (Highly Accurate), with the option of "Unsure." They were also asked to rate the extent to which the cases adequately addressed ethical considerations using a Likert scale of 1 (Completely Disagree) to 5 (Completely Agree). Reviewers were asked to indicate whether they would use each case for SBL (Yes, No, Unsure). Open-ended questions allowed evaluators to provide feedback on the quality of cases.

Potential Bias

Open-ended questions were used to allow evaluators to identify content in each case that they perceived as biased or stereotypical of the patient populations.

Process for Evaluating Cases

All case studies were generated in November 2024 and were evaluated between December 2024 and April 2025. Two coauthors (SS and UN) extracted the sociodemographic data from each case (eg, patient gender, race, ethnicity, and living situation). Six coauthors (Amie Brunson, ELB, CVFS, HC, RM, and HN) with expertise in interprofessional education in health care settings applied the SSET to evaluate the quality of cases generated. Each case was randomly assigned to 3 evaluators for review. The evaluation tool was uploaded to Qualtrics (Qualtrics International Inc), and each team member entered their assessments remotely and securely into a single database.

Data Analysis

All data were maintained in a Microsoft Excel spreadsheet for analysis. Descriptive statistics were generated for all of the sociodemographic data, including means, SDs, and CIs. The 3

reviewers' scores on the SSET for each case were averaged. Qualitative responses were extracted by 2 team members with expertise in qualitative health-related research methodologies (NR and HN) and used to triangulate patterns found in the quantitative data.

Post Hoc Examination

After the initial evaluation of the 50 case studies generated by AI for this project, a post hoc analysis was conducted to examine how alternative prompt engineering strategies informed by the study's findings could potentially improve the quality of AI-generated case studies. A detailed prompt was developed that addressed weaknesses and gaps observed in the original analysis and was used to generate a single case study using Claude (Anthropic AI). Details about this analysis and its implications for prompt engineering are provided in the "Discussion" section.

Ethical Considerations

This study did not involve human participants. All data were generated by the AI platforms from information drawn from the internet and other sources on which they had been trained by their developers. No data about real patients were entered into the platforms for this study.

Results

Overall Performance

All of the platforms demonstrated the ability to generate comprehensive case studies, though the quality of the cases varied within and across platforms. All cases were organized into content sections, often including a patient profile and demographics, primary presenting complaints of the patient, lists of current medications, past medical history, social history, and assessment results. Some cases generated learning objectives for students without additional prompting. However, for cases generated by Grok, several were incomplete, sometimes being cut off midsentence. For example, one case study generated by Grok featuring the fictitious patient, Mr Jack O'Connor,

included, “Current Presentation: Mr O’Connor was referred to...” (sic). In another example featuring Mrs Elena Rodriguez, the case included, “Psychosocial Factors:-...” (sic).

Variety of Scenario Details

The cases featured fictitious patients who varied in social and demographic backgrounds. There was some variation in the presenting complaints and medical history, though cases tended to emphasize the most common conditions impacting older adults, such as dementia, diabetes, heart disease, and hypertension. Interestingly, none of the 50 cases addressed COVID-19.

Table 3 provides the breakdown of sociodemographic information provided in the patient cases. There was about equal representation of women (n=26) and men (n=24), as well as a

variety of races and ethnicities. It should be noted that none of the cases identified the patient as being a member of the lesbian, gay, bisexual, transgender, queer, intersex, asexual, and more community, having undocumented immigration status, or identifying as Native Hawaiian, Alaska Native, or American Indian. There was also a lack of information regarding social determinants of health, most notably income and insurance information. Qualitative data provided by evaluators revealed that the topics of patient finances, culture, and religion were insufficiently addressed in most cases, which did not provide students with adequate context for addressing patient concerns. For example, patients’ decisions regarding treatment or medication, as well as behaviors such as medication noncompliance, may be influenced by their income, cultural background, or religious beliefs.

Table 3. Sociodemographic characteristics of the fictitious patients featured in the case studies generated by artificial intelligence (AI) platforms.

Characteristic	Value, n (%)
Sex	
Cisgender male	24 (48)
Cisgender female	26 (52)
Race and ethnicity	
Caucasian or White	9 (18)
African American, Black, or Afro-Caribbean	8 (16)
Latino or Latina	5 (12)
Asian or Pacific Islander	9 (18)
Native Hawaiian, Alaska Native, or American Indian	0 (0)
Other	2 (4)
Unknown	16 (32)
English proficiency	
Yes	21 (42)
No (has a family member interpreting)	3 (6)
Unknown	26 (52)
Sexual orientation	
Heterosexual	29 (58)
Unknown	21 (42)
Marital status	
Married	17 (34)
Divorced	5 (10)
Widowed	26 (52)
Committed partner	1 (2)
Unknown	1 (2)
Employment	
Never worked	1 (2)
Retired, not working	45 (90)
Working full-time	1 (2)
Other	3 (6)
Patient household income	
Identified as low-income	3 (6)
Identified as moderate-income	1 (2)
Identified as high-income	0 (0)
Unknown or not provided	46 (92)
Source of payment for health services	
Medicare	11 (22)
Medicare and Medicaid	5 (10)
Medicare and VA ^a benefits	2 (4)
Medicare with supplemental private insurance	2 (4)
Unknown	30 (60)
Living situation	
Lives at home alone	15 (30)

Characteristic	Value, n (%)
Lives with others in their own home	17 (34)
Lives with others in another person's home	3 (6)
Lives in an independent living facility	4 (8)
Lives in an assisted living facility	6 (12)
Lives in a nursing home or a similar setting	3 (6)
Unknown	2 (4)
Religious affiliation	
Christian (Catholic)	1 (2)
Jewish	1 (2)
Hindu	1 (2)
Other	2 (4)
Unknown	45 (90)
Immigration and citizenship	
Immigrant	4 (8)
Unknown	46 (92)

^aVA: Veterans Affairs.

Case Quality

Accuracy and Best Practices

For most cases, evaluators assessed the information provided as being generally accurate and reflecting best practices. However, they reported discrepancies with some case information, most often due to insufficient detail. For example, in reference to the SBL instructions for students, “*Develop a nursing care plan for hip fracture recovery, emphasizing mobility, prevention of pressure ulcers, and patient education on safety at home post*” in the case of Agness Muller (Grok), one evaluator provided qualitative feedback:

This instruction lacks detail and the usual patient progression. For example, a patient living alone with mild cognitive impairment (MCI)—based only on the Montreal Cognitive Assessment, which is not conclusive—and with a hip fracture would not be discharged home.

However, in a few cases, the evaluators felt that the case information was not realistic. For example, in another case of George Hawkins (Claude), one evaluator commented: “*I do not feel like the patient would be coming in stating that he is forgetting things. This information would be coming from the caregiver.*” Similarly, another evaluator commented on the same case: “*I think having the patient present with complaints of confusion would be a little unrealistic. We would likely see the caregiver reporting these findings to the health care provider.*”

SSET Scores

For most cases, enough details were provided for reviewers to evaluate the 6 elements represented in the SSET assessment. Yet, a small number of cases lacked details on 1 or more elements. Table 4 shows the average scores assigned by reviewers along with their SDs and CIs for each element across AI platforms.

Table 4. Evaluators' average scores, with SDs and 95% CIs, on the Simulation Scenario Evaluation Tool (SSET) and its 6 elements.

Platform	E ^a 1, mean, SD, 95% CI	E2, mean, SD, 95% CI	E3, mean, SD, 95% CI	E4, mean, SD, 95% CI	E5, mean, SD, 95% CI	E6, mean, SD, 95% CI	Overall, platform mean, SD, 95% CI
ChatGPT	4.22, 0.29, 4.01-4.42	3.89, 0.82, 3.30-4.47	3.71, 0.63, 3.26-4.15	3.28, 0.95, 2.60-3.96	1.78, 1.05, 1.02-2.53	2.74, 0.84, 2.14-3.34	3.27, 0.45, 2.95-3.59
Claude	3.74, 0.57, 3.33-4.15	3.59, 1.12, 2.79-4.39	3.34, 0.93, 2.68-4.00	3.50, 1.19, 2.65-4.35	1.58, 0.77, 1.03-2.13	1.70, 0.98, 1.00-2.40	2.91, 0.63, 2.46-3.36
Copilot	3.32, 0.42, 3.02-3.63	2.96, 0.83, 2.36-3.55	2.75, 0.93, 2.08-3.42	2.10, 0.67, 1.62-2.58	0.99, 0.28, 0.78-1.19	1.26, 0.71, 0.75-1.77	2.23, 0.39, 1.95-2.51
Gemini	3.30, 0.49, 2.95-3.66	3.15, 0.77, 2.60-3.71	2.87, 0.66, 2.39-3.35	2.39, 0.73, 1.87-2.91	1.32, 0.84, 0.72-1.92	1.20, 0.66, 0.73-1.67	2.37, 0.49, 2.02-2.72
Grok	2.15, 1.69, 0.93-3.37	2.52, 1.97, 1.10-9.3	1.90, 1.45, 0.87-2.94	1.53, 1.38, 0.53-2.52	0.70, 0.73, 0.18-1.22	0.84, 1.03, 0.11-1.58	1.61, 1.26, 0.71-2.51
Overall, E ^a ,	3.35, 1.08, 3.04-3.65	3.22, 1.24, 2.87-3.57	2.91, 1.11, 2.60-3.23	2.56, 1.23, 2.21-2.91	1.27, 0.84, 1.03-1.51	1.55, 1.05, 1.25-1.85	2.48, 0.90, 2.22-2.73

^aE: element.

Generally, the AI platforms performed best and most consistently at providing details related to element 1 (E1: learning objectives). While objectives varied in focus and scope, the platforms generated learning objectives that were detailed. Except for Grok, the mean scores for the platforms were relatively good (range of 3.30-4.22). ChatGPT particularly excelled at generating learning objectives with consistency,

given the narrow CIs (mean 4.22, SD 0.29, 95% CI 4.01-4.42). While Grok received lower scores for E1, there was also variability in their quality (mean 2.15, SD 1.69; CI 0.93-3.37).

In some cases, the AI generated a set of learning objectives tailored to the various disciplines that may be involved in the simulation learning activities. **Textbox 1** provides an example generated from Grok.

Textbox 1. Learning objectives generated by Grok for the case of Mrs Isabella Bianchi.

Medical students:
<ul style="list-style-type: none"> Manage complex medication regimen for multiple chronic conditions, focusing on interactions and side effects. Investigate causes of weight loss in older adults, considering both medical and psychological factors.
Nursing students:
<ul style="list-style-type: none"> Implement and teach strategies for medication adherence in patients with mild cognitive impairment. Develop a care plan for preventing falls and managing chronic constipation.
Social work students:
<ul style="list-style-type: none"> Address the emotional impact of transitioning to assisted living, focusing on preserving dignity and a sense of purpose. Explore community resources or programs that could engage Mrs Bianchi's interest in music, potentially improving her mental health.
Interdisciplinary goals:
<ul style="list-style-type: none"> Enhance Mrs Bianchi's quality of life by balancing her medical needs with her personal interests and psychological well-being. Coordinate a care plan that includes physical activity tailored to her condition, dietary adjustments for heart health and bone density, and social engagement to combat isolation.

The AI platforms also performed moderately well on element 2 (E2: clinical context and scenario overview), as evidenced by mean scores ranging from 2.96 to 2.52. Claude (mean 3.59, SD 1.12) and Grok (mean 2.52, SD 1.97) demonstrated the greatest variability in quality.

Evaluators generally felt that the platforms provided sufficient detail for students to understand the presenting case and address

the learning objectives. However, they also noted that most cases described clinical contexts occurring within a health care facility (eg, hospital and primary care office), which may limit applicability for students in professions that often interact with patients in other settings (eg, home visits and pharmacies). There were exceptions, such as the example generated by Gemini in **Textbox 2**, where a home setting was used as the clinical context.

Textbox 2. Clinical context description for the case Mr Arthur Chen generated by Gemini.

Simulation activities:

- Home visit: students can conduct a simulated home visit to assess Mr Chen's living environment and provide education on disease management and self-care.
- Caregiver support group: students can role-play a caregiver support group to provide Mr Chen with an opportunity to share his experiences and connect with others.
- Interprofessional case conference: students from different disciplines can participate in a case conference to discuss Mr Chen's care plan and coordinate services.
- Medication management simulation: students can practice medication reconciliation and develop strategies to improve adherence.
- Advance care planning role-play: students can role-play a conversation with Mr Chen and his family about advance care planning.

Element 3 (E3: critical actions) and element 4 (E4: patient states) were rated more modestly by evaluators, with average scores of 2.91 (SD 1.11) and 2.56 (SD 1.23), respectively. Evaluators generally felt that the cases offered enough detail for students

to make decisions about patient care and carry out critical actions to support patient care needs. In many cases, a progression of the scenario was presented, as exemplified in [Textbox 3](#).

Textbox 3. Scenario progression for the case of Mr Walter Freemen, generated by Claude (UTI: urinary tract infection).

Simulation scenario progression:

- Initial assessment: each discipline conducts its respective assessments in the emergency department or upon admission.
- Team huddle: interdisciplinary team meets to discuss findings and develop an initial care plan.
- Acute management: simulate management of UTI and delirium, including nonpharmacological interventions.
- Medication review: team collaborates on medication reconciliation, considering potential cognitive effects and drug interactions.
- Family meeting: simulate a meeting with Mrs Evelyn Chen (when more lucid) and her husband to discuss diagnosis, treatment plan, and support needs.
- Discharge planning: team develops a comprehensive plan for transition of care, including medication management, follow-up appointments, and caregiver support.
- Follow-up: simulate a postdischarge follow-up appointment to assess resolution of acute issues and manage chronic conditions.

Evaluators rated the cases the lowest for element 5 (E5: scenario materials) and element 6 (E6: debriefing plan). All platforms performed poorly on E5 (mean 1.27, SD 0.84, 95% CI 1.03-1.51) and E6 (mean 1.55, SD 1.05, 95% CI 1.25-1.85). For E5, ChatGPT led with an average score of 1.78 (SD 1.05), followed by Claude (mean 1.58, SD 0.77) and Gemini (mean 1.32, SD 0.84). Copilot (mean 0.99, SD 0.28) and Grok (mean 0.70, SD 0.73) had the lowest scores. Evaluators found it difficult to evaluate the supplies and materials needed for students to complete the simulation cases, as they had to make assumptions about what would typically be available in the specific setting (eg, emergency room). However, they acknowledged that greater detail should be provided by most of the platforms to facilitate learning for students who may be less familiar with the setting. For E6, Grok had the lowest score at 0.84 (SD 1.03), followed by Gemini (mean 1.20, SD 0.66), Copilot (mean 1.26, SD 0.71), and Claude (mean 1.70, SD 0.98). ChatGPT fared slightly better in this area with an average of 2.74 (SD 0.84), but still fell short of high performance. Evaluators noted that many cases lacked a debriefing plan, which would require that the educator create one to coincide with the cases.

Comparison of Platform Performance

When asked whether they would use the presenting case for simulation learning, evaluators most often stated that they would

use the cases generated by ChatGPT and were least likely to use cases generated by Copilot. When evaluators were asked to rate the extent to which they agreed that the case addressed ethical issues that students may experience in health care settings, ChatGPT and Gemini generally received the highest scores (mean 3.50, SD 0.61 and mean 3.51, SD 0.83, respectively), corresponding to "somewhat agree." Although evaluators thought many of the cases did not sufficiently address issues of culture, they did not report observing examples of gender or racial stereotypes.

[Table 4](#) provides the summary scores for the SSET across platforms. Overall, ChatGPT performed the best in generating quality case studies, while Grok was rated the lowest, mostly due to the incomplete content generated in many of its cases. However, the platforms varied in their strengths and weaknesses. Among all platforms, ChatGPT performed the best in element 1: learning objectives, with the highest average score of 4.22 (SD 0.29), followed by Claude at 3.74 (SD 0.57). Copilot and Gemini had almost identical scores, with averages of 3.32 (SD 0.42) and 3.30 (SD 0.49), respectively, showing only a marginal difference. Grok, however, had the lowest performance in this element, with an average score of just 2.15 (SD 1.69). For element 2, which focused on understanding the relevance and alignment of the clinical context and scenario across case studies, the average scores of ChatGPT (mean 3.89, SD 0.82),

Claude (mean 3.59, SD 1.12), and Gemini (mean 3.15, SD 0.77) were relatively close. In contrast, Copilot (mean 2.96, SD 0.83) and Grok (mean 2.52, SD 1.97) had lower averages.

Discussion

Principal Findings

This study aimed to evaluate the performance of 5 commercially available AI platforms in their ability to generate geriatric case studies that could be used for simulation learning within interprofessional education settings. Although there was variation in their performance, all 5 platforms evaluated for this study were able to generate comprehensive case studies that generally included accurate information and a variety of patient contexts. ChatGPT was found to be the most reliable of the AI platforms, as demonstrated by consistently high scores and narrower CIs. Grok performed unpredictably and consistently received the lowest scores. It is not surprising that the AI platforms varied in their performance, given that the content generated would be specific to the data used to train each platform and its algorithms. For example, Claude's algorithm emphasizes a Constitutional AI framework, which may result in it integrating different content than other platforms.

Implications

A central benefit of using AI for SBL is that educators have the ability to develop a large number of diverse case studies in a short period of time. However, there were some notable shortcomings of the platforms, including their underrepresentation of some vulnerable patient populations, limited attention to cultural sensitivity and religion in health care, and restricted patient settings. The findings have implications for educators who are interested in using AI tools for SBL within interprofessional education settings.

One of the biggest implications is that educators who generate SBL cases using generative AI must provide oversight of the content produced [36]. Research has found that content generated through AI platforms can, at times, be falsely fabricated (ie, hallucinations) or can draw on content that is outdated or biased [37]. This concern was demonstrated in this study, where evaluators felt that some content provided in cases did not adequately reflect real-world scenarios. Another example is that some vulnerable patient populations and medical conditions (such as COVID-19) were not represented in any of the 50 cases. For example, there was a complete absence of any mention of patients who identify as lesbian, gay, bisexual, transgender, queer, intersex, and asexual in any of the selected platforms. To address these shortcomings of generative AI, educators should seek training in “prompt engineering,” which refers to skills in generating prompts for AI tools that are more effective in addressing the goals of the AI user [38]. Meskó [38] indicates that prompt engineering is an important skill for health educators and health care practitioners and requires them to become knowledgeable about how individual AI platforms are designed and function so that they can develop the most effective prompts to meet their needs. In the case of health educators, prompt engineering could support them in developing AI-generated content (SBL case studies or otherwise) that better

aligns with the learning objectives they have established for their students.

The Importance of Prompt Engineering

For this study, the same general prompt was used to develop each case study. Prompt writing is recognized as a skill that can support LLM generative AI tools in tailoring responses for the user [39]. In a previous study by Rogers et al [24], the researchers evaluated the performance of ChatGPT in generating 2 case studies. However, rather than using a general prompt to generate a complete case, they used a series of prompts to guide the platform for each section of the case study (eg, presenting problem and social history). Based on findings from this study, prompt engineering skills could allow educators to more efficiently use generative AI tools that address gaps in content (eg, patient populations and conditions addressed) as well as the quality of the cases generated.

For example, in a post hoc examination, we generated an additional case study using Claude that was more specific regarding quality:

Generate a geriatric case study that can be used in simulation learning for students from various health professional disciplines, such as medical, nursing, and social work students. The content should reflect the five elements of case studies that are emphasized in the Simulation Scenario Evaluation Tool (SSET), developed by Hernandez et al [35]. For social work students, the activity should address the Educational Policy and Accreditation Standards developed by CSWE.

The resulting case study organized the content into learning objectives, clinical context, scenario overview, critical actions, patient states and progression, and scenario materials and resources. Thus, the prompt allowed Claude to build the case with established standards in mind. The resulting case study also linked content to specific Council on Social Work Education accreditation standards, such as asking social work students, “*How did you address social justice issues like financial barriers and access to care? (CSWE EPAS Competency 3)*” and “*What advocacy role did you take, and how did you respect Mrs. Martinez's right to self-determination? (CSWE EPAS Competency 2)*.”

This approach may help educators optimize the performance of AI tools in generating case studies that meet specific learning goals. Similarly, more tailored prompts could be used to generate cases featuring underrepresented patient populations or specific aspects of case studies that AI platforms may overlook (eg, ethical dilemmas, specific health conditions, or cases in less common health care settings).

Limitations

There were numerous strengths to the methodology used for this study, including the use of paid subscriptions to the AI platforms to generate cases and having 3 evaluators assess each case using a standardized assessment tool. There were also limitations. First, the case studies were all generated in November 2024, and advancements in AI platforms since then may affect current performance. Second, the team used the paid

subscriptions of each AI platform to reduce potential bias from varying access limitations across platforms. However, this may limit generalizability, as most users are more likely to access the free versions of these platforms. It is important to note that both paid and free versions of these platforms have continued to evolve since data collection, making it unclear how results would differ if the study were replicated today with either paid or free versions.

Also, some of the gaps in content that were identified by reviewers across case studies may not have been present if more specific prompts were used (eg, “*Create a case study that features a patient with COVID-19*”). It is also important to highlight that the only perspectives provided about the AI-generated case studies were those of experts in interprofessional health education. The study did not obtain perspectives from students participating in SBL activities. It also did not evaluate how the AI-generated case studies compared to case studies created by professional educators. While this analysis highlighted some of the weaknesses in the AI-generated case studies, they may still be of comparable or higher quality than those generated by humans. Future research should address these gaps.

Conclusion

The functions and accessibility of generative AI have increased dramatically over the past few years. Health care educators and providers are increasingly using these tools to create efficiencies in their work. In the case of interprofessional health education, AI can support SBL by creating a large variety of case studies in a short period of time. However, as with other applications of AI, human oversight is needed to make sure that the output from these platforms appropriately meets the learning needs of students. Although previous studies have examined AI’s potential role in SBL, this study is the first to systematically evaluate and compare multiple generative AI platforms for case study generation using a validated assessment tool (SSET). By quantifying quality differences across platforms and identifying systematic deficiencies, this research provides evidence-based guidance for educators on selecting and using AI tools effectively. The findings offer practical direction for educators navigating available generative AI tools to enhance training for health care professionals, including specific evidence-based strategies for prompt engineering that can improve the quality of SBL resources in interprofessional education. These insights enable educators to leverage AI capabilities while maintaining pedagogical rigor, ultimately supporting more efficient development of high-quality simulation training materials.

Acknowledgments

Generative AI was used to generate the case study data evaluated for this project, to assist in editing for grammar, and to assist in editing citations and formatting the content provided in [Tables 2](#) and [4](#). The data presented in the “Results” section for the cases consisted of the actual text generated by the platforms and were not content created or edited by the research team.

Funding

No external financial support or grants were received for this work.

Data Availability

A copy of all 50 case studies generated by the 5 AI platforms is available as a [Multimedia Appendix 1](#). The case studies are blinded to indicate which platform generated each case.

Conflicts of Interest

None declared.

Multimedia Appendix 1

All case studies generated by artificial intelligence (AI) and subsequently evaluated.

[[DOCX File, 446 KB - mededu_v12i1e83085_app1.docx](#)]

References

1. Kim S, Lee H, Connerton TP. How psychological safety affects team performance: mediating role of efficacy and learning behavior. *Front Psychol* 2020;11(1581):1-15 [FREE Full text] [doi: [10.3389/fpsyg.2020.01581](https://doi.org/10.3389/fpsyg.2020.01581)] [Medline: [32793037](https://pubmed.ncbi.nlm.nih.gov/32793037/)]
2. Keddington AS, Moore J. Simulation as a method of competency assessment among health care providers: a systematic review. *Nurs Educ Perspect* 2019;40(2):91-94. [doi: [10.1097/01.nep.0000000000000433](https://doi.org/10.1097/01.nep.0000000000000433)] [Medline: [30601272](https://pubmed.ncbi.nlm.nih.gov/30601272/)]
3. Sanko JS. A brief history of its use in nursing education. *Q Rev Distance Educ* 2017;18(2):77-85 [FREE Full text]
4. Zhang W, Cai M, Lee HJ, Evans R, Zhu C, Ming C. AI in medical education: global situation, effects and challenges. *Educ Inf Technol* 2023;29(4):4611-4633. [doi: [10.1007/s10639-023-12009-8](https://doi.org/10.1007/s10639-023-12009-8)]
5. Flanagan OL, Cummings KM. Standardized patients in medical education: a review of the literature. *Cureus* 2023;15(7):e42027 [FREE Full text] [doi: [10.7759/cureus.42027](https://doi.org/10.7759/cureus.42027)] [Medline: [37593270](https://pubmed.ncbi.nlm.nih.gov/37593270/)]

6. Englander R, Cameron T, Ballard AJ, Dodge J, Bull J, Aschenbrener CA. Toward a common taxonomy of competency domains for the health professions and competencies for physicians. *Acad Med* 2013;88(8):1088-1094. [doi: [10.1097/ACM.0b013e31829a3b2b](https://doi.org/10.1097/ACM.0b013e31829a3b2b)] [Medline: [23807109](#)]
7. Aebersold M. The history of simulation and its impact on the future. *AACN Adv Crit Care* 2016;27(1):56-61. [doi: [10.4037/aacnacc2016436](https://doi.org/10.4037/aacnacc2016436)] [Medline: [26909454](#)]
8. Sawaya RD, Mrad S, Rajha E, Saleh R, Rice J. Simulation-based curriculum development: lessons learnt in global health education. *BMC Med Educ* 2021;21(1):33 [FREE Full text] [doi: [10.1186/s12909-020-02430-9](https://doi.org/10.1186/s12909-020-02430-9)] [Medline: [33413346](#)]
9. Krishnan D, Keloth AV, Ubedulla S. Pros and cons of simulation in medical education: a review. *Int J Med Health Res* 2017;3(6):84-87 [FREE Full text]
10. Siew AL, Wong JW, Chan E. Effectiveness of simulated patients in geriatric education: a scoping review. *Nurse Educ Today* 2021;100:104856. [doi: [10.1016/j.nedt.2021.104856](https://doi.org/10.1016/j.nedt.2021.104856)] [Medline: [33740706](#)]
11. Weller JM, Nestel D, Marshall SD, Brooks PM, Conn JJ. Simulation in clinical teaching and learning. *Med J Aust* 2012;196(9):594. [doi: [10.5694/mja10.11474](https://doi.org/10.5694/mja10.11474)] [Medline: [22621154](#)]
12. Alanazi A, Nicholson N, Thomas S. The use of simulation training to improve knowledge, skills, and confidence among healthcare students: a systematic review. *Internet J Allied Health Sci Pract* 2017;15(3) [FREE Full text] [doi: [10.46743/1540-580X/2017.1666](https://doi.org/10.46743/1540-580X/2017.1666)]
13. Keiser MM, Turkelson C. Using simulation to evaluate clinical performance and reasoning in adult-geriatric acute care nurse practitioner students. *J Nurs Educ* 2019;58(10):599-603. [doi: [10.3928/01484834-20190923-08](https://doi.org/10.3928/01484834-20190923-08)] [Medline: [31573650](#)]
14. Mehdi Z, Ross A, Reedy G, Roots A, Ernst T, Jaye P, et al. Simulation training for geriatric medicine. *Clin Teach* 2014;11(5):387-392. [doi: [10.1111/tct.12156](https://doi.org/10.1111/tct.12156)] [Medline: [25041674](#)]
15. Braude P, Reedy G, Dasgupta D, Dimmock V, Jaye P, Birns J. Evaluation of a simulation training programme for geriatric medicine. *Age Ageing* 2015;44(4):677-682. [doi: [10.1093/ageing/afv049](https://doi.org/10.1093/ageing/afv049)] [Medline: [25953500](#)]
16. Langdalen H, Abrahamsen EB, Sollid SJM, Sørskår LIK, Abrahamsen HB. A comparative study on the frequency of simulation-based training and assessment of non-technical skills in the Norwegian ground ambulance services and helicopter emergency medical services. *BMC Health Serv Res* 2018;18(1):509 [FREE Full text] [doi: [10.1186/s12913-018-3325-1](https://doi.org/10.1186/s12913-018-3325-1)] [Medline: [29970079](#)]
17. Lateef F. Simulation-based learning: just like the real thing. *J Emerg Trauma Shock* 2010;3(4):348-352 [FREE Full text] [doi: [10.4103/0974-2700.70743](https://doi.org/10.4103/0974-2700.70743)] [Medline: [21063557](#)]
18. Bland AJ, Tobbell J. Towards an understanding of the attributes of simulation that enable learning in undergraduate nurse education: a grounded theory study. *Nurse Educ Today* 2016;44:8-13. [doi: [10.1016/j.nedt.2016.05.011](https://doi.org/10.1016/j.nedt.2016.05.011)] [Medline: [27429323](#)]
19. Lin Y, Cheng A, Hecker K, Grant V, Currie GR. Implementing economic evaluation in simulation-based medical education: challenges and opportunities. *Med Educ* 2018;52(2):150-160. [doi: [10.1111/medu.13411](https://doi.org/10.1111/medu.13411)] [Medline: [28949032](#)]
20. Glauberman G, Ito-Fujita A, Katz S, Callahan J. Artificial intelligence in nursing education: opportunities and challenges. *Hawaii J Health Soc Welf* 2023;82(12):302-305 [FREE Full text] [Medline: [38093763](#)]
21. 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](https://doi.org/10.1002/ase.2270)] [Medline: [36916887](#)]
22. Wu T, He S, Liu J, Sun S, Liu K, Han Q, et al. A brief overview of ChatGPT: the history, status quo and potential future development. *IEEE/CAA J Autom Sinica* 2023;10(5):1122-1136. [doi: [10.1109/jas.2023.123618](https://doi.org/10.1109/jas.2023.123618)]
23. 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](https://doi.org/10.3390/healthcare11060887)] [Medline: [36981544](#)]
24. Rodgers DL, Needler M, Robinson A, Barnes R, Brosche T, Hernandez J, et al. Artificial intelligence and the simulationists. *Simul Healthc* 2023;18(6):395-399. [doi: [10.1097/SIH.0000000000000747](https://doi.org/10.1097/SIH.0000000000000747)] [Medline: [37747487](#)]
25. Luo X, Tham YC, Giuffrè M, Ranisch R, Daher M, Lam K, GAMER Working Group. Reporting guideline for the use of generative artificial intelligence tools in medical research: the GAMER statement. *BMJ Evid Based Med* 2025;30(6):390-400 [FREE Full text] [doi: [10.1136/bmjebm-2025-113825](https://doi.org/10.1136/bmjebm-2025-113825)] [Medline: [40360239](#)]
26. Reporting guidelines. EQUATOR Network. 2025. URL: <https://www.equator-network.org/reporting-guidelines/> [accessed 2026-01-05]
27. Zheng L, Sheng Y, Chiang WL. Leaderboard overview. LMArena. 2024. URL: <https://lmarena.ai/leaderboard> [accessed 2025-08-13]
28. Ray PP. ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet Things Cyber-Phys Syst* 2023;3:121-154. [doi: [10.1016/j.iotcps.2023.04.003](https://doi.org/10.1016/j.iotcps.2023.04.003)]
29. Introducing Claude. Anthropic. 2023. URL: <https://www.anthropic.com/news/introducing-claude> [accessed 2025-08-13]
30. What is Claude AI? Claude AI. 2024. URL: <https://www.blockchain-council.org/ai/what-is-claude-ai/> [accessed 2025-08-13]
31. An overview of the Gemini app. Gemini. 2024. URL: <https://gemini.google/overview-gemini-app.pdf> [accessed 2025-08-13]
32. Perera PL, Lankathilake M. Preparing to revolutionize education with the multi-model GenAI tool google gemini? A journey towards effective policy making. *J Adv Educ Philos* 2023;7(08):246-253. [doi: [10.36348/jaep.2023.v07i08.001](https://doi.org/10.36348/jaep.2023.v07i08.001)]
33. Glover E. Grok: what we know about Elon Musk's AI chatbot. BuiltIn. 2024. URL: <https://builtin.com/articles/grok> [accessed 2025-08-13]

34. Spataro J. Introducing Microsoft 365 Copilot – your copilot for work. Official Microsoft Blog. 2023. URL: <https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/> [accessed 2025-08-13]
35. Hernandez J, Frallicciardi A, Nadir N, Gothard MD, Ahmed RA. Development of a Simulation Scenario Evaluation Tool (SSET): modified delphi study. BMJ Simul Technol Enhanc Learn 2020;6(6):344-350 [[FREE Full text](#)] [doi: [10.1136/bmjstel-2019-000521](https://doi.org/10.1136/bmjstel-2019-000521)] [Medline: [35515493](#)]
36. Mondal H, Mondal S. Artificial intelligence-generated content needs a human oversight. Indian J Dermatol 2024;69(3):284. [doi: [10.4103/ijd.ijd_163_24](https://doi.org/10.4103/ijd.ijd_163_24)] [Medline: [39119306](#)]
37. Fang X, Che S, Mao M, Zhang H, Zhao M, Zhao X. Bias of AI-generated content: an examination of news produced by large language models. Sci Rep 2024;14(1):5224 [[FREE Full text](#)] [doi: [10.1038/s41598-024-55686-2](https://doi.org/10.1038/s41598-024-55686-2)] [Medline: [38433238](#)]
38. Meskó B. Prompt engineering as an important emerging skill for medical professionals: tutorial. J Med Internet Res 2023;25:e50638 [[FREE Full text](#)] [doi: [10.2196/50638](https://doi.org/10.2196/50638)] [Medline: [37792434](#)]
39. Denny P, Leinonen J, Prather J, Luxton-Reilly A, Amarouche T, Becker BA, et al. Prompt problems: a new programming exercise for the generative AI era. 2024 Presented at: SIGCSE 2024: Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1; March 07, 2024; Portland, OR p. 296-302 URL: <https://doi.org/10.1145/3626252.3630909> [doi: <https://doi.org/10.1145/3626252.3630909>]

Abbreviations

AI: artificial intelligence

GAMER: Generative Artificial Intelligence Tools in Medical Research

LLM: large language model

SBL: simulation-based learning

SSET: Simulation Scenario Evaluation Tool

Edited by S Brini, T Leung; submitted 27.Aug.2025; peer-reviewed by Y Hu, H Fischle, F Lei; comments to author 27.Oct.2025; revised version received 29.Dec.2025; accepted 29.Dec.2025; published 30.Jan.2026.

Please cite as:

Ruggiano N, Sahoo S, Brashear A, Nwatu U, Brunson A, Noh H, Cole H, McKinney R, Framil Suarez CV, Brown EL, Prevost S. Evaluating AI-Generated Geriatric Case Studies for Interprofessional Education: Systematic Analysis Across 5 Platforms. JMIR Med Educ 2026;12:e83085

URL: <https://mededu.jmir.org/2026/1/e83085>

doi: [10.2196/83085](https://doi.org/10.2196/83085)

PMID:

©Nicole Ruggiano, Sudikshya Sahoo, Ava Brashear, Uche Nwatu, Amie Brunson, Hyunjin Noh, Heather Cole, Robert McKinney, C Victoria Framil Suarez, Ellen L Brown, Suzanne Prevost. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 30.Jan.2026. 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.

Digital Choice Architecture in Medical Education: Applying Behavioral Economics to Online Learning Environments

Victoria Ekstrom^{1,2}, MBBS, MRCP, Msc

¹Department of Gastroenterology and Hepatology, Singapore General Hospital, 20 College Road, Singapore, Singapore

²Medical Humanities Institute, SingHealth Duke-NUS Academic Medical Centre, Singapore, Singapore

Corresponding Author:

Victoria Ekstrom, MBBS, MRCP, Msc

Department of Gastroenterology and Hepatology, Singapore General Hospital, 20 College Road, Singapore, Singapore

Abstract

Health care has widely adopted behavioral economics to influence clinical practice, with documented success using defaults and social comparison feedback in electronic health records. However, online medical education, now the dominant modality for continuing professional development, remains designed on assumptions of rational learning that behavioral science has disproven in clinical contexts. This viewpoint examines the paradox of applying sophisticated behavioral insights to clinical work while designing digital learning environments as if learners are immune to cognitive limitations. We propose digital choice architecture for medical education: intentional integration of behavioral design principles into learning management systems and online platforms. Drawing from clinical nudge units and implementation science, we demonstrate how defaults, social norms, and commitment devices can be systematically applied to digital continuing education. As medical education becomes increasingly technology-mediated, behavioral science provides the theoretical foundation and practical tools for designing online learning environments that align with how clinicians actually make decisions.

(*JMIR Med Educ* 2026;12:e86497) doi:10.2196/86497

KEYWORDS

medical education; behavioral economics; digital learning; learning management systems; LMS; choice architecture; learning analytics; nudge theory; implementation science

Introduction

Despite a global continuing medical education (CME) market projected to reach US \$13.52 billion by 2028, continuing professional development (CPD) demonstrates remarkably limited effectiveness. A Cochrane systematic review of 215 studies with over 28,000 health care professionals found that educational meetings produce only a 6% median improvement in practice compliance and 3% improvement in patient outcomes [1]. Over 80% of CME activities rely on passive learning methods despite evidence showing that didactic presentations have little or no beneficial effect [2,3]. Health care organizations invest billions in educational technology while accepting single-digit effectiveness rates that would be unacceptable in any other clinical intervention.

The failure stems from flawed design assumptions. Current digital CPD platforms operate on an “information deficit model”—the assumption that providing knowledge changes behavior—which behavioral science has comprehensively debunked [2]. New information follows the Ebbinghaus forgetting curve, with 70% forgotten within 24 hours [3]. Physicians demonstrate poor self-assessment accuracy, with systematic reviews finding no correlation between self-assessed and measured competence in 65% of comparisons [4]. This

exposes the paradox: we design learning platforms as if clinicians are perfectly rational decision-makers while deploying behavioral interventions in clinical systems because we know they are not.

Digital platforms now dominate CPD. In 2022, accredited providers reported 230,000 total CME activities—a 13% increase—with online modules representing the most common format [5]. However, despite this technological transformation, most digital medical education assumes that clinicians will engage systematically, complete modules through intrinsic motivation, and independently translate knowledge into practice change.

Behavioral economics challenges this assumption. In clinical settings, health care organizations have proven that physician behavior responds predictably to choice architecture. When Penn Medicine redesigned electronic health record (EHR) defaults to favor generic medications, prescribing rates increased from 75% to 98% without additional education [6]. Health care nudge units have demonstrated that defaults, social comparison feedback, and environmental cues systematically shape clinical decisions [7]. These interventions succeed because clinicians, despite extensive training, exhibit status quo bias, present bias, decision fatigue, and susceptibility to social norms [8,9].

Decision fatigue alone has measurable effects: physicians prescribe antibiotics more frequently as the day progresses [10].

This paradox demands resolution. As medical education investment accelerates toward digital delivery, designing platforms on disproven assumptions wastes resources and perpetuates the knowing-doing gap. The behavioral science infrastructure that transformed clinical practice—nudge units, implementation scientists, and A/B testing—remains largely absent from educational technology development. Closing this gap requires systematic integration of behavioral economics into digital learning architecture.

This viewpoint argues that digital medical education should evolve toward behaviorally informed design, creating online learning architectures that work with human cognitive realities. We propose comprehensive integration of behavioral economics principles into digital CPD, offering both theoretical foundations and practical implementation frameworks. It is important to note that while these principles have demonstrated effectiveness in clinical contexts, their educational applications remain emerging and context-dependent, requiring systematic evaluation across diverse learning environments.

Terminology and Conceptual Clarity

Because behavioral economics terminology varies across disciplines, we establish the following definitions:

- Digital choice architecture refers to the systematic design of decision environments in digital learning platforms, encompassing defaults, interface structures, information displays, and feedback mechanisms that influence learner behavior while preserving choice. This term describes the overall design philosophy.
- Choice architecture represents the broader concept from behavioral economics, referring to how choices are organized and presented across any context [11].
- Behavioral design describes the practical application of behavioral economics and psychology principles to create systems supporting desired behaviors. We use this term when discussing the design process itself.
- Behavioral economics-informed design specifically indicates design decisions drawn from behavioral economics research, emphasizing the evidence-based grounding recommendations.
- Nudges are specific interventions that alter behavior predictably without forbidding options or significantly changing incentives [11]. Examples include defaults, social comparison feedback, and timely reminders.

These terms share conceptual overlap but maintain distinct emphases. Throughout this paper, we apply them consistently, distinguishing between broad design philosophy (choice architecture), practical implementation (behavioral design), specific interventions (nudges), and evidence-based approaches (behavioral economics-informed design).

Theoretical Foundations: Core Behavioral Economics Concepts

Behavioral economics integrates psychology and economics to understand how people actually make decisions—often deviating predictably from rational actor models. For digital medical education, these insights may explain why well-designed platforms achieve dramatically different learning outcomes than behaviorally naive alternatives, though direct empirical evidence in educational contexts remains limited.

Choice Architecture and Nudges

Choice architecture—defined by Thaler and Sunstein [11] as “organizing the context in which people make decisions”—recognizes that no choice presentation is neutral. Every interface element, default setting, and information display influences behavior. A comprehensive meta-analysis examining over 200 studies with 440+ effect sizes (N=2,148,439) found that choice architecture interventions promote behavior change with Cohen $d=0.43$, a small-to-medium effect practically significant for population-level interventions [8].

For digital learning platforms, choice architecture operates through every design decision. Whether modules appear in alphabetical order or by relevance to practice gaps shapes engagement. Whether completion tracking displays absolute progress (“3 of 10 modules”) or social comparison (“You’ve completed fewer modules than 67% of peers”) influences motivation through different behavioral mechanisms. The question is whether design reflects evidence or intuition.

Defaults and status quo bias leverage the tendency to stick with current states through multiple mechanisms: loss aversion, cognitive effort minimization, and implied endorsement [11]. Johnson and Goldstein [12] found that the organ donation consent rates were approximately twice as high in opt-out versus opt-in countries. Platforms that auto-enroll clinicians in modules addressing documented practice gaps leverage defaults while preserving opt-out capability.

Present bias and temporal discounting explain persistent procrastination. Clinicians understand CPD improves patient care, but immediate opportunity costs outweigh psychologically distant benefits. Laibson’s [13] quasi-hyperbolic discounting model demonstrates that people systematically prefer smaller immediate rewards over larger delayed rewards. On the basis of these findings, we propose that digital platforms may counteract this through temporal optimization—delivering microlearning at the point of clinical relevance, which minimizes the gap between learning effort and application reward.

Decision fatigue and cognitive depletion have measurable effects on clinical judgment. As noted, physicians prescribe antibiotics more frequently as the day progresses [10]. Judicial rulings decline from 65% favorable at session start to near zero before breaks, returning to baseline after rest [14]. For digital learning, this suggests that platforms should minimize unnecessary choices that deplete cognitive resources rather than exhausting capacity on navigation decisions.

Social norms and social comparison powerfully influence behavior. Social comparison feedback showing how behavior compares to peers consistently drives behavior change across clinical contexts. Digital dashboards displaying peer prescribing comparisons reduce inappropriate antibiotic use [15]. On the basis of the available evidence, we propose that learning dashboards could leverage similar mechanisms, though whether clinical effects translate to educational outcomes requires systematic investigation. Studies of collaborative learning modules show that embedding social cues could enhance motivation and learning clarity [16], suggesting that digital platforms could achieve similar benefits through discussion forums and peer-review features integrated into the learning management systems (LMSs).

Integration With Complementary Theories

Behavioral economics does not operate in isolation. Cognitive load theory (CLT) provides mechanistic insights into working memory limitations—working memory can process only 4 to 7 information chunks simultaneously [17,18]. CLT distinguishes between intrinsic (inherent task difficulty), extraneous (imposed by suboptimal presentation), and germane (devoted to learning) cognitive loads. Choice architecture principles directly reduce extraneous cognitive load, freeing capacity for germane processing. The AMEE (Association for Medical Education in Europe) Guide No. 86 establishes CLT’s “particular relevance to medical education because many professional activities require simultaneous integration of multiple knowledge sets” [19].

Self-determination theory identifies 3 psychological needs essential for intrinsic motivation: autonomy, competence, and relatedness [20]. Williams et al [21] demonstrated that “autonomously motivated learning leads to better educational outcomes.” This creates productive tension with nudging approaches, as defaults can reduce perceived locus of causality. However, autonomy-preserving designs can reconcile these tensions through transparent nudges that explain rationale, self-nudges where learners choose their own interventions, and “boosts” that build competence rather than merely steering behavior. Hertwig and Grüne-Yanoff [17] distinguish boosting—which aims to “foster people’s competence to make their own choices”—from nudging, arguing that boosts create lasting cognitive “capital stock” and necessarily preserve autonomy.

These complementary frameworks converge on similar design principles. Behavioral economics explains why interventions work (cognitive biases and heuristics), CLT specifies how much information learners can process, and self-determination theory identifies what motivates sustained engagement.

Digital Transformation and Clinical Lessons

The shift from classroom to technology-mediated education has fundamentally altered how clinicians access professional development. LMSs track every interaction—login times, completion rates, assessment attempts, and module abandonment points. Mobile apps deliver microlearning during clinical

workflows. Artificial intelligence algorithms personalize content sequences. This digital infrastructure generates unprecedented behavioral data while creating new decision points where choice architecture can be applied. Every interface element implicitly shapes learning behavior. The question is not whether digital platforms incorporate choice architecture, but whether they do so intentionally and based on evidence.

Clinical decision support systems demonstrate what behaviorally informed digital design can achieve. EHRs now routinely incorporate default order sets favoring evidence-based practices. Digital dashboards provide peer comparison feedback on prescribing patterns. Predictive analytics trigger just-in-time reminders when clinical decisions deviate from best practice. A systematic review found robust evidence that digital defaults and social proof could meaningfully change a physician’s behavior across multiple domains [21]. Notably, these effects persist among experienced practitioners—behavioral tendencies are not eliminated by expertise. Digital choice architecture in clinical systems represents the convergence of behavioral science with technological capability, creating decision environments that systematically support better choices at scale.

Behavioral Design Gap in Digital Medical Education

Applications of behavioral economics to online medical education exist but remain limited. Emerging evidence demonstrates both promise and important limitations.

Gurley et al [22] found that email reminders increased residency assessment completion rates, though the intervention required repeated prompts and the effect sizes were modest. More comprehensive interventions showed mixed results at scale. Kizilcec et al’s [23] landmark study tested 5 behavioral interventions across 269,169 students in 247 massive open online courses. Interventions showing effectiveness at a small scale (planning prompts: +29% completion for committed students) demonstrated minimal impact when implemented broadly. Only value-relevance interventions showed modest improvements for disadvantaged students in courses with existing achievement gaps. This large-scale evidence reveals critical constraints: behavioral interventions may require targeting and personalization, and what works in controlled trials may not translate to population-level deployment.

More promising results emerge from targeted interventions. A 2023 study combining learning analytics with behavioral nudges found personalized reminders significantly increased student resource clicks (planning courses: 415 to 778 clicks, $P<.001$; education courses: 304-446 clicks, $P<.001$) [24]. Research on tailored digital nudges matched to students’ motivation profiles showed improved engagement when interventions aligned with learner characteristics [25].

Medical education-specific applications remain rare. Van Gaalen et al’s [26] systematic review found that gamification in health professional education showed no negative outcomes, though most studies were descriptive with few using theory or well-defined control groups. Kerfoot et al’s [27] high-quality randomized controlled trial with 141 medical students found

that gamified and spaced learning improved hypertension management knowledge and produced modest reductions in days to reach target blood pressure—demonstrating transfer to patient outcomes.

Recent evidence from anatomy education demonstrated how behavioral principles could be successfully embedded. Specialized study modules (SSMs) incorporating intentional pathway design, collaborative learning structures, and clear learning objectives produced significant improvements in student motivation, teamwork, and clarity of learning goals [16,28]. These findings parallel mechanisms that digital choice architecture seeks to operationalize: both approaches rely on well-designed pathways that make desired behaviors easier, clearer, and more rewarding. Just as SSMs improved learning through structured guidance, digital platforms can leverage defaults, social nudges, and collaborative cues to streamline learning decisions and reduce friction. The observed heterogeneity in learning strategies across educational contexts [16] reinforces the need for adaptive nudges and personalized digital pathways accommodating diverse learner preferences.

Systematic reviews note that education-related studies represent only 4% of choice architecture interventions tested across all domains [29], with learning analytics in medical education “still in infancy” with “systematic applications limited” [30]. Platforms rarely use predictive analytics to identify at-risk learners or deploy targeted nudges. Research demonstrates that only half of people who form intentions actually take action [31], highlighting the critical intention-behavior gap. The majority of accredited CPD activities do not target clinical behavior change [32], with behavior change techniques appearing in only a subset of CME [32].

Considering typical design patterns in online CPD platforms, the default pathway after completing a module is returning to the dashboard—no implementation planning, no commitment to practice change, and no follow-up accountability. Courses are structured as optional selections requiring active search rather than defaults based on practice gaps. Progress tracking emphasizes completion percentages without peer comparison. Educational content is delivered months before clinical application, despite evidence that temporal distance undermines behavior change. The field that transformed clinical practice through digital behavior change has barely touched the digital systems used to educate clinicians.

A Framework for Digital Choice Architecture in Online Medical Education

Digital choice architecture for medical education means intentionally designing online learning environments using behavioral principles proven effective in clinical contexts. We propose a comprehensive framework integrating established implementation science models with practical design principles. It is critical to emphasize that educational applications of these principles remain emerging, requiring context-specific adaptation and rigorous evaluation.

Established Implementation Frameworks

Multiple frameworks can guide the systematic application of behavioral insights to educational platforms. The UK Behavioral Insights Team’s EAST (easy, attractive, social, and timely) framework provides the following accessible design heuristics [33] as : Make it Easy (reduce effort through defaults, simplified messages, and minimal “hassle factors”), Make it Attractive (use visual design, immediate rewards, and salient feedback), Make it Social (leverage peer comparison, network effects, and commitment to others), and Make it Timely (deliver prompts when learners are most receptive).

The COM-B (capability, opportunity, motivation, and behavior) model provides diagnostic capacity by conceptualizing behavior as arising from the interaction of 3 core components: capability (physical and psychological), opportunity (environmental and social enablers), and motivation (reflective and automatic processes) [34]. This diagnostic approach links to the behavior change wheel’s 9 intervention functions, enabling systematic selection of appropriate behavioral strategies based on identified barriers. The theoretical domains framework extends analysis across 14 granular domains, including knowledge, skills, professional identity, and beliefs about capabilities [35]. The Behavioral Insights Team’s 4-step method provides a cyclical process: define the outcome, understand the context, build your intervention, and test or learn or adapt [33].

Stepwise Design Principles for LMS Platforms

Translating frameworks into practice requires systematic steps.

Step 1: Audit Current Choice Architecture

Before implementing behavioral design, organizations must document existing implicit architecture. This includes mapping all default settings. For example, what happens when learners complete a module? What enrollment pathways exist? Organizations should identify every decision point where learners make choices, catalog implicit assumptions embedded in platform design, and establish baseline behavioral metrics including completion rates, time-to-completion, drop-off points, and peak engagement times. This audit reveals where behavioral insights can add value while providing baseline data for measuring improvement.

Step 2: Apply Behavioral Design Principles

Defaults leverage status quo bias. Auto-enrollment in required modules based on practice type or documented knowledge gaps preserves choice while making evidence-based learning the path of least resistance. Research demonstrates that providing clear, guided pathways—similar to defaults—improves motivation and learning clarity while reducing cognitive burden [16,28]. Evidence-based pathway before selection prioritizes high-impact content. Postmodule defaults should route learners to implementation planning tools rather than simply returning to catalogs. Assessment scheduling can prepopulate optimal spacing intervals. Each default preserves learner autonomy through transparent opt-out.

Friction reduction addresses that every additional click creates abandonment opportunity. One-click enrollment minimizes barriers. Single sign-on eliminates authentication friction.

Streamlined navigation reduces unnecessary steps. Form simplification eliminates nonessential fields. These micro-optimizations cumulatively reduce extraneous cognitive load.

Social comparison and peer feedback leverage descriptive norms without public shaming. Learning dashboards can display “You have completed 60% of required CPD; 80% of peers in your specialty have completed more.” Anonymized leaderboards show rankings for voluntary activities. Specialty-specific benchmarks compare engagement to relevant peer groups. Completion milestones highlight social norms: “Join the 73% of cardiologists who completed this module.” Peer-supported learning structures enhance motivation and learning clarity [16], suggesting that platforms can achieve benefits through discussion forums, peer-review features, and collaborative assignments integrated into LMSs.

Timely prompts and momentum nudges address present bias. Deadline reminders sent 3 days before, 1 day before, and on the day counteract procrastination. Momentum messages like “You’re 80% through this module—finish now?” capitalize on progress. Peak engagement timing delivers notifications at moments when individual users are most likely to engage, based on their historical behavior. Calendar integration schedules protected learning time directly into clinicians’ calendars.

Personalization based on behavioral data increases relevance while reducing cognitive load. Role-based recommendations curate content by specialty. Practice-gap-informed suggestions link learning to quality metrics. Adaptive difficulty adjusts complexity based on assessment performance. Temporal optimization delivers notifications when individual learners historically show the highest engagement.

Step 3: Implement With Transparency

Ethical behavioral design requires disclosure and opt-out mechanisms. Platform documentation should clearly explain that the system uses behavioral design principles. Opt-out mechanisms must allow users to disable social comparison features, modify notification frequency, or change default pathways. Transparent nudges make intent reconstructable—for example, “This module was recommended based on your practice gaps in diabetes management.” All behavioral design choices should be documented and reviewed through appropriate oversight mechanisms. Research shows transparent nudges remain effective [36,37].

Step 4: Measure and Iterate

Continuous measurement enables data-driven improvement. Organizations should track engagement metrics (logins, completion rates, time-on-platform, and drop-off points), learning outcomes (assessment scores and knowledge retention), and practice change where feasible (quality metrics and chart audits). A/B testing allows systematic comparison of behavioral interventions. User feedback through surveys and qualitative interviews provides insight into learner experience. On the basis of these data, organizations should refine interventions quarterly, abandoning approaches that do not work and scaling those that do.

Digital Commitment Devices and Implementation Intentions

The intention-behavior gap represents a critical barrier. Research demonstrates that only half of the people who form intentions actually take action [31]. Digital platforms can incorporate commitment mechanisms grounded in implementation intention research, which shows medium-to-large effect sizes (Cohen $d=0.65$) when individuals create specific if-then plans [38]. Practical applications include implementation planning prompts after module completion requiring learners to specify “When will you apply this? In what clinical situation?” Calendar commitments directly schedule follow-up learning or application time blocks. Automated accountability check-ins follow up, asking “Did you apply what you learned? What barriers did you encounter?”

Implementation Roadmap for Health Care Organizations

Health care organizations can follow these phased approaches.

Phase 1: Assessment and Planning (Months 1-3)

Phase 1 includes the following:

1. Learning analytics specialists conduct comprehensive choice architecture audits documenting all platform defaults, decision points, and implicit assumptions.
2. Learner surveys identify motivation barriers and preferences while behavioral analytics reveal completion patterns and peak engagement times.
3. Deliverables include current state reports, priority intervention lists ranked by impact-effort ratio, baseline metrics dashboards, and stakeholder engagement plans. This requires a learning analytics specialist, behavioral scientist consultant (0.2 - 0.5 full-time equivalents), platform administrator, end-user clinicians, and an instructional designer.

Phase 2: Pilot Interventions (Months 4-9)

Phase 2 consists of the following:

1. Implement 3 to 5 behavioral design changes in a limited scope, typically targeting a single department.
2. Conduct A/B testing where feasible. Measurement includes quantitative metrics (completion rates, time-to-completion, and assessment performance) and qualitative methods (user feedback and focus groups). Example pilots: changing default postmodule action to “create implementation plan,” adding peer comparison feedback to learner dashboards, implementing smart reminders based on individual engagement patterns, testing auto-enrollment versus opt-in, or adding commitment devices.
3. Track completion rates, time-to-completion, assessment performance, user satisfaction, opt-out rates, and qualitative feedback themes.

Phase 3: Scale and Sustain (Months 10+)

Phase 3 consists of the following:

1. Roll out successful interventions system-wide.

2. Build ongoing A/B testing infrastructure.
3. Train staff in behavioral design principles.
4. Establish governance structures for ethical oversight.
5. Integrate behavioral metrics into routine quality improvement cycles.
6. Share results through publications and presentations.
7. Sustainability requires behavioral science expertise (0.5 - 1.0 FTE), learning analytics infrastructure, regular review cycles, ethical oversight committees, and continuous professional development.

Resource Requirements and Common Pitfalls

Resource requirements scale with organizational size. Small organizations serving fewer than 5000 learners typically invest US \$50,000 to US \$100,000 annually, including part-time behavioral science consultation, analytics platform subscriptions, and platform customization. Medium organizations serving 5000 to 50,000 learners typically invest US \$200,000 to US \$400,000 annually, including full-time behavioral science expertise, dedicated analytics platforms, testing infrastructure, and platform development. Large organizations serving over 50,000 learners typically invest US \$500,000 to US \$1 million or more annually, including behavioral science teams, advanced analytics platforms, dedicated nudge unit infrastructure, and platform development teams. These figures represent ongoing investments necessary for effective educational delivery in the digital age.

Organizations should avoid common implementation pitfalls. Overcomplication from starting with too many simultaneous interventions makes isolating effects impossible; organizations should begin with 2 to 3 high-priority changes, demonstrate value, and then expand. Metrics myopia from optimizing purely for completion rates without monitoring quality or satisfaction can lead to gaming behaviors; use balanced scorecards instead. Transparency failures from implementing behavioral interventions without disclosure violate ethical principles and risk backlash; always explain design rationale up front. Ethical shortcuts from rushing implementation without ethical review create institutional risks; establish oversight early. Siloed work with behavioral science operating independently from instructional design creates a fragmented user experience; integrate from the start. Technology determinism assumes that building features automatically change behavior without attention to psychology; focus on behavioral science first and technology second. Insufficient testing by scaling interventions before demonstrating effectiveness wastes resources; pilot rigorously before scaling. Ignoring heterogeneity by assuming all learners respond identically misses personalization opportunities; conduct segment analyses and adapt interventions accordingly.

Research Agenda

Translating behavioral economics from clinical to educational digital systems requires systematic investigation. Key priorities include conducting comprehensive audits of existing online CPD platforms to document current choice architecture, cataloging defaults and behavioral assumptions. Randomized controlled trials should compare traditional digital learning

designs with behaviorally optimized alternatives, measuring not just completion but practice change outcomes. Implementation frameworks need development to provide practical guidance for educational institutions.

The key research questions are as follows:

1. Do behavioral interventions remain effective in clinical EHRs transfer to educational LMS contexts with similar effect sizes?
2. How does professional identity moderate responses to digital educational nudges?
3. What is the optimal balance between algorithmic personalization and learner autonomy?
4. How do clinicians respond to transparent disclosure that platforms incorporate behavioral design?

Technical implementation requires interdisciplinary collaboration [39]. Educational technologists must partner with behavioral scientists to translate abstract principles into specific interface elements. Data scientists need frameworks for analyzing learning analytics through a behavioral lens. Software developers require training in ethical behavioral design to avoid dark patterns that maximize compliance at the expense of learning quality or user autonomy.

Ethical Implications of Behavioral Design in Digital Learning

Applying behavioral economics to digital education raises legitimate concerns about manipulation and autonomy. Well-designed nudges can preserve freedom of choice while supporting goal achievement [40]. The distinction between facilitation and manipulation lies in transparency, alignment with learner objectives, and preservation of autonomy.

The Hansen and Jespersen Framework

Hansen and Jespersen [36] provide essential ethical distinctions through a 2D typology. Nudges vary along cognitive processing (type 1 automatic vs type 2 reflective) and transparency (transparent vs nontransparent intent and mechanism). Transparent type 2 nudges engage reflective processing while making intent and mechanism clear—these “do not comprise the unethical aspect of manipulating or secretly influencing individuals’ behavior” and represent the gold standard [37]. For example, “This module was recommended based on your documented practice gaps in diabetes management. View criteria and opt out if not relevant.” Transparent type 1 nudges operate through automatic processing but are disclosed. For example, “We’ve pre-selected the most commonly chosen learning path for your specialty. You can customize at any time.” Nontransparent nudges raise manipulation concerns and are generally impermissible in educational contexts.

Supportive Versus Coercive Design

Supportive nudges preserve choice through easy, clear opt-out mechanisms. They maintain transparent intent, align with learners’ stated goals, build capability alongside behavioral steering, and respect autonomy. In contrast, coercive manipulation obscures alternatives or makes opt-out difficult. It uses hidden intent, optimizes for platform engagement metrics

rather than learner welfare, and uses dark patterns such as fake urgency, hidden costs, forced continuity, or obstruction.

Research decisively shows that transparent nudges are not less effective than nontransparent alternatives [36,37]. This evidence suggests that transparency serves as an “ethical filter” without sacrificing outcomes, meaning platforms can—and should—openly explain their behavioral design principles while maintaining effectiveness.

Ethical Decision Framework: 5-Step Process

Organizations implementing digital choice architecture should apply the following framework before deploying any intervention:

1. Is the nudge transparent? Can learners reconstruct the intent?
2. Does it engage reflective processing (type 2)? Prefer interventions that prompt conscious consideration.
3. Is opt-out easy and clear? Can learners disable the intervention in 3 clicks or fewer?
4. Does it align with learner objectives? Is the intervention designed to help learners achieve their stated goals?
5. Does it build competence alongside behavior change? Does the intervention include educational components explaining why the behavior is beneficial?

If an intervention fails any criteria, it should be redesigned or rejected.

Dark Patterns: Explicitly Prohibited

Educational platforms must avoid deceptive design practices including fake scarcity (“Only 3 spots left!” when capacity is unlimited), hidden costs, forced continuity, obstruction (making cancellation unreasonably difficult), misdirection, nagging (excessive notifications), and sneaking (adding items without consent). These practices violate trust, undermine autonomy, and have no place in professional education.

The Future of Behaviorally Informed Digital Medical Education

The integration of behavioral economics into digital medical education represents a paradigm shift from information-deficit models to evidence-based decision support. As artificial intelligence and adaptive learning technologies mature, behavioral design principles will become increasingly critical. Machine learning algorithms can personalize nudges based on individual response patterns, learning analytics can identify optimal timing for interventions, and natural language processing can generate personalized feedback. However, technological sophistication must be matched by ethical sophistication. The more powerful these tools become, the more important transparency and oversight become.

Looking forward, digital choice architecture may extend beyond traditional CPD to include just-in-time clinical decision support that simultaneously educates and guides practice, simulation-based training environments that use defaults and social comparison to accelerate skill acquisition, and longitudinal professional development pathways that adapt to

career stages and practice evolution. The convergence of behavioral science, educational technology, and implementation science offers unprecedented opportunity to close the knowing-doing gap that has plagued medical education for decades.

Limitations and Future Research

This viewpoint has important limitations. First, we extrapolate from clinical behavioral interventions to educational contexts without comprehensive evidence that effect sizes transfer. Educational effects of behavioral design principles are inferred from clinical contexts and remain context-dependent, requiring systematic empirical validation across diverse learning environments before widespread adoption. The Kizilcec et al’s [23] study demonstrates that small-scale educational successes may not replicate at the population level, suggesting caution in assuming that clinical behavioral economics directly translates to educational platforms. Second, we focus primarily on CPD rather than undergraduate or postgraduate training, where different considerations may apply. Third, ethical frameworks proposed require empirical validation—we need evidence on how transparent nudges affect trust, engagement, and learning outcomes specifically in medical education contexts.

Fourth, there exists a risk of overoptimization or technocratic control, where excessive focus on measurable behavioral metrics (completion rates and click-through rates) may overshadow harder-to-quantify educational outcomes such as critical thinking, professional identity formation, or intrinsic motivation. Behavioral design could inadvertently create systems that optimize for compliance rather than meaningful learning.

Fifth, resistance from clinicians who value professional autonomy represents a significant implementation barrier. Physicians may perceive behavioral interventions—even transparent ones—as patronizing or as threats to professional independence. This resistance may be particularly strong among senior clinicians with established practice patterns who view educational decisions as matters of professional judgment rather than targets for behavioral modification.

Sixth, institutional constraints pose substantial challenges. These include cost barriers (significant upfront investment in behavioral science expertise, analytics infrastructure, and platform redesign), governance complexities (establishing ethical oversight, determining acceptable intervention boundaries, and managing stakeholder concerns), and legacy system limitations (many organizations operate on established LMS platforms with limited customization capabilities, vendor dependencies, and lengthy procurement cycles that impede rapid iteration). Health care organizations must navigate these practical constraints while attempting implementation.

Future research should prioritize large-scale randomized trials comparing behaviorally optimized platforms to traditional designs across multiple outcomes (completion, learning, satisfaction, and practice change). Studies should examine heterogeneity in treatment effects—which clinician subgroups respond most strongly to which interventions? Implementation science research should document barriers and facilitators to

adopting behavioral design in health care organizations. Ethical research should examine learner perceptions of nudges, optimal disclosure methods, and long-term effects on professional development autonomy.

Conclusions

The question facing medical education is not whether to design learning environments but whether to design them well or badly. Every digital interface embodies choice architecture—defaults, menu structures, notification timing, and progress displays—that shapes learning behavior. Currently, most platforms incorporate architecture based on intuition rather than evidence. This represents a missed opportunity of staggering proportions.

Health care organizations should conduct choice architecture audits of existing CPD platforms, identifying defaults and decision points; pilot behavioral interventions with rigorous

measurement before scaling; and partner with behavioral scientists during platform design, not as an afterthought. Academic medical centers should deploy implementation science frameworks to systematically translate behavioral insights into educational practice. Accreditation bodies should incentivize behavioral design through quality metrics that reward practice change, not just completion rates. Vendors should partner with behavioral scientists from the design stage, building behavioral principles into product architecture. Professional societies should develop best practice guidelines for ethical behavioral design in medical education, facilitate sharing through communities of practice, and advocate for research funding.

The evidence base exists. The tools are available. The imperative is urgent. The path is clear. The question is whether we will take it.

Acknowledgments

The author thanks colleagues at the SingHealth Duke-NUS Medical Humanities Institute for their feedback on earlier drafts.

Authors' Contributions

VE is the sole author and conceived, researched, and wrote this paper.

Conflicts of Interest

None declared.

References

1. Forsetlund L, O'Brien MA, Forsén L, et al. Continuing education meetings and workshops: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2021 Sep 15;9(9):CD003030. [doi: [10.1002/14651858.CD003030.pub3](https://doi.org/10.1002/14651858.CD003030.pub3)] [Medline: [34523128](#)]
2. Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 1999 Oct 20;282(15):1458-1465. [doi: [10.1001/jama.282.15.1458](https://doi.org/10.1001/jama.282.15.1458)] [Medline: [10535437](#)]
3. Murre JMJ, Dros J. Replication and analysis of Ebbinghaus' forgetting curve. *PLoS One* 2015;10(7):e0120644. [doi: [10.1371/journal.pone.0120644](https://doi.org/10.1371/journal.pone.0120644)] [Medline: [26148023](#)]
4. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA* 2006 Sep 6;296(9):1094-1102. [doi: [10.1001/jama.296.9.1094](https://doi.org/10.1001/jama.296.9.1094)] [Medline: [16954489](#)]
5. ACCME data report: renewal and growth in accredited continuing education – 2022. : Accreditation Council for Continuing Medical Education; 2022 URL: https://accme.org/wp-content/uploads/2024/05/2022-ACCME-Annual-Data-Report-1003_20230808.pdf?utm_source=chatgpt.com [accessed 2026-01-24]
6. Patel MS, Day SC, Halpern SD, et al. Generic medication prescription rates after health system-wide redesign of default options within the electronic health record. *JAMA Intern Med* 2016 Jun 1;176(6):847-848. [doi: [10.1001/jamainternmed.2016.1691](https://doi.org/10.1001/jamainternmed.2016.1691)] [Medline: [27159011](#)]
7. Patel MS, Volpp KG, Asch DA. Nudge units to improve the delivery of health care. *N Engl J Med* 2018 Jan 18;378(3):214-216. [doi: [10.1056/NEJMmp1712984](https://doi.org/10.1056/NEJMmp1712984)] [Medline: [29342387](#)]
8. Mertens S, Herberz M, Hahnel UJJ, Brosch T. The effectiveness of nudging: a meta-analysis of choice architecture interventions across behavioral domains. *Proc Natl Acad Sci U S A* 2022 Jan 4;119(1):e2107346118. [doi: [10.1073/pnas.2107346118](https://doi.org/10.1073/pnas.2107346118)] [Medline: [34983836](#)]
9. Samuelson W, Zeckhauser R. Status quo bias in decision making. *J Risk Uncertainty* 1988 Mar;1(1):7-59. [doi: [10.1007/BF00055564](https://doi.org/10.1007/BF00055564)]
10. Linder JA, Doctor JN, Friedberg MW, et al. Time of day and the decision to prescribe antibiotics. *JAMA Intern Med* 2014 Dec;174(12):2029-2031. [doi: [10.1001/jamainternmed.2014.5225](https://doi.org/10.1001/jamainternmed.2014.5225)] [Medline: [25286067](#)]
11. Thaler RH, Sunstein CR. Nudge: Improving Decisions About Health, Wealth, and Happiness: Yale University Press; 2008. URL: <https://books.google.co.in/books/about/Nudge.html?id=dSJQn8egXvUC> [accessed 2026-01-28]

12. Johnson EJ, Goldstein DG. Medicine. Do defaults save lives? *Science* 2003 Nov 21;302(5649):1338-1339. [doi: [10.1126/science.1091721](https://doi.org/10.1126/science.1091721)] [Medline: [14631022](#)]
13. Laibson D. Golden eggs and hyperbolic discounting. *Q J Econ* 1997 May 1;112(2):443-478. [doi: [10.1162/003355397555253](https://doi.org/10.1162/003355397555253)]
14. Danziger S, Levav J, Avnaim-Pesso L. Extraneous factors in judicial decisions. *Proc Natl Acad Sci U S A* 2011 Apr 26;108(17):6889-6892. [doi: [10.1073/pnas.1018033108](https://doi.org/10.1073/pnas.1018033108)] [Medline: [21482790](#)]
15. Meeker D, Linder JA, Fox CR, et al. Effect of behavioral interventions on inappropriate antibiotic prescribing among primary care practices: a randomized clinical trial. *JAMA* 2016 Feb 9;315(6):562-570. [doi: [10.1001/jama.2016.0275](https://doi.org/10.1001/jama.2016.0275)] [Medline: [26864410](#)]
16. Ogut E, Yildirim FB, Senol Y, Senol AU. Comprehensive evaluation of the educational impact and effectiveness of specialized study modules in cross-sectional anatomy: a study on student engagement and learning outcomes. *BMC Med Educ* 2025 Apr 10;25(1):514. [doi: [10.1186/s12909-025-07050-9](https://doi.org/10.1186/s12909-025-07050-9)] [Medline: [40211255](#)]
17. Hertwig R, Grüne-Yanoff T. Nudging and boosting: steering or empowering good decisions. *Perspect Psychol Sci* 2017 Nov;12(6):973-986. [doi: [10.1177/1745691617702496](https://doi.org/10.1177/1745691617702496)] [Medline: [28792862](#)]
18. Blumenthal-Barby JS, Krieger H. Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy. *Med Decis Making* 2015 May;35(4):539-557. [doi: [10.1177/0272989X14547740](https://doi.org/10.1177/0272989X14547740)] [Medline: [25145577](#)]
19. Young JQ, Van Merriënboer J, Durning S, Ten Cate O. Cognitive load theory: Implications for medical education: AMEE Guide No. 86. *Med Teach* 2014 May;36(5):371-384. [doi: [10.3109/0142159X.2014.889290](https://doi.org/10.3109/0142159X.2014.889290)] [Medline: [24593808](#)]
20. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am Psychol* 2000;55(1):68-78. [doi: [10.1037/0003-066X.55.1.68](https://doi.org/10.1037/0003-066X.55.1.68)] [Medline: [11392867](#)]
21. Williams GC, Saizow RB, Ryan RM. The importance of self-determination theory for medical education. *Acad Med* 1999 Sep;74(9):992-995. [doi: [10.1097/00001888-199909000-00010](https://doi.org/10.1097/00001888-199909000-00010)] [Medline: [10498090](#)]
22. Gurley A, Jenkins C, Nguyen T, Woodall A, An J. Nudge theory: effectiveness in increasing emergency department faculty completion of residency assessments. *West J Emerg Med* 2024 Jan;25(1):40-42. [doi: [10.5811/westjem.57721](https://doi.org/10.5811/westjem.57721)] [Medline: [38205983](#)]
23. Kizilcec RF, Reich J, Yeomans M, et al. Scaling up behavioral science interventions in online education. *Proc Natl Acad Sci U S A* 2020 Jun 30;117(26):14900-14905. [doi: [10.1073/pnas.1921417117](https://doi.org/10.1073/pnas.1921417117)] [Medline: [32541050](#)]
24. Brown A, Basson M, Axelsen M, Redmond P, Lawrence J. Empirical evidence to support a nudge intervention for increasing online engagement in higher education. *Educ Sci* 2023;13(2):145. [doi: [10.3390/educsci13020145](https://doi.org/10.3390/educsci13020145)]
25. Plak S, Cornelisz I, Meeter M, van Klaveren C. Early warning systems for more effective student counselling in higher education: evidence from a Dutch field experiment. *High Educ Q* 2022 Jan;76(1):131-152. [doi: [10.1111/hequ.12298](https://doi.org/10.1111/hequ.12298)]
26. van Gaalen AEJ, Brouwer J, Schönrock-Adema J, Bouwkamp-Timmer T, Jaarsma ADC, Georgiadis JR. Gamification of health professions education: a systematic review. *Adv in Health Sci Educ* 2021 May;26(2):683-711. [doi: [10.1007/s10459-020-10000-3](https://doi.org/10.1007/s10459-020-10000-3)]
27. Kerfoot BP, Baker H, Pangaro L, et al. An online spaced-education game to teach and assess medical students: a multi-institutional prospective trial. *Acad Med* 2012 Oct;87(10):1443-1449. [doi: [10.1097/ACM.0b013e318267743a](https://doi.org/10.1097/ACM.0b013e318267743a)] [Medline: [22914524](#)]
28. Bhatti F, Mowforth O, Butler M, et al. Meeting the shared goals of a student-selected component: pilot evaluation of a collaborative systematic review. *JMIR Med Educ* 2023 Mar 15;9:e39210. [doi: [10.2196/39210](https://doi.org/10.2196/39210)] [Medline: [36920459](#)]
29. Szaszi B, Palinkas A, Palfi B, Szollosi A, Aczel B. A systematic scoping review of the choice architecture movement: toward understanding when and why nudges work. *J Behav Decis Mak* 2018 Jul;31(3):355-366. [doi: [10.1002/bdm.2035](https://doi.org/10.1002/bdm.2035)]
30. Bojic I, Mammadova M, Ang CS, et al. Empowering health care education through learning analytics: in-depth scoping review. *J Med Internet Res* 2023 May 17;25:e41671. [doi: [10.2196/41671](https://doi.org/10.2196/41671)] [Medline: [37195746](#)]
31. Gollwitzer PM, Sheeran P. Implementation intentions and goal achievement: a meta - analysis of effects and processes. *Adv Exp Soc Psychol* 2006;38:69-119. [doi: [10.1016/S0065-2601\(06\)38002-1](https://doi.org/10.1016/S0065-2601(06)38002-1)]
32. Légaré F, Freitas A, Thompson-Leduc P, et al. The majority of accredited continuing professional development activities do not target clinical behavior change. *Acad Med* 2015 Feb;90(2):197-202. [doi: [10.1097/ACM.0000000000000543](https://doi.org/10.1097/ACM.0000000000000543)] [Medline: [25354076](#)]
33. EAST: four simple ways to apply behavioural insights. : The Behavioural Insights Team; 2014 URL: https://www.behaviourchange.net/docs/bit-publication-east_fa_web.pdf [accessed 2026-01-24]
34. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement Sci* 2011 Apr 23;6:42. [doi: [10.1186/1748-5908-6-42](https://doi.org/10.1186/1748-5908-6-42)] [Medline: [21513547](#)]
35. Atkins L, Francis J, Islam R, et al. A guide to using the theoretical domains framework of behaviour change to investigate implementation problems. *Implement Sci* 2017 Jun 21;12(1):77. [doi: [10.1186/s13012-017-0605-9](https://doi.org/10.1186/s13012-017-0605-9)] [Medline: [28637486](#)]
36. Hansen PG, Jespersen AM. Nudge and the manipulation of choice: a framework for the responsible use of the nudge approach to behaviour change in public policy. *Eur J Risk Regul* 2013;4(1):3-28. [doi: [10.1017/S1867299X00002762](https://doi.org/10.1017/S1867299X00002762)]
37. Gollwitzer PM. Implementation intentions: strong effects of simple plans. *Am Psychol* 1999;54(7):493-503. [doi: [10.1037/0003-066X.54.7.493](https://doi.org/10.1037/0003-066X.54.7.493)]

38. Bieleke M, Keller L, Gollwitzer PM. If-then planning. *Eur Rev Soc Psychol* 2021 Jan 2;32(1):88-122. [doi: [10.1080/10463283.2020.1808936](https://doi.org/10.1080/10463283.2020.1808936)]
39. Hodson N, Powell BJ, Nilsen P, Beidas RS. How can a behavioral economics lens contribute to implementation science? *Implement Sci* 2024 Apr 26;19(1):33. [doi: [10.1186/s13012-024-01362-y](https://doi.org/10.1186/s13012-024-01362-y)] [Medline: [38671508](https://pubmed.ncbi.nlm.nih.gov/38671508/)]
40. Sunstein CR. The ethics of nudging. *Yale J Regul* 2015;32(2):413-450 [[FREE Full text](#)]

Abbreviations

AMEE: Association for Medical Education in Europe

CLT: cognitive load theory

CME: continuing medical education

COM-B: capability, opportunity, motivation, and behavior

CPD: continuing professional development

EAST: easy, attractive, social, and timely

EHR: electronic health record

LMS: learning management system

Edited by J Gentges; submitted 25.Oct.2025; peer-reviewed by E Ogut, O Ng; revised version received 07.Jan.2026; accepted 14.Jan.2026; published 06.Feb.2026.

Please cite as:

Ekstrom V

Digital Choice Architecture in Medical Education: Applying Behavioral Economics to Online Learning Environments

JMIR Med Educ 2026;12:e86497

URL: <https://mededu.jmir.org/2026/1/e86497>

doi: [10.2196/86497](https://doi.org/10.2196/86497)

© Victoria Ekstrom. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 6.Feb.2026. 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.

Implementing Large Language Models to Support Misconception-Based Collaborative Learning in Health Care Education

Brandon C J Cheah^{1,2}; Shefaly Shorey³, RN, RM, PhD; Jun Hong Ch'ng¹, BSc, PhD; Chee Wah Tan¹, BSc, PhD

¹Department of Microbiology and Immunology, Yong Loo Lin School of Medicine, National University of Singapore, 5 Science Drive 2, Singapore, Singapore

²Department of Pharmacy and Pharmaceutical Sciences, National University of Singapore, Singapore, Singapore

³Alice Lee Centre of Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Corresponding Author:

Chee Wah Tan, BSc, PhD

Department of Microbiology and Immunology, Yong Loo Lin School of Medicine, National University of Singapore, 5 Science Drive 2, Singapore, Singapore

Abstract

This paper proposes a framework for leveraging large language models (LLMs) to generate misconceptions as a tool for collaborative learning in health care education. While misconceptions—particularly those generated by AI—are often viewed as detrimental to learning, we present an alternative perspective: that LLM-generated misconceptions, when addressed through structured peer discussion, can promote conceptual change and critical thinking. The paper outlines use cases across health care disciplines, including both clinical and basic science contexts, and a practical 10-step guidance for educators to implement the framework. It also highlights the need for medium- to long-term research to evaluate the impact of LLM-supported learning on student outcomes. This framework may support health care educators globally in integrating emerging AI technologies into their teaching, regardless of the disciplinary focus.

(*JMIR Med Educ* 2026;12:e81875) doi:10.2196/81875

KEYWORDS

large language model; refutation text; collaborative learning; health care education; misconception-based learning; LLM

Introduction

In an age where the availability of information is the norm and not the exception, misinformation from a range of information sources—including social media—can dominate the information landscape. Potential misinformation exists in the medical world, given the increasing rates of knowledge dissemination and the presence of potentially misleading information sources.

Medical practitioners and health professionals are not exempt from its effects, with the consequences having wide-reaching implications. An example of a high-profile case where physicians generated harm due to misinformation includes the physician-authored Great Barrington Declaration during the COVID-19 pandemic, which may have potentially increased resistance to lockdown policies in the United States and Europe [1]. Other examples include a Houston-based physician tweeting in 2022 to her account of more than 50,000 followers that there were no medically valid reasons for vaccination at the time [2], or physician-promoted anti-diet movements supporting the consumption of discretionary foods [3]. Medical training must therefore explicitly cultivate the capacity to detect and dismantle misinformation.

To address misconceptions stemming from misinformation, examples of current approaches by medical educators include manual assessment of student answers in examinations [4] and in-class polling [5] to identify misconceptions for use in misconception-based learning. However, these methods assume that (1) various student cohorts over time share similar misconceptions, without potential idiosyncratic variability in thought patterns; (2) educators have a breadth of experience with multiple cohorts; and (3) educators have the time capacity to generate multiple misconceptions in-depth. These assumptions may not hold true in practice, ultimately limiting the scalability of traditional misconception-based learning to educator experience.

The advent of ChatGPT and other large language models (LLMs) fueled by artificial intelligence (AI) presents opportunities to augment misconception-based learning. LLMs have demonstrated potential to personalize learning experiences, offer instant feedback and enhance student engagement [6]. Within misconception-based learning, LLMs offer the potential to both assist educators in generating misconceptions and students in identifying inaccurate thought patterns through feedback. However, limited research exists on LLM-assisted misconception-based learning in health care education.

Therefore, key questions arise: (1) how do LLMs fit within the process of misconception-based learning; and (2) how can LLMs be applied in the context of misconception-based learning in health care education? To address the research gap, this paper serves as a theoretical foundation and a conceptual framework with potential cases to apply LLMs in health care education.

Theoretical Premise of Misconception-Based Learning

Misinformation, and later entrenched misconceptions, can result from cognitive biases [7] and the lack of critical thinking [8]. It is important to note that scholarly debate exists on the use of the word misconception, particularly surrounding its definition and appropriateness of use [9]. Here misconceptions are referred to in the lens of the Piagetian constructivist theory, which sees misconceptions as a part of active constructions of students' mental models [10]. Misconceptions, in this context, are not deficiencies in knowledge but rather a gap between the student's perceptions and scientifically accurate information. To be more effective, educators could elucidate and challenge misconceptions early in the process of knowledge acquisition.

One of the foundations of health care professions education lies in conceptual acquisition. Conceptual acquisition could range from logic-derived processes (ie, mechanics of inspiration and breathing in physiology) to memory-based internalization (ie, eukaryotic transcription factors in molecular biology). Misconceptions would thus be the students' prior beliefs and constructed mental models that contradict accepted scientific concepts [11]. In health care professions, not addressing misconceptions originating from education may have ramifications on clinical reasoning and patient safety [12].

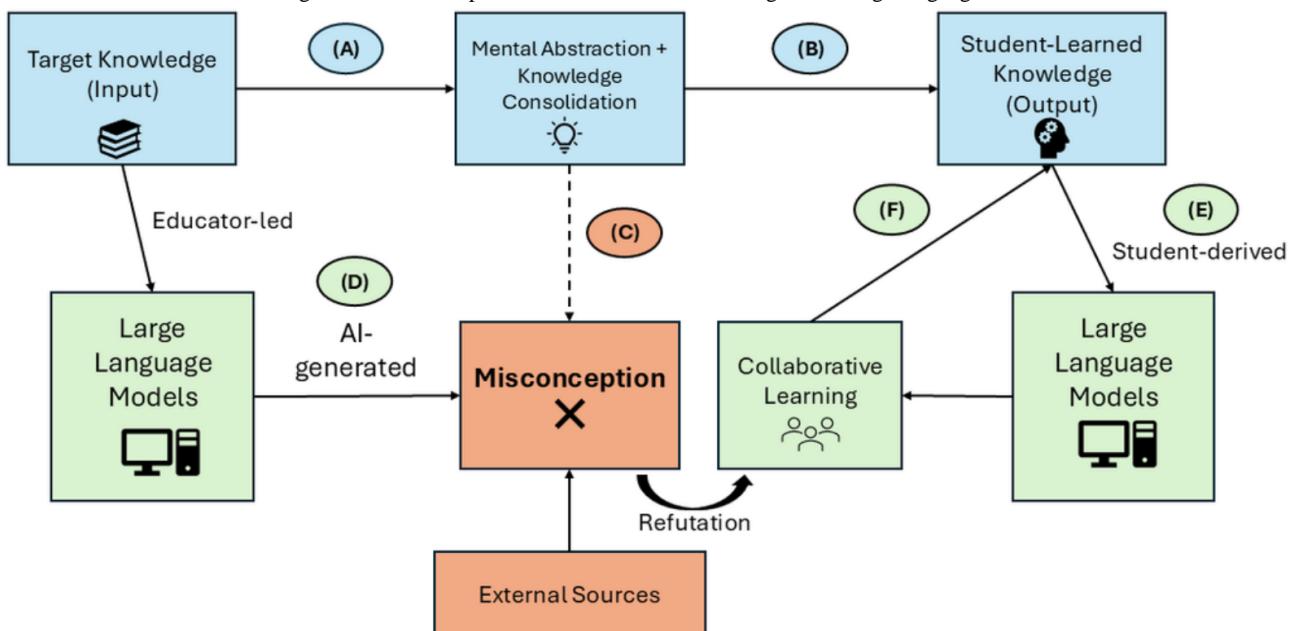
Figure 1 presents a simplified model of the process surrounding scientific conceptual acquisition. The assumed educational goal is the possession of student-learned knowledge supported by factual accuracy (output). The student is first given an input of target knowledge either through didactic or autodidactic methods. Given that the target knowledge input in health sciences can be highly conceptual, which has little to no attribution to sensorimotor information [13], the student is likely to begin the process of mental abstraction that includes

simplification and consolidation of knowledge (labeled (A) in **Figure 1**). Conceptual acquisition could occur in a mechanistic process, with the assembly of smaller and cognitively plausible concepts as a bootstrap for developing larger concepts [14].

Ideally, the iterative process of conceptual acquisition is factually accurate, with the student utilizing prior learned knowledge to acquire new concepts, a core component of critical thinking (labeled (B) in **Figure 1**). However, a student is likely to develop misconceptions due to a continuously evolving, plausible intuitive interpretation of the concept according to lived experiences (denoted by the dotted arrow labeled (C) in **Figure 1**) [15]. Misconceptions can interfere or inhibit with the further process of mental abstraction, given that the student is likely to form a false belief that the concept learned is accurate, resulting in a persistent misconception.

One of the ways to manage misconceptions arising from the process of scientific conceptual acquisition is to pre-emptively expose students to misconceptions in the classroom and systematically debunk such misconceptions. A method to which misconception-based learning can be applied in practice is using refutation text. Refutation text is a misconception-centric pedagogical method first rooted in theories of conceptual accommodation in science [16]. Conceptual accommodation is the internalization of new concepts upon facing limitations in utilizing their existing conceptions to solve problems [17]. Conceptual accommodation, therefore, is the goal of refutation text and misconception-based learning.

Three components work together to form the refutation text pedagogical method: (1) stating common but inaccurate knowledge held by the student; (2) explicit indication of the incorrect aspects of said knowledge; and (3) stating factually correct information with supporting explanation and applications [18]. The advantages of refutation text over traditional text or didactic instruction include explicit mention of assumed prior incorrect beliefs, and increased propensity for conceptual acquisition even for students with low prior topic knowledge [19]. Given that students may find it challenging to state their own misconceptions held, the educator would be the primary source of providing the refutation text. Misception exposure and the scientific method have already been in practice within the National University of Singapore common curriculum [20].

Figure 1. A framework for LLM-augmented misconception-based collaborative learning. LLM: large language model.

Integrating LLMs: An LLM-Augmented Misconception-Based Collaborative Learning Approach

There are two potential areas for LLM augmentation: educator-led or student-derived inputs of misconceptions into LLMs (denoted by the parts labeled (D) and (E), respectively, in **Figure 1**). In the educator-led case, LLMs can be used to simulate common student misconceptions. Educators can input a defined set of assessment-relevant concepts and prompt the model to generate plausible errors, providing a basis for targeted instruction. Techniques in prompt engineering include few-shot prompting [21], which is the provision of additional examples of desired tasks by the user in follow-up prompts. Given that LLMs predict appropriate outputs based on patterns in training data, providing structured follow-up prompts such as through few-shot prompting can help refine the scope and relevance of these generated misconceptions. For more complex cognitive processes, such as clinical reasoning, advanced strategies like chain-of-thought prompting—which simulates step-by-step heuristic thinking—can be employed to trace and identify potential misconceptions within the reasoning pathways of clinicians-in-training [21].

In the student-derived case, LLMs can drive deeper engagement between students and the course material. Prior student examination answers, or reference texts for the concepts, could be supplied to an LLM along with a pre-defined prompt to role-play as a junior trainee. The students could then begin a Socratic-style conversation with the LLM, identifying misconceptions based on the LLM's questions. Afterwards, the students could collaboratively correct the misconception and explain the correct understanding to the LLM. Such dialogic engagement could encourage metacognitive awareness, sustain curiosity due to the personalized nature of responses, and foster deeper conceptual understanding.

LLM-generated misconceptions complement student misconceptions revealed through assessment responses. Importantly, however, LLM-generated misconceptions may not be universally applicable across all student populations or educational contexts. As such, the lecturer's role includes adapting LLM-generated content to suit specific curricular needs, positioning the LLM as a quasi-teaching assistant in the generation of discussion prompts.

Overall, this framework in **Figure 1** outlines the approach of LLM-augmented, misconception-based collaborative learning (CL). In addition to saving educator time, reducing reliance on educator experience, and identifying idiosyncratic student misconceptions, LLMs can also drive deeper engagement between students and the learning material. Educators could then shift their focus toward fostering higher order thinking skills aligned with the upper levels of Bloom's Taxonomy, though the extent of this shift will depend on the course's learning objectives and institutional priorities. Additionally, student-led presentation on evaluating misconceptions generated by AI can integrate into existing pedagogical approaches on the flipped classroom, encouraging student autonomy. Ultimately, the flexibility in this framework's extent of implementation potentially allows for a variety of use cases.

The Utility of CL in LLM-Assisted Misconception-Based Learning

Refuting misconceptions generated by LLMs could be achieved in CL settings, such as tutorials or workshops. CL emphasizes consensus building and collective problem-solving within a peer group, in contrast to individualistic or competitive approaches. By shifting learners from passive to active modes of engagement, CL promotes critical thinking, deeper conceptual understanding, and the development of social support networks [22].

In the context of addressing misconceptions, CL provides a structured environment for students to identify, articulate, and refute conceptual errors. A three-step approach can be used to scaffold this process: (1) students first work individually to pinpoint and justify any conceptual flaws within a statement; (2) small groups are then allocated time to discuss and contrast these ideas with factually accurate knowledge, refining their reasoning through peer interaction; and (3) groups present their consensus and rationale to the larger class, facilitating the synthesis and reinforcement of accurate understanding. After the group presentation, educators could identify strengths and weaknesses in the student groups' arguments. Additionally, LLMs could provide supplemental feedback alongside educators' comments on the group's quality of arguments. CL can thus have positive feedback on student-learned knowledge and conceptual retention (denoted by the part labeled (F) in Figure 1).

CL can present a risk of unmoderated reinforcement of misconceptions in LLM-supported environments through distractor information. Distractor information generated by LLMs adds unnecessary extraneous cognitive load to a student's working memory. This generates a quasi-redundancy effect that involves processing unnecessary or misleading information, possibly reducing learning efficiency [23].

However, the presence of CL here can be a crucial first line "guardrail" to the accuracy of knowledge. Through peer discussion, learners can identify and correct hallucination errors, making CL an essential rather than peripheral component of the learning process. CL has shown to be resource efficient in digitally enabled classroom settings when addressing misconceptions [24]. Educators could then review the understanding of the student group, acting as a second "guardrail" to address problematic misconceptions. Educators could further explore initiatives to scaffold understanding through diagnosis of student understanding and reinforcing understanding within CL groups with learning checkpoints [25].

Operational Considerations and Obstacles for Implementation

A key challenge in applying LLMs to misconception-based CL lies in data quality. LLMs learn from data patterns, and their performance depends critically on the quality and reliability of training materials [26]. Low-quality or inappropriate data can result in hallucinations or the generation of misconceptions that misrepresent core concepts. To mitigate this, domain-specific LLM architectures could be trained to draw exclusively from vetted, purpose-built educational resources such as notes, textbooks, or slides. A cost-efficient alternative for developing domain-specific LLMs is to input relevant educational resources into general-purpose models, such as GPT-4o, prior to prompt generation—though this approach may yield less domain-specific or contextually relevant outputs.

Providing high-quality data for LLM training is especially vital in health care education, where complex competencies such as clinical reasoning and clinical history taking have significant patient safety implications. In these contexts, custom LLM

pipelines that utilize high-quality data—including pre-existing manual diagnostic frameworks currently in use by clinicians—are required. However, given present concerns on data security and ethical use, such data must be rigorously vetted for content, suitability, and compliance with privacy regulations before it is used to support learning and assessment. Privacy breaches in AI-assisted learning environments could compromise patient confidentiality and institutional trust in patient data security. Any adoption of LLMs in medical curricula must therefore follow the same rigor as clinical data governance.

Additionally, given that there is an absence of gold standards for clinical reasoning pipelines, it may be challenging in practice to develop an AI model specifically for clinical reasoning. Such pipelines of reasoning could be obtained from experienced clinicians for database construction, which would lead to greater quality of input data for LLMs to generate misconceptions for use in education.

While LLMs can generate customized feedback, pitfalls of LLMs in education include its potential to mislead students with false answers under the pretense of conversational and easy-to-understand responses [27]. However, a learning opportunity exists amidst the discourse, that is, how might we, as educators (1) utilize such misconceptions as a teaching opportunity, and (2) demonstrate the value of critical thinking and source evaluation in the process? A shift in the paradigm of conceptual acquisition can be achieved by seeing LLMs not as a source of harm, but as a pedagogical opportunity. This is pertinent in the health care education context, as hallucinated misinformation generated by LLMs may reinforce clinically unsafe assumptions if students are not trained to critically verify references.

Recent advances in ChatGPT-4o functionality include a deep research mode, which attempts to provide source references for the content it generates. However, because many publications remain behind paywalls and inaccessible to LLMs, the tool may still produce inaccurate citations or "hallucinated" references despite this feature. This limitation presents an opportunity for educators to help students critically evaluate the credibility of retrieved materials, distinguishing peer-reviewed from non-peer-reviewed sources and identifying instances of fabricated references.

Use Cases for LLM-Assisted Misconception Generation

Educators seeking to use LLMs such as ChatGPT for generating misconceptions can employ role-play within a single prompt, followed by targeted clarification prompts to leverage the few-shot learning capabilities of LLMs. Table 1 illustrates example use cases across disciplines including pharmacy, nursing, medicine, molecular biology, and microbiology, highlighting the potential application of this approach to both the practice-oriented and basic science domains of health care education. Table 1 only displays the initial prompt and excludes any subsequent prompts for elaboration based on the individual educator's expertise and use needs.

Table . Example use cases generated by GPT-4o for large language model (LLM)–assisted misconception-based collaborative learning.

Discipline	Concept(s)	Prompt	Learning goal	ChatGPT-4o output
Pharmacy	Hepatic extraction ratio	I am an educator within the pharmacy department of my university. I would like to teach the concept of hepatic extraction ratio. Come up with likely misconceptions that students may possess.	<ul style="list-style-type: none"> Differentiate between pharmacokinetic parameters of hepatic extraction ratio and hepatic clearance. 	<ul style="list-style-type: none"> Misconception about hepatic extraction ratio: “The hepatic extraction ratio is the same as hepatic clearance.” Why it’s wrong: Hepatic clearance depends on both hepatic blood flow and extraction ratio, while the hepatic extraction ratio (HER) is only the fraction of drug removed from the blood during a single pass through the liver. So, clearance can vary even if HER is constant, and vice versa.

Discipline	Concept(s)	Prompt	Learning goal	ChatGPT-4o output
Nursing	High-flow oxygen therapy	I am a nurse educator in my university's school of medicine. I am trying to convey that not all patients benefit from high-flow oxygen therapy, and that post-operation patients should be given fluid intake. Come up with likely misconceptions that students may have about this concept.	<ul style="list-style-type: none"> Apply clinical reasoning to determine when high-flow oxygen is indicated or contraindicated. Explain the effects of the lack of fluid intake post-surgery. 	<ul style="list-style-type: none"> Common misconception about high-flow oxygen therapy: "All patients with low SpO_2^a need high-flow oxygen." Why it's wrong: Not all hypoxic patients benefit from aggressive oxygen therapy (eg, patients with COPD^b at risk for hypercapnia). Reinforce understanding by highlighting clinical scenarios where high-flow oxygen is beneficial (eg, severe hypoxia or ARDS^c) versus when it can be detrimental (eg, certain COPD exacerbations). Common misconception about fluid intake in post-operation patients: "Patients should remain NPO^d for as long as possible post-surgery." Why it's wrong: Prolonged NPO status can increase the risk of ileus, dehydration, and delays in recovery. Early feeding (as tolerated) is now recommended in many ERAS (enhanced recovery after surgery) protocols.
Medicine	Dengue immunity and waning antibody titres	I am a clinician educator in my university hospital. I am delivering a lecture to a group of medical students. I would like to convey the concept of waning immunity and the fact that having dengue once does not guarantee long-term immunity. Provide some common student misconceptions.	<ul style="list-style-type: none"> Explain serotype-specific immunity and the effect of waning antibody titres in dengue reinfection. 	

Discipline	Concept(s)	Prompt	Learning goal	ChatGPT-4o output
				<ul style="list-style-type: none"> • Common misconception on dengue immunity: "Having dengue once guarantees long-term immunity against it." • Why it's wrong: Dengue is caused by four distinct serotypes (DENV1, DENV2, DENV3, and DENV4). Infection with one serotype only provides long-term immunity to that specific serotype. It does not confer complete protection against the others. Vaccines must balance immunity across all four serotypes. • Common misconception on dengue immunity: "Once infected, dengue antibodies stay strong forever." • Why it's wrong: Antibody levels wane over time, reducing neutralization and potentially allowing more severe disease upon reinfection.
Molecular Biology	DNA replication	I am teaching an introductory biology class to freshmen in Pharmacy. I am trying to convey the concept that DNA replication begins at multiple points of origin and DNA Polymerase uses dNTPs ^e for DNA replication. Come up with likely misconceptions that students may have about these two concepts.		<ul style="list-style-type: none"> • Explain that DNA replication begins at multiple points. • Explain how dNTPs provide energy for phosphodiester bond formation during DNA synthesis.

Discipline	Concept(s)	Prompt	Learning goal	ChatGPT-4o output
				<ul style="list-style-type: none"> • Common misconception about DNA replication at multiple points (in eukaryotes): “Replication starts at only one point, like in bacteria.” • Why it’s wrong: Students often generalize what they learn from prokaryotic replication (which starts at a single origin) to eukaryotic systems. • Common misconception about dNTPs in DNA replication: “The dNTPs are added in their final form directly into DNA.” • Why it’s wrong: Students may not understand that the energy for bond formation comes from the triphosphate tail of the dNTP, which is hydrolyzed during incorporation.
Microbiology	Viral spillover effects in humans, and viral evolution	I am an educator within my university’s microbiology and immunology department. I am delivering a lecture on SARS-CoV-2 spillover effects and viral evolution. Generate a list of plausible student misconceptions that I can debunk and explain.	<ul style="list-style-type: none"> • Explain how evolutionary pressures might shape viral transmissibility without assuming linear directionality. • Describe viral mutation as a process of random mutation. 	<ul style="list-style-type: none"> • Common misconception about SARS-CoV-2 spillover: “Viruses always evolve to become weaker over time.” • Why it’s wrong: Evolution is not directional; some variants evolve for higher transmissibility, some for immune escape, and others may evolve increased or decreased virulence. • Common misconception about viral evolution: “Viruses evolve intentionally to bypass immunity or treatments.” • Why it’s wrong: Evolution is a process of random mutations and natural selection — there’s no ‘intent’ behind it.

^aSpO₂: oxygen saturation.

^bCOPD: chronic obstructive pulmonary disease.

^cARDS: acute respiratory distress syndrome.

^dNPO: nil per os (nothing by mouth).

^edNTP: deoxynucleotide triphosphate.

Of interest are the inherent assumptions made by ChatGPT about a student's prior knowledge in the subject. For example, in the molecular biology misconception on DNA replication (Table 1), ChatGPT assumes that students in an introductory molecular biology class can contrast DNA replication between prokaryotes and eukaryotes, despite the prompt only indicating misconceptions on multiple points of DNA replication. Such an assumption may not hold true in an average cohort. This may indicate a form of metacognitive hindsight bias intrinsic within GPT-related architectures. Educators could then include clarification prompts on the existing knowledge base of the target audience and modify AI-generated misconceptions to suit teaching needs.

In addition, educators can draw students' attention to instances of hasty generalization within LLM-generated misconceptions such as in Table 1. For example, absolute qualifiers such as "all" (nursing prompt), "guarantees" (medical prompt), or "always" (microbiology prompt) imply universal causation with no exceptions. By explicitly highlighting these terms, educators can introduce counterexamples that challenge such overgeneralizations, providing a method for debunking misconceptions.

Finally, educators with prior experience of running past editions of the course in previous years could provide students' examination answers as an input for LLM analysis, after anonymizing student answer scripts. This could clarify the scope and customize the misconceptions generated to a specific educator's need. However, given excess specific input, the LLM runs a risk of factoring in irrelevant information, potentially leading to erroneous outputs. Educators could potentially mitigate this by limiting excessive irrelevant information and vetting the input contents to prevent overloading of the LLM's context window.

A 10-Step Guide for Implementing LLM-Augmented Misconception-Based CL

For educators who intend on implementing the framework in Figure 1 in practice, the following is a suggested 10-step guide with the concept of serotype-specific immunity in Table 1 used as an example.

Educator-Led Steps

Step 1: Setup

Input lecture slides and/or reference textbook chapters into the LLM. Include specifics of your role (eg, clinician educator) and audience (eg, medical students) in the prompt.

In this step, educators should input lecture slides and/or reference textbook chapters into the LLM. Include specifics of

your role (eg, clinician educator) and audience (eg, medical students) in the prompt.

Step 2: Review

Educators review the misconceptions provided. LLMs such as ChatGPT-4o may occasionally provide misconceptions based on content outside the scope of coverage.

Step 3: Tailor

Educators tailor the subsequent prompts to vary the length of misconceptions provided by the LLM. For example, if the concepts covered pertain to the nature of serotype-specific immunity and its relationship to long-term immunity such as antibody-dependent enhancement, a prompt instructing the LLM to integrate these concepts into a single misconception statement may be necessary. The integration of multiple related concepts into one misconception can test students' attention to detail.

Student-Derived Steps

Step 4: Setup

Upload anonymized student examination answers and lecture notes into an institutional pre-trained LLM chatbot.

Step 5: Role-play

Prompt the chatbot to: (1) assume the role of a junior medical student, (2) engage in a Socratic-style conversational dialogue, (3) ask questions individually and reply to responses with a follow-up question, and (4) provide some misconceptions within its train of thought based on the uploaded student examination answers.

Alternatively, if institutional chatbots are not available, ChatGPT-4o may be used. Only input lecture notes and/or reference textbook chapters and not student examination answers would be used in this case, as sending past student examination answers to current students may potentially infringe upon data privacy.

Step 6: Group

Depending on the cohort size and availability of educators, split the students into groups of 3 - 5 to discuss the generated misconceptions for CL purposes. CL should preferably be implemented after students have gone through the content at least once, either through pre-reading or classroom instruction.

Step 7: Present

Student groups should conduct short presentations to identify and explain the LLM-generated misconceptions.

Step 8: Converse

Instruct the students to access the prompted institutional chatbot or ChatGPT-4o and initiate the conversation.

Step 9: Collaborate

Allow the students to work collaboratively in groups to address the questions posed by the LLM and “teach” it the concepts. Students could download the chat history or transcripts after the conclusion of the “lesson.”

Step 10: Assess

Students’ misconceptions could be assessed afterwards by reviewing their responses within the downloaded chat transcripts.

Limitations and Directions for AI-Assisted Pedagogy in Health Care Education Studies

It should be noted that only 2 quantitative studies [28,29] have been conducted on misconception-based reasoning and the wider misconception-based learning literature within the context of health care education. More than half of these studies are focused on the fields of Newtonian physics, genetically modified organisms, and evolution [30]. While misconception-based learning was determined to be robust across disciplines independent of specific text design [31], generalizing misconception-based learning’s efficacy from other fields to health care education may not be fully applicable. More studies are required to examine the impact of this pedagogical method on student outcomes in health care education.

A common methodological critique of this pedagogical approach is its vulnerability to pre-test priming effects. In typical designs, students complete a pre-test before exposure to the instructional intervention and a post-test afterward. However, evidence suggests that notifying students about an impending test may activate prior knowledge and trigger conceptual change, making it challenging to distinguish between the effects of pre-test priming and those of LLM-assisted, misconception-based CL. To address this, researchers could track individual student responses and apply statistical controls—such as paired *t* tests or ANCOVA analyses—to adjust for pretest influences and better isolate the intervention’s true impact.

CL applied in LLM-generated misconception generation assumes that there is a form of homogeneity in conceptual understanding within and between groups. In a group comprised only of weaker students, false explanations of the LLM-generated misconceptions may be negatively reinforced instead. While ensuring the distribution of students with varying backgrounds and degrees of conceptual understanding within groups may be ideal to fully capitalize on the benefits of CL, such methods are difficult to implement in practice and are a key limitation of this pedagogical method.

Given the limited number of studies present to examine AI-assisted pedagogy in health care education, multiple avenues for follow-up research studies exist. Few studies have assessed cognitive attitudes and student perspectives on CL using AI [32]. Importantly, the critical nature of such studies stems from the increasingly interconnected working environment. Studying cognitive attitudes on CL using AI from students’ perspectives

could better associate AI technology implementation in the classroom with student outcomes.

Additionally, there has been a lack of medium- to long-term research on student outcomes using AI [33]. The prevalence of AI technologies and its impact on the educational setting have far-reaching implications for changing pedagogy and student learning. However, since the rise of LLMs to prominence in 2023, studies have largely been cross-sectional, capturing only a snapshot of AI’s influence at a specific point in its rapid development.

Cross-sectional studies are likely only able to capture a snapshot of AI’s impact on student learning at a specific point of time in AI’s development. Such studies often overlook how a student’s learning trajectory evolves alongside advances in AI itself. For example, an educational experience supported by GPT-3 would likely differ substantially from one shaped by GPT-4o, which introduced capabilities such as image generation. Longitudinal, mixed-methods studies are needed to track these evolving dynamics, providing deeper insights into both shifts in student perceptions and longer-term learning outcomes as AI technologies continue to evolve. Such an approach could also assist to identify key factors that support or hinder the long-term implementation of AI in teaching practice.

Yet, it should be noted that the fast-evolving nature of LLMs—which at the time of writing include Gemini 2.5 and the upcoming GPT-5 by OpenAI—could impact the reproducibility of generated misconceptions. With larger token sizes and greater architecture complexity, prompts provided to newer models may not produce the same outputs. Future educational research involving LLMs could ensure the reporting of the specific model version used, sampling conditions, and prompts used for an effort towards greater reproducibility. While the stochastic nature of LLMs carries a risk of low reproducibility of outputs with the same prompt, the role of educators to mitigate such a risk could be to verify and vet the LLM-generated outputs. Thus, LLM-generated misconceptions should serve to complement curriculum design, and the educator should not be removed from the process.

Additionally, greater data privacy risk comes with the evolving dynamics of new LLM models. The extent to which newer models retain data inputs by users through larger context windows are not well understood. Therefore, such risks could be mitigated by educators through anonymization and the removal of original file metadata. On the end of educational technology developers, methods to protect data security such as adversarial attack testing or encryption specific to LLM modalities could be explored [34].

Finally, explicit indications of the misconception and subsequent explanations by educators may not always work if the misconception is entrenched in the students’ mental models. Greater emphasis needs to be placed on understanding the root of the falsehood, perhaps through analysis of past lived experiences of students that led to this network of beliefs, to facilitate effective conceptual change. Addressing misconceptions in education would therefore likely not be a one-size-fits-all solution. Supplementing refutation-based learning with other educational approaches such as concept

maps and audiovisual representation may improve coherence of concept representation [29].

Conclusion – LLM’s Potential for Augmenting Misconception-Based Learning in Health Care Education

Misconceptions, which could stem from naive misinformation, have implications on critical thinking and student learning. Yet, when leveraged intentionally—for example, through LLM-augmented, misconception-based CL—they can catalyze transformative shifts in teaching practice. This approach not only deepens students’ understanding of subject matter but also provides a valuable teaching opportunity for exploring both the benefits and limitations of LLMs.

Additionally, the framework serves as a springboard for future empirical research, generating several potentially testable

hypotheses. For example, does pre-bunking with LLM-generated misconceptions elicit stronger propensity for conceptual accommodation as compared to educator-generated refutation text? Does CL around LLM-generated misconceptions improve reasoning accuracy and quality compared to individual learning? Does LLM-generated misconceptions change classroom dynamics for CL? These are important questions for identifying and testing the value of LLMs in misconception-based CL.

However, the true value of LLMs in misconception-based learning exists not to replace educator involvement and expertise. Rather, its purpose is to augment student learning with personalized feedback, synthesize “edge-case” concepts across topics, and reduce the time taken for educators by supplementing misconceptions. The educator’s role is still important in context-setting, the facilitation of CL, and critical review of the AI output. Ultimately, the proposed framework positions LLMs not as an endpoint for teaching and learning, but to foster self-directed, critical, and reflective thinking.

Acknowledgments

We used the generative AI tool ChatGPT by OpenAI to generate examples of use cases in [Table 1](#). Additionally, ChatGPT was used to refine the grammatical structure of the main text to enhance clarity and readability. We would like to thank Dr Ronald Boyd Anderson (National University of Singapore) for his kind comments to improve the paper on the concepts surrounding LLMs and reproducibility.

Funding

This work was funded by the National University of Singapore Startup Grant (NUHSRO/2023/018/Startup/10).

Data Availability

No original data were generated or collected for the purpose of this paper.

Authors' Contributions

Conceptualization: BCJC (lead), JHC (supporting), CWT (supporting)

Investigation: BCJC

Methodology: BCJC

Evidence curation: BCJC

Writing – Original draft: BCJC

Writing – Review and editing: BCJC (lead), JHC (supporting), CWT (supporting), SS (supporting)

Supervision: JHC (equal), CWT (equal), SS (supporting)

Funding acquisition: CWT

Conflicts of Interest

None declared.

References

1. Tandar CE, Lin JC, Stanford FC. Combating medical misinformation and rebuilding trust in the USA. *Lancet Digit Health* 2024 Nov;6(11):e773-e774. [doi: [10.1016/S2589-7500\(24\)00197-3](https://doi.org/10.1016/S2589-7500(24)00197-3)] [Medline: [39379267](https://pubmed.ncbi.nlm.nih.gov/39379267/)]
2. Rubin R. When physicians spread unscientific information about COVID-19. *JAMA* 2022 Mar 8;327(10):904. [doi: [10.1001/jama.2022.1083](https://doi.org/10.1001/jama.2022.1083)]
3. Denniss E, Lindberg R. Social media and the spread of misinformation: infectious and a threat to public health. *Health Promot Int* 2025 Mar 5;40(2):daaf023. [doi: [10.1093/heapro/daaf023](https://doi.org/10.1093/heapro/daaf023)] [Medline: [40159949](https://pubmed.ncbi.nlm.nih.gov/40159949/)]
4. Olde Bekkink M, Donders A, Kooloos JG, de Waal RMW, Ruiter DJ. Uncovering students’ misconceptions by assessment of their written questions. *BMC Med Educ* 2016 Aug 24;16(1):221. [doi: [10.1186/s12909-016-0739-5](https://doi.org/10.1186/s12909-016-0739-5)] [Medline: [27558400](https://pubmed.ncbi.nlm.nih.gov/27558400/)]
5. Bordes SJ, Manyevitch R, Huntley JD, Li Y, Murray IVJ. Medical student misconceptions in cardiovascular physiology. *Adv Physiol Educ* 2021 Jun 1;45(2):241-249. [doi: [10.1152/advan.00220.2020](https://doi.org/10.1152/advan.00220.2020)] [Medline: [33825522](https://pubmed.ncbi.nlm.nih.gov/33825522/)]

6. Al-Zahrani AM, Alasmari TM. Exploring the impact of artificial intelligence on higher education: the dynamics of ethical, social, and educational implications. *Humanit Soc Sci Commun* 2024 Dec 1;11(1). [doi: [10.1057/s41599-024-03432-4](https://doi.org/10.1057/s41599-024-03432-4)]
7. Britt MA, Rouet JF, Blaum D, Millis K. A reasoned approach to dealing with fake news. *Policy Insights Behav Brain Sci* 2019 Mar;6(1):94-101. [doi: [10.1177/2372732218814855](https://doi.org/10.1177/2372732218814855)]
8. List JA, Ramirez LM, Seither J, Unda J, Vallejo BH. Critical thinking and misinformation vulnerability: experimental evidence from Colombia. *PNAS Nexus* 2024 Oct;3(10):pgae361. [doi: [10.1093/pnasnexus/pgae361](https://doi.org/10.1093/pnasnexus/pgae361)] [Medline: [39411082](https://pubmed.ncbi.nlm.nih.gov/39411082/)]
9. Leonard MJ, Kalinowski ST, Andrews TC. Misconceptions yesterday, today, and tomorrow. *LSE* 2014 Jun;13(2):179-186. [doi: [10.1187/cbe.13-12-0244](https://doi.org/10.1187/cbe.13-12-0244)]
10. Rowell JA, Dawson CJ, Lyndon H. Changing misconceptions: a challenge to science educators. *Int J Sci Educ* 1990 Apr;12(2):167-175. [doi: [10.1080/0950069900120205](https://doi.org/10.1080/0950069900120205)]
11. Boshuizen HPA, Marambe KN. Misconceptions in medicine, their origin and development in education and working life. *Int J Educ Res* 2020;100:101536. [doi: [10.1016/j.ijer.2020.101536](https://doi.org/10.1016/j.ijer.2020.101536)]
12. Södervik I, Mikkilä-Erdmann M, Chi MTH. Conceptual change challenges in medicine during professional development. *Int J Educ Res* 2019;98:159-170. [doi: [10.1016/j.ijer.2019.07.003](https://doi.org/10.1016/j.ijer.2019.07.003)]
13. Yee E. Abstraction and concepts: when, how, where, what and why? *Lang Cogn Neurosci* 2019 Nov 26;34(10):1257-1265. [doi: [10.1080/23273798.2019.1660797](https://doi.org/10.1080/23273798.2019.1660797)]
14. diSessa AA. Toward an epistemology of physics. *Cogn Instr* 1993 Apr;10(2-3):105-225. [doi: [10.1080/07370008.1985.9649008](https://doi.org/10.1080/07370008.1985.9649008)]
15. Vosniadou S, Ioannides C. From conceptual development to science education: a psychological point of view. *Int J Sci Educ* 1998 Dec;20(10):1213-1230. [doi: [10.1080/0950069980201004](https://doi.org/10.1080/0950069980201004)]
16. Posner GJ, Strike KA, Hewson PW, Gertzog WA. Accommodation of a scientific conception: toward a theory of conceptual change. *Sci Educ* 1982 Apr;66(2):211-227. [doi: [10.1002/sce.3730660207](https://doi.org/10.1002/sce.3730660207)]
17. Nussbaum J, Novick S. Alternative frameworks, conceptual conflict and accommodation: toward a principled teaching strategy. *Instr Sci* 1982 Dec;11(3):183-200. [doi: [10.1007/BF00414279](https://doi.org/10.1007/BF00414279)]
18. Danielson RW, Jacobson NG, Patall EA, et al. The effectiveness of refutation text in confronting scientific misconceptions: a meta-analysis. *Educ Psychol* 2025 Jan 2;60(1):23-47. [doi: [10.1080/00461520.2024.2365628](https://doi.org/10.1080/00461520.2024.2365628)]
19. Hynd CR. Refutational texts and the change process. *Int J Educ Res* 2001 Jan;35(7-8):699-714. [doi: [10.1016/S0883-0355\(02\)00010-1](https://doi.org/10.1016/S0883-0355(02)00010-1)]
20. National University of Singapore, Faculty of Science. Stay curious, question everything: 'Citizen scientists' explore how and why science works. 2024. URL: <https://www.science.nus.edu.sg/blog/2024/04/innovative-educators-assoc-prof-ryan-bettens/> [accessed 2025-07-09]
21. Chen B, Zhang Z, Langrené N, Zhu S. Unleashing the potential of prompt engineering for large language models. *Patterns* (N Y) 2025 Jun 13;6(6):101260. [doi: [10.1016/j.patter.2025.101260](https://doi.org/10.1016/j.patter.2025.101260)] [Medline: [40575123](https://pubmed.ncbi.nlm.nih.gov/40575123/)]
22. Laal M, Ghodsi SM. Benefits of collaborative learning. *Procedia - Social and Behavioral Sciences* 2012;31:486-490. [doi: [10.1016/j.sbspro.2011.12.091](https://doi.org/10.1016/j.sbspro.2011.12.091)]
23. Sweller J. Cognition in education. In: Ross BH, Mestre JP, editors. *Psychol Learn Motiv*: Elsevier; 2011:313 URL: <https://shop.elsevier.com/books/cognition-in-education/mestre/978-0-12-387691-1> [accessed 2026-01-05]
24. Jansen M, Kim Cavdar I. Adapting team-based learning for medical education: a case study with scalable and resource-efficient implementation. *MedSciEduc* ;35(2):883-892. [doi: [10.1007/s40670-024-02246-y](https://doi.org/10.1007/s40670-024-02246-y)]
25. van de Pol J, Mercer N, Volman M. Scaffolding student understanding in small-group work: students' uptake of teacher support in subsequent small-group interaction. *Journal of the Learning Sciences* 2019 Mar 15;28(2):206-239. [doi: [10.1080/10508406.2018.1522258](https://doi.org/10.1080/10508406.2018.1522258)]
26. Cheah BCJ, Vicente CR, Chan KR. Machine learning and artificial intelligence for infectious disease surveillance, diagnosis, and prognosis. *Viruses* 2025 Jun 23;17(7):882. [doi: [10.3390/v17070882](https://doi.org/10.3390/v17070882)] [Medline: [40733500](https://pubmed.ncbi.nlm.nih.gov/40733500/)]
27. Li R, Wu T. Delving into the practical applications and pitfalls of large language models in medical education: narrative review. *Adv Med Educ Pract* 2025;16:625-636. [doi: [10.2147/AMEP.S497020](https://doi.org/10.2147/AMEP.S497020)] [Medline: [40271151](https://pubmed.ncbi.nlm.nih.gov/40271151/)]
28. Holford D, Mäki KO, Karlsson LC, Lewandowsky S, Gould VC, Soveri A. A randomized controlled trial of empathetic refutational learning with health care professionals. *BMC Public Health* 2025 Feb 12;25(1):583. [doi: [10.1186/s12889-025-21787-4](https://doi.org/10.1186/s12889-025-21787-4)] [Medline: [39939961](https://pubmed.ncbi.nlm.nih.gov/39939961/)]
29. Versteeg M, van Loon MH, Wijnen-Meijer M, Steendijk P. Refuting misconceptions in medical physiology. *BMC Med Educ* 2020 Aug 5;20(1):250. [doi: [10.1186/s12909-020-02166-6](https://doi.org/10.1186/s12909-020-02166-6)] [Medline: [32758215](https://pubmed.ncbi.nlm.nih.gov/32758215/)]
30. Zengilowski A, Schuetze BA, Nash BL, Schallert DL. A critical review of the refutation text literature: methodological confounds, theoretical problems, and possible solutions. *Educ Psychol* 2021 Jul 3;56(3):175-195. [doi: [10.1080/00461520.2020.1861948](https://doi.org/10.1080/00461520.2020.1861948)]
31. Schroeder NL, Kucera AC. Refutation text facilitates learning: a meta-analysis of between-subjects experiments. *Educ Psychol Rev* 2022;34(2):957-987. [doi: [10.1007/s10648-021-09656-z](https://doi.org/10.1007/s10648-021-09656-z)] [Medline: [35095236](https://pubmed.ncbi.nlm.nih.gov/35095236/)]
32. Yu JH, Chauhan D, Iqbal RA, Yeoh E. Mapping academic perspectives on AI in education: trends, challenges, and sentiments in educational research (2018–2024). *Education Tech Research Dev* 2025 Feb;73(1):199-227. [doi: [10.1007/s11423-024-10425-2](https://doi.org/10.1007/s11423-024-10425-2)]

33. Adiguzel T, Kaya MH, Cansu FK. Revolutionizing education with AI: exploring the transformative potential of ChatGPT. *CONT ED TECHNOLOGY* 2023;15(3):ep429. [doi: [10.30935/cedtech/13152](https://doi.org/10.30935/cedtech/13152)]
34. Yan B, Li K, Xu M, et al. On protecting the data privacy of large language models (llms): a survey. 2024 Mar 14 Presented at: 2024 International Conference on Meta Computing (ICMC). [doi: [10.1109/ICMC60390.2024.00008](https://doi.org/10.1109/ICMC60390.2024.00008)]

Abbreviations

AI: artificial intelligence
CL: collaborative learning
LLM: large language model
MBC: misconception-based collaborative

Edited by S Tsuei; submitted 09.Aug.2025; peer-reviewed by A Alhur, JYW Lee, MNH Mohamad SaidnAff; revised version received 05.Nov.2025; accepted 06.Nov.2025; published 16.Jan.2026.

Please cite as:

Cheah BCJ, Shorey S, Ch'ng JH, Tan CW
Implementing Large Language Models to Support Misconception-Based Collaborative Learning in Health Care Education
JMIR Med Educ 2026;12:e81875
URL: <https://mededu.jmir.org/2026/1/e81875>
doi: [10.2196/81875](https://doi.org/10.2196/81875)

© Brandon C J Cheah, Shefaly Shorey, Jun Hong Ch'ng, Chee Wah Tan. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 16.Jan.2026. 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.

Global Learner Feedback on Continuing Medical Education—Accredited e-Learning Modules in Pediatric Endocrinology and Diabetes: Cross-Sectional Study

Yvonne G van der Zwan¹, MD, PhD; Conny van Wijngaard-de Vugt², MScBA; Abdulsalam Abu-Libdeh^{3,4}, MD, PhD; Evangelia Kalaitzoglou⁵, MD, PhD; Zacharoula Karabouta⁶, MD, PhD; Stenvert L S Drop⁷, MD, PhD; Annemarieke M Boot⁸, MD, PhD; Sze May Ng^{9,10}, MD, PhD; Jan Idkowiak^{11,12,13}, MD, PhD

¹Department of Paediatrics, Isala, Zwolle, The Netherlands

¹⁰Department of Pediatric Endocrinology, Mersey and West Lancashire Teaching Hospitals NHS Trust, Southport, United Kingdom

¹¹Department of Metabolism and Systems Science, College of Medicine and Health, University of Birmingham, Edgbaston, IBR tower, level 2, Birmingham, United Kingdom

¹²Centre for Endocrinology, Diabetes and Metabolism, Birmingham Health Partners, University of Birmingham, Birmingham, United Kingdom

¹³Department of Endocrinology and Diabetes, Birmingham Children's Hospital, Birmingham Women's and Children's NHS Foundation Trust, Birmingham, United Kingdom

²WV Research, Advice and Management, Brunssum, The Netherlands

³Department of Pediatrics, Division of Pediatric Endocrinology, Makassed Hospital and Al-Quds Medical School, Jerusalem, Palestinian Territory, Occupied

⁴Division of Pediatric Endocrinology, Hadassah University Medical Center & Hadassah Medical School, Jerusalem, Israel

⁵Department of Pediatrics, University of Kentucky, Barnstable Brown Diabetes Center, Lexington, KY, United States

⁶2nd Paediatric Department, AHEPA University General Hospital, Thessaloniki, Greece

⁷Division of Endocrinology, Department of Pediatrics, Sophia Children's Hospital, Erasmus MC, Rotterdam, The Netherlands

⁸Department of Pediatric Endocrinology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

⁹Edge Hill University, Ormskirk, England, United Kingdom

Corresponding Author:

Jan Idkowiak, MD, PhD

Department of Metabolism and Systems Science, College of Medicine and Health, University of Birmingham, Edgbaston, IBR tower, level 2, Birmingham, United Kingdom

Abstract

Background: The European Society for Paediatric Endocrinology (ESPE) e-Learning website is a free, globally accessible online resource to enhance learning in pediatric endocrinology and pediatric diabetes. The content is created by world-leading experts in pediatric endocrinology and pediatric diabetes and is closely aligned with published international consensus guidelines. In August 2022, 30 hours of e-learning courses received accreditation from the European Accreditation Council for Continuing Medical Education (CME). These CME courses cover three categories: (1) pediatric endocrinology, (2) pediatric diabetes, and (3) pediatric endocrinology in resource-limited settings.

Objective: This study aimed to assess learners' demographics and feedback from mandatory surveys after completion of CME e-learning courses and to identify areas for improvement.

Methods: The ESPE e-learning committee created a mandatory survey for each CME e-learning module. The survey includes baseline demographics and feedback on the quality of the learning content, assessed using a 5-point Likert scale. Data were extracted from the start of the CME modules in August 2022 until September 2025.

Results: A total of 567 surveys were completed: 286 (50.4%) in the category pediatric endocrinology, 225 (39.7%) in the category pediatric diabetes based on the International Society for Pediatric and Adolescent Diabetes guidelines, and 56 (9.9%) in the category pediatric endocrinology in resource-limited settings. There was global participation, with most learners practicing in Europe (n=333, 59%), followed by Asia (n=124, 22%), Africa (n=53, 9%), the Americas (North America, n=45, 8%; and South America, n=11, 2%), and Oceania (n=1, 0%). Most of the users indicated that they were medical experts (n=210, 37%), followed by fellows or residents (n=223, 39%), and medical students and nurses (n=29, 5% and n=32, 6%, respectively); overall, 10% (n=56) of learners practice in resource-limited countries. Overall, the learning content was well received for all modules regarding accessibility, organization, level of interest, improvement of learners' clinical practice, appropriateness of content, and provision of feedback (median Likert score 4, IQR 4-5). Learners' free-text feedback identified some areas for improvement, including

reducing text-heavy content and providing more graphical content and more interactive case reports. Most learners' free-text feedback consisted of encouraging and thankful comments.

Conclusions: The ESPE CME-accredited e-learning modules are well received, providing globally free CME education in pediatric endocrinology and pediatric diabetes. These findings support the continued development and promotion of open-access CME platforms, with the aim of improving global equity in specialist medical education and focusing on educational impact.

(*JMIR Med Educ* 2026;12:e67332) doi:[10.2196/67332](https://doi.org/10.2196/67332)

KEYWORDS

e-learning; continuing medical education; endocrinology; European Society for Paediatric Endocrinology; ESPE; International Society for Pediatric and Adolescent Diabetes; ISPAD; pediatric endocrinology; pediatrics; diabetes; eHealth; digital health; medical education

Introduction

The European Society for Paediatric Endocrinology (ESPE) e-Learning website [1] is a free, globally accessible online tool to enhance learning in pediatric endocrinology (PE) and pediatric diabetes (PD) worldwide [2-4].

The ESPE e-Learning website was first published online in 2012. Since then, the content and use have expanded with courses available in PE and PE in resource-limited settings (RLSs). The e-learning committee of ESPE and the International Society for Pediatric and Adolescent Diabetes (ISPAD) are collaboratively responsible for establishing and maintaining the e-learning platform, as well as ensuring the development and quality of its content. This freely accessible online portal allows medical students, fellows, specialists, nurses, and tutors from around the world to share, contribute to, and expand their knowledge through interactive chapters and case-based learning, covering both core- and advanced-level modules [5,6]. During the COVID-19 pandemic, e-learning became an essential resource for continuous professional development, and online learning demonstrated that the effects of e-learning are equivalent to traditional learning [7-9]. e-Learning has now become a solution for overcoming training barriers posed by social distancing rules, offering education and learning opportunities to students, trainees, and physicians to help them maintain essential competencies and continue professional development. It is especially beneficial for individuals working residential shifts with work hour restrictions, which frequently conflict with in-person attendance at didactic lectures, and for those living in remote areas.

The ESPE e-learning website contains 80 chapters and more than 130 cases covering core and advanced learning courses on the most common endocrine disorders, including diabetes; of these, 16 chapters and 25 cases target health care professionals based in RLSs. The content includes normal physiology, pathophysiological mechanisms underlying endocrine diseases, diagnostic approaches or algorithms, and management based on international expert consensus and published guidelines. Direct feedback on questions from users is encouraged. Recently, two video masterclasses delivered by internationally recognized experts in the field have been added. Steering, conceptual, and content oversight is provided by the ESPE-ISPAD e-learning committee, which is responsible for creating and maintaining the e-learning website, ensuring up-to-date information and high-quality standards.

All courses are available in English, and all courses in the category *health care in resource-limited settings*, which is specifically intended for practitioners working in primary, secondary, and tertiary health care centers in resource-limited countries, are available in four additional languages (ie, French, Spanish, Swahili, and Chinese) [10].

Since August 2022, 30 credit hours of ESPE-ISPAD continuing medical education (CME) e-learning courses created by world-leading experts in PE and PD have received accreditation from the European Accreditation Council for CME (EACCME). These CME courses are typically based on consensus guidelines. They are organized in three categories with 10 courses each in the categories *pediatric endocrinology*, *pediatric diabetes*, and *health care in resource-limited settings*. The latter module is available in five different languages (ie, English, French, Spanish, Swahili, and Chinese). Each CME course takes approximately 1 to 1.5 hours of educational time, and completion of a course provides one European CME credit.

To promote global awareness and uptake of the CME-accredited modules, the ESPE e-learning platform was actively disseminated through multiple channels. These included presentations at international society meetings, such as ESPE and ISPAD, regular features in quarterly ESPE newsletters and bimonthly ISPAD newsletters, and targeted outreach via social media platforms, including Facebook and X. These efforts aimed to ensure broad visibility and accessibility of the platform across diverse geographic and professional audiences.

To date, little is known about how these accredited modules are used and perceived by health care professionals globally. Herein, we aimed to evaluate the geographical reach, user demographics, and learner feedback on the accredited CME modules, specifically examining how health care professionals from diverse regions and professional backgrounds engage with and evaluate the CME-accredited modules. Therefore, we conducted a descriptive cross-sectional analysis of the mandatory postcourse survey data collected between August 2022 and September 2025. We hypothesized that the modules would be positively received across user groups and that learners would identify specific areas for improvement. The findings may help inform educators, module developers, and professional societies in improving their approach to expand equitable access to high-quality, guideline-based CME in PE and PD.

Methods

Ethical Considerations

Participation in the feedback survey was entirely voluntary, and informed consent was implied by completion of the survey. Respondents were clearly informed that they could opt out at any time by not submitting the survey. No additional consent was required because this was a secondary analysis of anonymized data collected under the original consent process for CME module participation, which included agreement to provide feedback for service evaluation purposes.

Ethical approval was not deemed necessary as the analysis of the survey was conducted as an evaluation and quality improvement exercise to evaluate the educational offering. Data were collected anonymously from adults/professionals (no vulnerable population); no identifiable information was collected; there was no intervention, randomization, or direct implications on patient care; and the participation was entirely voluntary.

No financial or material compensation was provided to participants for completing the survey.

Survey Design and Setting

The reporting of this study adhered to the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) to ensure methodological transparency and reproducibility [11]. The checklist was applied to all stages of the online evaluation process, including survey design, delivery, and data handling. The evaluation questionnaire was developed by the ESPE e-learning team and pretested with a small group of users to assess clarity, face validity, and technical functionality. The final version was implemented within the online education platform, which automatically invited participants to complete the anonymous feedback survey upon module completion (items 5 - 8). Participation was mandatory upon completion of the course; no personal data, as per General Data Protection Regulation requirements, were collected. No incentives were offered. Measures to prevent multiple submissions included limiting one response per user account and restricting survey access to registered participants only. Data completeness checks were built into the survey logic, and incomplete responses were excluded from the main analysis but retained for sensitivity assessment. All data were collected through a secure institutional server with encrypted transmission and stored in accordance with international data governance standards.

The survey was created by the ESPE e-learning committee to obtain information on learners' demographics and to evaluate the use and quality of the courses. The content, in accordance with Kirkpatrick's evaluation type level 1 (*reaction*) [12,13], included questions about learners' professional backgrounds and countries of residence. Feedback on the quality of the learning content, presentation, accessibility, and the anticipated impact on clinical practice was assessed using a 5-point Likert scale, ranging from strongly agree (Likert scale=1), agree (Likert scale=2), neutral (Likert scale=3), disagree (Likert scale=4), and strongly disagree (Likert scale=5). We used a different Likert scale for one specific question asking for the degree of

difficulty of quiz questions provided in the modules (Likert scale 1=too easy, 2=easy, 3=appropriate, 4=difficult, and 5=too difficult). Likert scales were chosen to allow nuanced assessment of attitudes and perceptions, are well understood by respondents, and facilitate both descriptive and inferential statistical analyses. Additional feedback on content, user interface, or other issues was encouraged through an open free-text question. The detailed survey is provided in [Multimedia Appendix 1](#). Data were analyzed from the launch of the CME module in August 2022 until September 2025. The professional background was categorized based on professional training and experience in PE and grouped as (1) medical expert or consultant, (2) fellow, (3) resident, (4) medical student, (5) nurse, and (6) other (free text). Health care officers from research-limited countries were asked if they were based in primary, secondary, or tertiary level settings to understand resources and facilities available (primary level: basic or rural with very limited laboratory and imaging facilities; secondary level: district or regional hospitals with limited laboratory and imaging facilities; and tertiary level: main or national referral hospitals with most but not all laboratory and imaging facilities [10]).

Visualization of Global Participation

Geographical distribution of participants completing the online CME modules was visualized using *Plotly* (version 5.24.0; Plotly Technologies Inc). Country names from the registration dataset were matched to International Organization for Standardization-recognized country centroids, and each country was represented by a proportional bubble plotted on a world map in a natural-earth projection.

Statistical Analysis

Descriptive statistics were applied. When analyzing the 5-point Likert scale, we used the median (IQR) as appropriate for an ordinal scale; however, we have also provided the mean and SD to describe the level of dispersion. As Likert responses are ordinal and nonnormally distributed, comparisons across the three modules were performed using the Kruskal-Wallis H test. Where relevant, *P* values *<.05* were considered statistically significant. A stacked Likert plot was generated to visualize the proportional distribution of responses across all survey questions. Free-text feedback was independently coded into four categories by two authors. Discrepancies were discussed, and final classifications were reached by consensus. We have used Microsoft Excel and GraphPad Prism for descriptive statistics and graphical illustration of the data.

Results

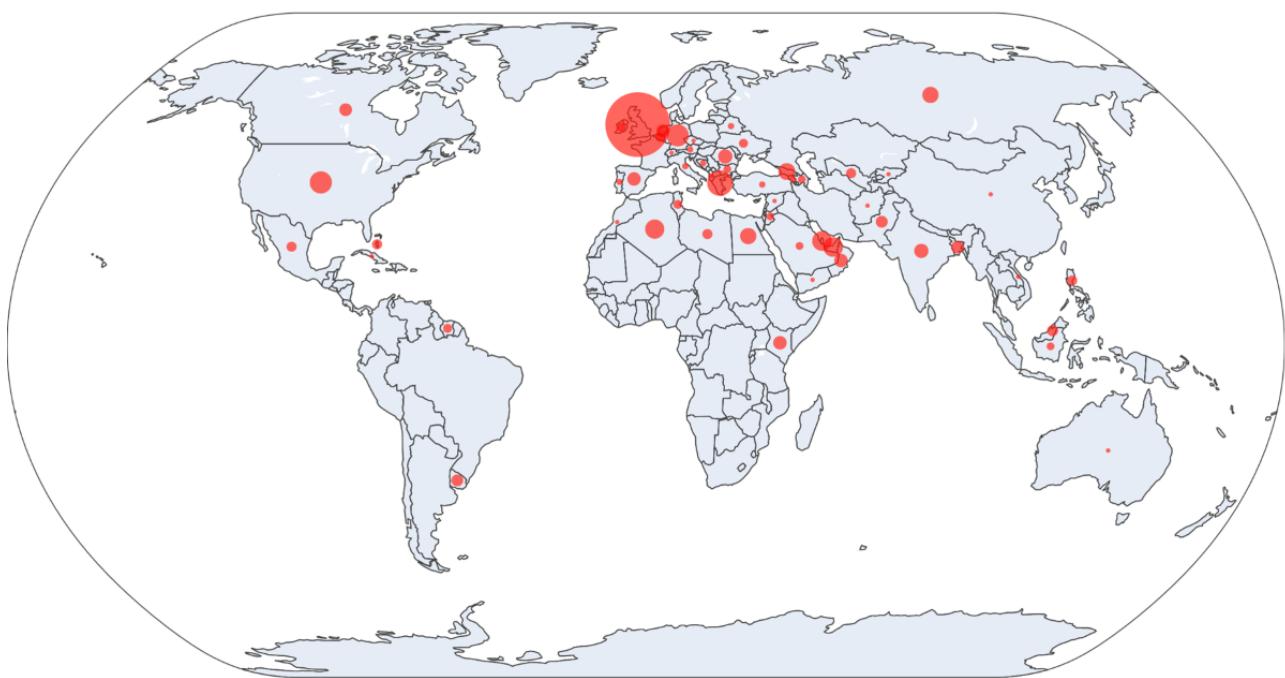
From August 2022 to September 2025, a total of 567 surveys were completed. Of those, 286 (50.4%) were in the category pediatric endocrinology, 225 (39.7%) were in the category pediatric diabetes, and 56 (9.9%) were in the category pediatric endocrinology in RLS.

Learner's Background

There was global participation, but most learners were practicing in Europe (n=333, 59%), followed by Asia (n=124, 22%), Africa (n=53, 9%), the Americas (North America, n=45, 8%; and South America, n=11, 2%), and Oceania (n=1, 0%; [Figure 1](#) and

Multimedia Appendix 2). Although the RLS courses are available in Mandarin, there was no participation from China.

Figure 1. Global distribution of participants completing online continuing medical education modules. Each bubble represents a country, with bubble size proportional to the number of completed modules. Red semitransparent bubbles are projected onto a natural-earth world map, with representative bubble sizes (n=5, 10, and 100) shown for scale.

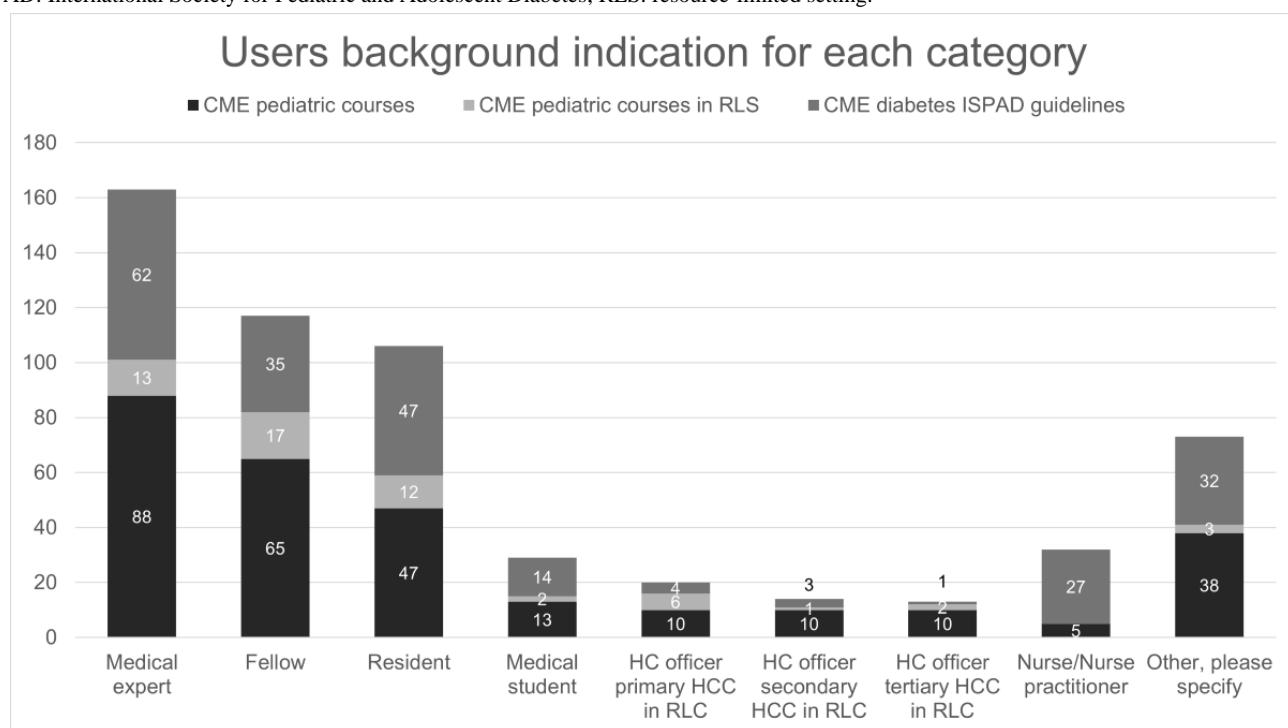


Most users indicated that they were medical experts (210/567, 37%), followed by fellows or residents (223/567, 39%). Medical students and nurses formed 5% and 6%, respectively, of the total group, and 13% were categorized under *other* (Figure 2). Relevant specifications given in the category *other* include

lecturer, psychologist, patient advocate, diabetes educator, nutritionist, and education support worker.

Twelve percent (n=70/567) of learners practice in RLS countries, of which 20 work as health care officers in primary health care centers, 14 work in secondary health care centers, and 13 work in tertiary health care centers [14].

Figure 2. Representation of learners according to training background. CME: continuing medical education; HC: health care; HCC: health care center; ISPAD: International Society for Pediatric and Adolescent Diabetes; RLS: resource-limited setting.



Feedback on CME Modules

Across all survey items, overall learner satisfaction was high, with most responses falling within the *agree* or *strongly agree* categories (Table 1, Figure 3). The median scores for the majority of questions were 4.0, with means consistently above 4.0 for questions relating to content quality, clarity, motivation, competence, and perceived educational value. The feedback on the questions' difficulty level was also narrow, with the majority of responses falling into the *appropriate* category. The PE module generally received the highest mean ratings across items, whereas the RLS and PD modules showed similar, though slightly lower, response patterns. IQRs were narrow for most

questions, indicating highly compressed distributions with limited variability. Kruskal-Wallis testing revealed statistically significant differences between modules for most survey items ($P<.05$), primarily due to a higher proportion of strongly positive responses in the PE module. Exceptions included the question on perceived improvement in performance, where module differences were not statistically significant. Items relating to preferences for additional open-ended questions (eg, bullet list and teacher feedback formats) received comparatively lower mean scores and showed greater variability. Overall, the findings indicate strong learner engagement and perceived educational value, with some variation in response patterns across modules (Table 1, Figure 3).

Table . Descriptive and comparative statistics for all Likert-scale survey questions across the three modules (pediatric endocrinology [PE], n=283; PE in resource-limited settings [RLS], n=56; and pediatric diabetes [PD], n=228).

Survey question ^a and module	Likert score responses, n					Comparative statistics		
	1	2	3	4	5	Median (IQR)	Mean (SD)	P value
The online format was appropriate for the subject matter and I was able to access all components of the activity without difficulty. ^b								<.001
PE	0	5	9	137	132	4.0 (4.0-5.0)	4.40 (0.64)	
RLS	0	0	5	34	17	4.0 (4.0-5.0)	4.21 (0.59)	
PD	0	1	16	146	65	4.0 (4.0-5.0)	4.21 (0.58)	
The material was organized clearly for learning to occur. ^b								<.001
PE	0	4	9	138	132	4.0 (4.0-5.0)	4.41 (0.63)	
RLS	0	0	5	37	14	4.0 (4.0-4.2)	4.16 (0.57)	
PD	0	2	17	143	66	4.0 (4.0-5.0)	4.20 (0.60)	
The content of this chapter or case is interesting to me. ^b								<.001
PE	0	1	18	122	142	5.0 (4.0-5.0)	4.43 (0.63)	
RLS	0	1	5	35	15	4.0 (4.0-5.0)	4.14 (0.65)	
PD	0	1	14	139	74	4.0 (4.0-5.0)	4.25 (0.58)	
After studying this chapter or case, I feel motivated to learn more on the subject. ^b								<.001
PE	0	4	18	118	143	5.0 (4.0-5.0)	4.41 (0.68)	
RLS	0	0	5	34	17	4.0 (4.0-5.0)	4.21 (0.59)	
PD	0	1	18	141	68	4.0 (4.0-5.0)	4.21 (0.59)	
This activity will assist in the improvement of my competence. ^b								.02
PE	0	0	13	140	130	4.0 (4.0-5.0)	4.41 (0.58)	
RLS	0	0	2	38	16	4.0 (4.0-5.0)	4.25 (0.51)	
PD	0	1	11	134	82	4.0 (4.0-5.0)	4.30 (0.58)	
This activity will assist in the improvement of my performance. ^b								.12
PE	0	0	20	139	124	4.0 (4.0-5.0)	4.37 (0.61)	
RLS	0	0	2	38	16	4.0 (4.0-5.0)	4.25 (0.51)	
PD	0	1	12	135	80	4.0 (4.0-5.0)	4.29 (0.58)	
This activity will assist in the improvement of my patient outcomes. ^b								<.001
PE	0	0	26	128	129	4.0 (4.0-5.0)	4.36 (0.65)	
RLS	0	0	2	38	16	4.0 (4.0-5.0)	4.25 (0.51)	
PD	0	2	19	125	82	4.0 (4.0-5.0)	4.26 (0.64)	
The content and questions align with my knowledge level. ^b								<.001
PE	2	8	25	176	72	4.0 (4.0-5.0)	4.09 (0.72)	
RLS	0	0	5	45	6	4.0 (4.0-4.0)	4.02 (0.45)	
PD	0	4	19	156	49	4.0 (4.0-4.0)	4.10 (0.60)	
I like self-assessment with multiple choice questions. ^b								.02
PE	0	5	17	145	116	4.0 (4.0-5.0)	4.31 (0.67)	
RLS	0	3	4	38	11	4.0 (4.0-4.0)	4.02 (0.70)	
PD	0	1	27	142	58	4.0 (4.0-5.0)	4.13 (0.61)	
The feedback that is given after answering the questions is appropriate. ^b								<.001

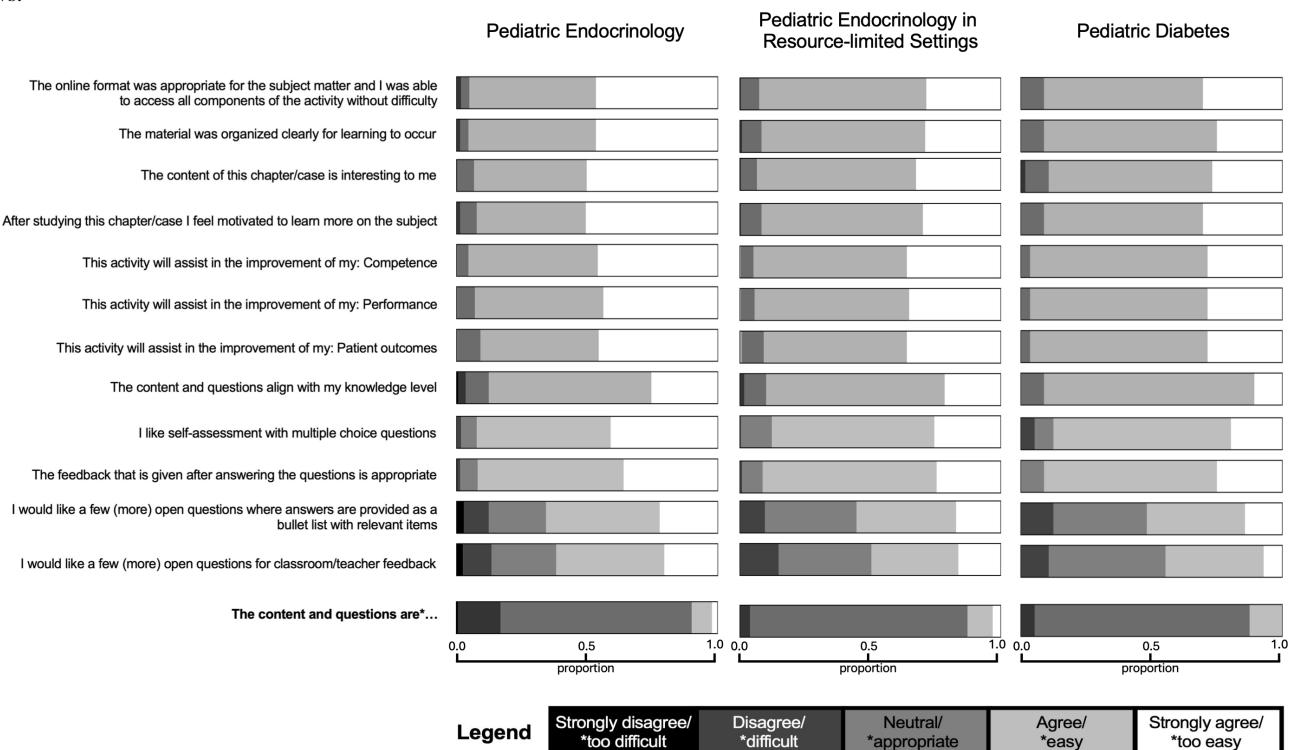
Survey ques- tion ^a and mod- ules	Likert score responses, n					Comparative statistics		
	1	2	3	4	5	Median (IQR)	Mean (SD)	P value
PE	1	3	19	158	102	4.0 (4.0-5.0)	4.26 (0.66)	
RLS	0	0	5	37	14	4.0 (4.0-4.2)	4.16 (0.57)	
PD	0	2	18	152	56	4.0 (4.0-4.0)	4.15 (0.58)	
I would like a few (more) open questions where answers are provided as a bullet list with relevant items. ^b								<.001
PE	8	27	62	123	63	4.0 (3.0-4.0)	3.73 (1.00)	
RLS	0	7	20	21	8	4.0 (3.0-4.0)	3.54 (0.89)	
PD	0	22	80	87	39	4.0 (3.0-4.0)	3.63 (0.88)	
I would like a few (more) open questions for classroom/teacher feedback. ^b								.02
PE	7	31	70	117	58	4.0 (3.0-4.0)	3.66 (1.00)	
RLS	0	6	25	21	4	3.0 (3.0-4.0)	3.41 (0.78)	
PD	0	34	81	76	37	3.0 (3.0-4.0)	3.51 (0.94)	
The content and questions are: ^c								<.001
PE	2	46	207	22	6	3.0 (3.0-3.0)	2.94 (0.59)	
RLS	0	3	46	7	0	3.0 (3.0-3.0)	3.07 (0.42)	
PD	0	9	190	22	7	3.0 (3.0-3.0)	3.12 (0.50)	

^aFor each question, response frequencies on a 5-point Likert scale were expanded into individual-level datasets to compute descriptive statistics, including total responses, median, mean, IQR, and SD. Differences in score distributions across modules were assessed using the Kruskal-Wallis H test, appropriate for ordinal, nonnormally distributed data. *P* values <.05 were considered statistically significant and are indicated by italics. Higher scores reflect more favorable evaluations.

^bLikert scale responses: 1=strongly disagree; 2=disagree; 3=neutral; 4=agree; and 5=strongly agree.

^cLikert scale responses: 1=too easy; 2=easy; 3=appropriate; 4=difficult; and 5=too difficult.

Figure 3. Stacked Likert distribution of responses across all survey questions. This figure displays the proportional distribution of Likert-scale responses, separate for each module (pediatric endocrinology, pediatric endocrinology in resource-limited settings, and pediatric diabetes). Each horizontal bar represents one survey question, subdivided into gray-shaded coded segments corresponding to the five Likert response categories. Segment widths reflect the proportion of total responses in each category. The plot illustrates the overall positive skew of responses across the curriculum, with most items demonstrating high levels of agreement (Likert 4 - 5), while specific questions show greater dispersion, indicating more heterogeneous participant views.



Not all participants provided free-text feedback. [Table 2](#) illustrates the distribution of free-text feedback received, categorized into four main categories, with the largest being various forms of appraisal. Improvement suggestions, such as requesting more interactive cases and algorithms, are encouraging. Layout issues are helpful to guide improvements,

and similarly, critical comments about the difficulty of the content and excessive text are incentives to indicate more clearly the level of the course (core vs advanced), and replacing text-heavy chapters with bullet styles using artificial intelligence diligently.

Table . Free-text feedback from completed continuing medical education modules.

Domains	Overall (N=384), n (%)	PE ^a (n=197), n (%)	RLS ^b (N=26), n (%)	PD ^c (n=161), n (%)	Example comments
Praise or approval	274 (71.3)	145 (73.6)	21 (80.8)	108 (67.1)	Content is appropriate, good, excellent, perfect, helpful, informative, useful, adds knowledge and scientific thinking
Content suggestions	47 (12.2)	27 (13.7)	3 (11.5)	17 (10.6)	Needs more interactive cases, more algorithms, more detailed information, better explanations of multiple-choice answers, and more video content
Layout suggestions	28 (7.3)	13 (6.6)	2 (7.7)	13 (8.1)	Buttons need to be bigger, multiple replies in multiple-choice questions unclear, minimize menu left side of screen, and typos and missing sentences
Criticism or disapproval	35 (9.1)	12 (6.1)	0	23 (14.2)	Slightly <i>dry</i> , too difficult, too much text, too many references, and inappropriate for my work setting

^aPE: pediatric endocrinology.

^bRLS: PE in resource-limited setting.

^cPD: pediatric diabetes.

Discussion

In an era where equitable access to CME is increasingly important, this study highlights the global uptake and positive reception of the ESPE CME-accredited e-learning modules made freely available in PE and PD, particularly among professionals in both high-resource and RLSs.

Principal Findings

Over a 3-year period, 567 learners from five continents completed postcourse surveys. The majority of participants were medical experts or trainees. Encouragingly, there were also learners practicing in RLSs. Feedback was overall very positive, with the majority of responses scoring 4 or 5 on a 5-point Likert scale. Free-text feedback suggested areas for improvement, including reducing text-heavy content and increasing interactivity and visual elements.

Implications of Findings

These findings support the value of free, guideline-based CME resources in PE and PD, particularly in expanding access to continuing education globally. The content is not only in line with the training requirements in PE and PD [15] but is also complementary to the requirements of the neonatology training curriculum [16]. Moreover, it is relevant for internists dealing with former pediatric patients with chronic congenital or acquired pediatric disorders [17].

A prerequisite of self-directed learning is that learners are internally motivated to take responsibility for their learning through a process in which they identify their own learning needs, use a variety of resources to meet these needs, and evaluate their learning to ensure that their learning needs have been met [18].

The high satisfaction ratings suggest that the ESPE platform meets these self-directed learning needs of a diverse professional audience. The engagement of health professionals, such as nurses and dietitians, also reflects the growing multidisciplinary nature of pediatric endocrine care. The feedback received provides actionable insights to improve the user experience and tailor content to different learning styles and clinical contexts.

Learners' free-text feedback has provided actionable insights for improving the ESPE e-learning modules. Common suggestions, such as adding more interactive cases, in particular when content is text-heavy; simplifying algorithms; and reducing text density, have informed content updates. The ESPE e-Learning website is moving to a new online platform, which allows for more visually engaging content and the use of artificial intelligence, which we anticipate will enhance learners' experience, engagement, and accessibility across varying levels of training. Categorizing feedback into themes also enables the committee to track trends and evaluate the impact of changes over time, ensuring the platform evolves in response to user needs.

Comparison With Prior Work

The EACCME implemented criteria for the accreditation of e-learning materials in 2009 and conducted an audit of CME—Continuous Professional Development (events between 2017 and 2019), which included 385 e-learning materials [19]. However, no details on the assessment of these e-learning materials are provided, and studies focusing on accreditation of self-directed interactive asynchronous online learning designs are scarce [20,21]. Blomberg et al [22] highlighted a post-COVID-19 pandemic shift toward hybrid and digital CME formats, emphasizing accessibility and learner-centered design. Curran et al [20] concluded that an accredited, asynchronous e-learning module provides flexibility, accessibility, and scalability, particularly for geographically dispersed health professionals and described a structured approach to developing accredited e-learning modules, underscoring the importance of stakeholder engagement and iterative content development. Additionally, Tudor Car et al [23] demonstrated that digital problem-based learning is as effective as traditional methods for knowledge acquisition and may be superior for skill development.

Our results are consistent with earlier evaluations of the ESPE platform. Ng et al [4] reported sustained global engagement with the nonaccredited modules over a 10-year period. Kalaitzoglou et al [10] demonstrated the effectiveness of multilingual modules in improving access in RLSs, which aligns with our finding that 9% of users came from such contexts. Prior studies by Drop et al [5] and Kranenburg-van Koppen et al [6] emphasized the importance of interactive, case-based learning, a preference echoed in our learners' feedback. Our study adds to this literature by focusing specifically on the CME-accredited modules and providing structured postcourse feedback data.

This supports the notion that digital CME platforms, such as the ESPE e-learning portal, are well positioned to meet evolving global educational needs in PE and PD.

Strengths and Limitations

A key strength of this study is its global scope and inclusion of diverse professional roles, offering a comprehensive view of how the CME modules are used and perceived. The survey's mandatory nature ensured a high response rate. However, the study has limitations. It relies on self-reported data, which may be subject to response bias. The absence of pre- and post-module assessments limits our ability to measure learning outcomes objectively. Additionally, participation from certain regions (eg, China, Brazil, and Scandinavia) was lacking. Several factors may explain the lack of participation from certain regions. First, national CME accreditation policies vary widely; some countries may not recognize EACCME credits or may not require CME for specialist revalidation, reducing the incentive to engage with external platforms. Second, language barriers may persist despite the facilitating multilingual content, particularly if promotional materials or user interfaces are not localized. Third, awareness and visibility of the platform may be limited in regions where ESPE or ISPAD has fewer members. Fourth, cultural preferences for in-person or locally developed educational resources may influence uptake. Finally, technological access

and digital literacy may still pose challenges in some areas, particularly in rural or underserved settings [24]. These insights underscore the need for targeted outreach strategies, local ambassadors, and region-specific promotional campaigns to improve awareness and adoption. Several authors emphasize the importance of partnerships with international societies, such as ESPE and ISPAD, and the support by their organizational infrastructure as of critical importance [4,22,25,26].

The accreditation credit points (European CME credit) are based on educational time spent and not on educational impact. In several papers, transitioning CME crediting from being time based to impact based is advocated with a plea for more rigorous quantitative or qualitative assessments. The assessment in the format of surveys should be appropriate to the intended goals or outcomes of the accredited education, measuring improvements in learner knowledge, skills, and competencies, professional performance, and ultimately in changes in patient health status [19,27,28]. The International Academy for Continuing Professional Development Accreditation created a shared set of international standards for accrediting CME [28]. De Leeuw et al [27] constructed a validated 7-step Medical E-Learning Evaluation Survey. The authors conclude that the Medical E-Learning Evaluation Survey is useful and understandable, adding value for e-learning creators. However, it is impossible to predict how motivated users will be to provide useful feedback. Therefore, a survey should always be accompanied by an in-depth focus group evaluation with the users. The adoption of international standards would establish a global framework, and substantive equivalency would enable international collaboration, reciprocity of credits, and broader access to high-quality education for health care professionals [28].

To enhance the platform's impact, future efforts should focus on developing pre- and postmodule assessments to evaluate knowledge gain, increasing interactivity through multimedia content and case simulations, strengthening outreach in underrepresented regions by partnering with local societies and ambassadors, and exploring the long-term impact of the modules on clinical practice and patient outcomes. The ESPE-ISPAD e-Learning platform has the potential to serve as a model for other specialties seeking to deliver equitable, high-quality CME globally.

Conclusions

The ESPE-ISPAD CME-accredited e-learning modules are widely accessed and well received by health care professionals globally, offering free, high-quality education in PE and PD. Moreover, the asynchronous self-paced designs afford greater convenience and flexibility for providers in accessing the CME courses at times that are best for them [26].

Learners reported high satisfaction with the content and delivery, and feedback highlighted opportunities to further enhance interactivity and accessibility.

These findings support the continued development and promotion of open-access CME platforms to improve global equity in specialist medical education. Ongoing efforts are needed to expand outreach, refine content based on user

feedback, and make accreditation more dynamic, evidence-based, and focused on educational impact.

Acknowledgments

The authors would like to thank contributing authors from many countries who have given many hours of their valuable time to provide this outstanding educational material. The authors gratefully acknowledge the support of the European Society for Paediatric Endocrinology (ESPE) and the International Society for Pediatric and Adolescent Diabetes (ISPAD) for facilitating the provision of the European Union of Medical Specialists (EUMS) CME certifications for the participants free of charge. They would also like to acknowledge QuestionPro for providing a nonprofit license. Apart from the free survey software, they also obtained access to QuestionPro's free survey templates.

Funding

The espe-elearning.org online learning website has received financial support from ESPE; ISPAD; Kyowa Kirin, Japan; Sandoz International, Germany; Alexion Pharmaceuticals, United States; and the European Union's Seventh Framework Program (FP7/2007-2013) under grant agreement 2014444. RLC Project has received financial support from ESPE, ISPAD, and Eli Lilly & Co, United States.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: YGVDZ, CVW-DV, JI

Data curation: CVW-DV

Formal analysis: CVW-DV, YGVDZ, JI

Funding acquisition: SMN, SLS, JI

Methodology: YGVDZ, CVW-DV, JI

Project administration: CVW-DV

Software: CVW-DV, JI

Validation: YGVDZ, JI

Visualization: CVW-DV, JI, YGVDZ

Writing – original draft: YGVDZ, CVW-DV, SLS, JI

Writing – review & editing: all authors

Conflicts of Interest

SMN is the current Chair of both ESPE and ISPAD E-learning Committees.

AMB is a past chair of the ESPE e-Learning Committee. SLS is a past member, treasurer, and president of ESPE Council; a former member of the ESPE Caucasus and Central Asia School; and a former chief editor of espe-elearning.org

Multimedia Appendix 1

Content of the continuing medical education feedback survey.

[[DOCX File, 27 KB - mededu_v12i1e67332_app1.docx](#)]

Multimedia Appendix 2

Continuing medical education courses completed.

[[XLSX File, 10 KB - mededu_v12i1e67332_app2.xlsx](#)]

References

1. ESPE e-Learning. URL: www.espe-elearning.org [accessed 2025-09-30]
2. Grijpink-van den Biggelaar K, Drop SLS, Schuwirth L. Development of an e-learning portal for pediatric endocrinology: educational considerations. *Horm Res Paediatr* 2010;73(3):223-230. [doi: [10.1159/000284366](https://doi.org/10.1159/000284366)] [Medline: [20197677](#)]
3. Ng SM, de Beaufort C, Becker D, Wolfsdorf JI. A global e-learning initiative for pediatric diabetes and endocrinology: introduction and description. *Pediatr Diabetes* 2021 Jun;22(4):692-694. [doi: [10.1111/pedi.13214](https://doi.org/10.1111/pedi.13214)] [Medline: [33904237](#)]
4. Ng SM, Kalaitzoglou E, Utari A, et al. Ten-year experience of a global and freely accessible e-learning website for pediatric endocrinology and diabetes. *Horm Res Paediatr* 2023;96(4):366-375. [doi: [10.1159/000527984](https://doi.org/10.1159/000527984)] [Medline: [36349753](#)]
5. Drop SLS, Kranenburg L, Reerds S, Grijpink K, Iotova V. The ESPE e-learning webportal: a global tool for instruction and formative assessment of pediatric endocrinology. *Scr Sci Med* 2014;46(3):11-18. [doi: [10.14748/ssm.v46i3.760](https://doi.org/10.14748/ssm.v46i3.760)]

6. Kranenburg-van Koppen LJC, Grijpink-van den Biggelaar K, Drop SLS. An interactive E-learning portal in pediatric endocrinology: practical experience. *Pharmacy (Basel)* 2013;1(2):160-171. [doi: [10.3390/pharmacy1020160](https://doi.org/10.3390/pharmacy1020160)]
7. Alzahrani M. Traditional learning compared to online learning during the COVID-19 pandemic: lessons learned from faculty's perspectives. *Sage Open* 2022 Apr;12(2):21582440221091720. [doi: [10.1177/21582440221091720](https://doi.org/10.1177/21582440221091720)]
8. Brika SKM, Chergui K, Algamdi A, Musa AA, Zouaghi R. E-learning research trends in higher education in light of COVID-19: a bibliometric analysis. *Front Psychol* 2022;12:762819. [doi: [10.3389/fpsyg.2021.762819](https://doi.org/10.3389/fpsyg.2021.762819)] [Medline: [35308075](#)]
9. 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](https://doi.org/10.1001/jama.300.10.1181)] [Medline: [18780847](#)]
10. Kalaitzoglou E, Majaliwa E, Zacharin M, et al. Multilingual global E-learning pediatric endocrinology and diabetes curriculum for front line health care providers in resource-limited countries: development study. *JMIR Form Res* 2020 Nov 5;4(11):e18555. [doi: [10.2196/18555](https://doi.org/10.2196/18555)] [Medline: [33151156](#)]
11. 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](https://doi.org/10.2196/jmir.6.3.e34)] [Medline: [15471760](#)]
12. Smidt A, Balandin S, Sigafoos J, Reed VA. The Kirkpatrick model: a useful tool for evaluating training outcomes. *J Intellect Dev Disabil* 2009 Sep;34(3):266-274. [doi: [10.1080/13668250903093125](https://doi.org/10.1080/13668250903093125)] [Medline: [19681007](#)]
13. Heydari MR, Taghva F, Amini M, Delavari S. Using Kirkpatrick's model to measure the effect of a new teaching and learning methods workshop for health care staff. *BMC Res Notes* 2019 Jul 10;12(1):388. [doi: [10.1186/s13104-019-4421-y](https://doi.org/10.1186/s13104-019-4421-y)] [Medline: [31292006](#)]
14. Khan T, Abimbola S, Kyobutungi C, Pai M. How we classify countries and people-and why it matters. *BMJ Glob Health* 2022 Jun;7(6):e009704. [doi: [10.1136/bmjgh-2022-009704](https://doi.org/10.1136/bmjgh-2022-009704)] [Medline: [35672117](#)]
15. Busiah K, Peet A, Tornese G, et al. The 2021 European training requirements in paediatric endocrinology and diabetes. *Horm Res Paediatr* 2021;94(11-12):441-447. [doi: [10.1159/000520073](https://doi.org/10.1159/000520073)] [Medline: [34638127](#)]
16. Santoro D, Zibulsky DA, Roehr CC, et al. Meeting the need for effective and standardized neonatology training: a pan-European Master's Curriculum. *Pediatr Res* 2024 Oct;96(5):1195-1200. [doi: [10.1038/s41390-024-03182-8](https://doi.org/10.1038/s41390-024-03182-8)] [Medline: [38702380](#)]
17. Luger A, Durkan M, Christ-Crain M, Marques P. European society of Endocrinology curriculum and training recommendation in Endocrinology. *Eur J Endocrinol* 2024 Oct 29;191(5):R71-R77. [doi: [10.1093/ejendo/lvae141](https://doi.org/10.1093/ejendo/lvae141)] [Medline: [39582372](#)]
18. Zhu M, Bonk CJ, Berri S. Fostering self-directed learning in MOOCs: motivation, learning strategies, and instruction. *Online Learn* 2022;26(1):153-173. [doi: [10.24059/olj.v26i1.2629](https://doi.org/10.24059/olj.v26i1.2629)]
19. Sideris M, Rallis KS, Nicolaides M, et al. The UEMS experience in continuous medical education accreditation process: a "quo vadis" analysis of our global database. *Ann Med Surg (Lond)* 2024 Feb;86(2):689-696. [doi: [10.1097/MS9.0000000000001583](https://doi.org/10.1097/MS9.0000000000001583)] [Medline: [38333329](#)]
20. Curran V, Glynn R, Whitton C, Hollett A. An approach to the design and development of an accredited continuing professional development e-learning module on Virtual Care. *JMIR Med Educ* 2024 Aug 8;10:e52906. [doi: [10.2196/52906](https://doi.org/10.2196/52906)] [Medline: [39119741](#)]
21. Fiuzzi M. Outcomes and observations of on-line CME activities during the pandemic. *J CME* 2023;12(1):2167286. [doi: [10.1080/28338073.2023.2167286](https://doi.org/10.1080/28338073.2023.2167286)] [Medline: [36969485](#)]
22. Blomberg D, Stephenson C, Atkinson T, et al. Continuing medical education in the post COVID-19 pandemic era. *JMIR Med Educ* 2023 Nov 15;9:e49825. [doi: [10.2196/49825](https://doi.org/10.2196/49825)] [Medline: [37966881](#)]
23. Tudor Car L, Kyaw BM, Dunleavy G, et al. Digital problem-based learning in health professions: systematic review and meta-analysis by the digital health education collaboration. *J Med Internet Res* 2019 Feb 28;21(2):e12945. [doi: [10.2196/12945](https://doi.org/10.2196/12945)] [Medline: [30816846](#)]
24. Fraillon J, Rožman M, editors. IEA International Computer and Information Literacy Study 2023: Assessment Framework: Springer; 2025. [doi: [10.1007/978-3-031-61194-0](https://doi.org/10.1007/978-3-031-61194-0)]
25. Craig A, Lawford H, Miller M, et al. Use of technology to support health care providers delivering care in low- and lower-middle-income countries: systematic umbrella review. *J Med Internet Res* 2025 Jun 18;27:e66288. [doi: [10.2196/66288](https://doi.org/10.2196/66288)] [Medline: [40533075](#)]
26. Roth MJ, Maggio LA, Costello JA, Samuel A. E-learning interventions for quality improvement continuing medical education—a scoping review. *J Contin Educ Health Prof* 2025 Jan 1;45(1):35-43. [doi: [10.1097/CEH.0000000000000564](https://doi.org/10.1097/CEH.0000000000000564)] [Medline: [39028318](#)]
27. de Leeuw RA, Westerman M, Walsh K, Scheele F. Development of an instructional design evaluation survey for postgraduate medical e-learning: content validation study. *J Med Internet Res* 2019 Aug 9;21(8):e13921. [doi: [10.2196/13921](https://doi.org/10.2196/13921)] [Medline: [31400102](#)]
28. Regnier K, Campbell C, Griebenow R, et al. Standards for substantive equivalency between continuing professional development/continuing medical education (CPD/CME) accreditation systems. *J Eur CME* 2021 Jan 29;10(1):1874644. [doi: [10.1080/21614083.2021.1874644](https://doi.org/10.1080/21614083.2021.1874644)] [Medline: [33552679](#)]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys

CME: continuing medical education

EACCME: European Accreditation Council for Continuing Medical Education

ESPE: European Society for Paediatric Endocrinology

ISPAD: International Society for Pediatric and Adolescent Diabetes

PD: pediatric diabetes

PE: pediatric endocrinology

RLS: resource-limited setting

Edited by B Lesselroth; submitted 12.Oct.2024; peer-reviewed by J Sarvestan, J Woldorf, Y Shen; revised version received 24.Nov.2025; accepted 03.Dec.2025; published 30.Jan.2026.

Please cite as:

van der Zwan YG, van Wijngaard-de Vugt C, Abu-Libdeh A, Kalaitzoglou E, Karabouta Z, Drop SLS, Boot AM, Ng SM, Idkowiak J. Global Learner Feedback on Continuing Medical Education—Accredited e-Learning Modules in Pediatric Endocrinology and Diabetes: Cross-Sectional Study. JMIR Med Educ 2026;12:e67332

URL: <https://mededu.jmir.org/2026/1/e67332>

doi:10.2196/67332

© Yvonne G van der Zwan, Conny van Wijngaard-de Vugt, Abdulsalam Abu-Libdeh, Evangelia Kalaitzoglou, Zacharoula Karabouta, Stenvert L S Drop, Annemieke M Boot, Sze May Ng, Jan Idkowiak. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 30.Jan.2026. 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.

Original Paper

Impact of Community-Oriented Medical Education on Medical Students' Perceptions of Community Health Care: Qualitative Study

Kiyoshi Shikino¹, MD, MHPE, PhD; Kazuyo Yamauchi^{1,2}, MD, MHPE, PhD; Nobuyuki Araki¹, MD, PhD; Naoto Ozaki¹, MD; Yu Kamata¹, MD; Shinya Aoki^{1,3}, MD; Yota Katsuyama^{1,4}, MD, PhD; Daichi Sogai^{1,5}, MD; Mai Miyamoto^{1,6}, MD; Kensuke Yoshimura⁷, MD, MPH, PhD; Takeshi Oki⁸, MD, PhD; Shoichi Ito^{1,9}, MD, PhD

¹Department of Community-Oriented Medical Education, Graduate School of Medicine, Chiba University, Chiba, Japan

²Department of Medical Education, Dokkyo Medical University, Tochigi, Japan

³Department of General Medicine, Shioda Hospital, Katsuura, Japan

⁴Sanbunomori Clinic, Sanmu, Japan

⁵Department of General Medicine, Sanmu Medical Center, Sanmu, Japan

⁶Department of Internal Medicine, Isumi Medical Center, Isumi, Japan

⁷Center for Next Generation of Community Health, Chiba University Hospital, Chiba, Japan

⁸Trinity Neurology Clinic, Sakura, Japan

⁹Department of Medical Education, Graduate School of Medicine, Chiba University, Chiba, Japan

Corresponding Author:

Kiyoshi Shikino, MD, MHPE, PhD

Department of Community-Oriented Medical Education

Graduate School of Medicine

Chiba University

1-8-1, Inohana

Chu-ou-ku

Chiba, 2608677

Japan

Phone: 81 43 222 7171

Email: kshikino@gmail.com

Abstract

Background: Physician maldistribution remains a global challenge, with Japan's rural regions facing critical health care shortages. Regional quota programs aim to attract medical students to underserved areas; however, their effectiveness in fostering long-term commitment is uncertain. Community-oriented medical education (COME) programs aim to address this issue by developing students' understanding and dedication to rural health care.

Objective: This study investigated the impact of an enhanced COME program, featuring increased early clinical exposure and faculty development, on first-year regional quota medical students' perception of community health care at Chiba University.

Methods: We conducted a cross-sectional qualitative study comparing 2 cohorts, 20 students enrolled from the existing COME course (April-December 2021) and 20 from the revised course (April-December 2022). The revised course included an additional day of community-based clinical exposure supervised by COME-trained attending physicians. Students' written reflections were analyzed using qualitative content analysis and categorized according to the Fink Taxonomy of significant learning, comprising 6 domains, including foundational knowledge, application, integration, human dimension, caring, and learning how to learn. Reflections were synthesized into higher-order themes crosswalked to the Fink domains.

Results: Demographics were similar between the 2021 and 2022 cohorts. In 2021, 311 learning codes were identified across foundational knowledge (n=128), application (n=91), integration (n=40), human dimension (n=16), caring (n=30), and learning how to learn (n=6). In 2022, codes increased to 385, with notable growth in caring (n=58) and human dimension (n=57), alongside increases in learning how to learn (n=15) and integration (n=45). Theme-based synthesis identified four overarching themes: (1) community health care as an interconnected, resource-constrained system; (2) patient-centered relationships and trust through communication and teamwork; (3) emerging professional identity and responsibility toward community service; and (4) developing

a self-directed learning orientation for community practice. Qualitative analysis revealed that students gained a deeper understanding of patient-centered care, interprofessional collaboration, and social challenges in rural health care. The consistency in the foundational knowledge domain underscored a stable conceptual foundation, while the increase in affective and reflective domains reflected greater emphasis on interpersonal, value-oriented, and reflective learning in the revised cohort.

Conclusions: Enhancements of the COME program, including additional early clinical exposure and faculty development, were associated with improved students' perceptions of community health care. The increased focus on the caring and human dimension domains underscores the role of practical experiences in fostering collaboration, communication, and patient-centered care. The theme-based synthesis further suggests that the revised program prompted more frequent reflections on professional identity formation and self-directed learning while maintaining a stable foundation of community health care concepts. Mentorship by community hospital attendings, alongside structured clinical exposure, appears crucial in shaping medical students' understanding and commitment to rural medicine. Ongoing longitudinal evaluations are warranted to assess the sustained impact of COME programs on career trajectories in underserved areas.

(*JMIR Med Educ* 2026;12:e84406) doi:10.2196/84406

KEYWORDS

community health care; community-based medical education; community-oriented medical education; physician maldistribution; career perceptions; community hospital; community education; faculty development

Introduction

Physician maldistribution is a global concern affecting many nations [1-6]. In Japan, rural regions are facing a critical shortage of health care providers [7,8] due to urbanization [9-11]. To address this issue, numerous Japanese medical schools have introduced regional quota programs designed to motivate students to practice community or rural medicine [7,8,12]. These programs offer scholarships tied to service agreements in designated regions to address the scarcity of physicians. As of 2015, 70 out of 79 Japanese medical schools offered regional quota programs, enrolling 1541 students [7].

However, the effectiveness and ethical implications of these programs are debatable. Students assigned to regional quotas may become less inclined to work in their designated regions as they progress through their academic programs [8]. Therefore, it is essential to develop community-oriented medical education (COME) courses that not only impart knowledge but also cultivate students' commitment to serving in underserved communities [13].

The Japanese government has high expectations for community-based medical education (CBME) and has made it a compulsory component of medical training [13-15]. Through direct participation in CBME, medical students and resident physicians can acquire practical skills, knowledge, and attitudes pertinent to clinical practice [16]. Nonetheless, questions remain regarding the long-term sustainability and impact of these programs, including student retention in rural areas [14]. Thus, continuous evaluation and enhancement are critical to effectively mitigate health care provider shortages in rural areas in Japan [8].

Herein, we aimed to investigate the educational impact of an enhanced COME program, which incorporated increased early clinical exposure and faculty development for community hospital attendings, on first-year regional quota medical students at Chiba University. Specifically, we explored students' "career perceptions," defined as their developing understanding of professional identity, motivation toward community service,

and long-term orientation toward rural medical practice. We hypothesized that participation in the enhanced COME program would strengthen students' sense of professional responsibility and commitment to working in underserved areas by deepening their reflection on real-world community health care experiences. In this study, the term "career perceptions" refers to students' evolving understanding of their future professional roles, motivations, and attitudes toward community health care. These perceptions encompass not only students' intentions to work in rural or underserved settings but also their sense of social responsibility, empathy, and professional identity as future physicians.

Methods

Study Design

A cross-sectional survey was conducted to assess the influence of a COME program, designed by the Chiba University School of Medicine, on medical students' career perceptions. In this study, "career perceptions" were defined as students' reflective understanding of their professional identity, motivation toward community service, and long-term orientation toward rural medical practice. The survey was designed to collect qualitative data from students' written reflections rather than quantitative responses, allowing for an in-depth exploration of their perceptions and learning experiences. Qualitative research typically involves data in the form of words instead of numbers and is particularly appropriate for understanding participants' experiences, thoughts, and motivations [17].

This study encompassed 2 cohorts, including 20 first-year regional quota medical students from April to December 2021, representing the existing course structure, and 20 students from April to December 2022, representing a revised course with additional early clinical exposure. Before participating in the COME program, the students had already completed the interprofessional education (IPE) step 1 course earlier in the academic year, which may have influenced their perceptions and learning outcomes in the COME program.

Participants and Settings

Participants were first-year medical students enrolled in Chiba University under a regional quota system, participating in a year-long COME program. All participants were Japanese medical students enrolled in the regional quota program at Chiba University. The 2021 (n=20) and 2022 cohorts (n=20) each consisted of first-year students aged 18-24 years and 18-21 years, respectively. Sex assigned at birth (male or female) was collected as a demographic variable. This program included didactic sessions and practical experiences designed to enhance students' understanding of and commitment to community health care.

Selection Process for Regional Quota Students

Students applying to Chiba University School of Medicine must choose between applying under the general admission track or the regional quota track. Admission decisions are based on the scores from the common test for university admissions, a written examination, and an interview. Students admitted under the regional quota track are obligated to work within Chiba Prefecture for 9 years after graduation, fulfilling a service agreement. In return, they receive financial support through educational subsidies as an incentive. This system is intended to attract and retain health care professionals in underserved areas of the prefecture.

COME at Chiba University

Chiba Prefecture is one of the regions in Japan with a significant disparity in the geographical distribution of physicians, creating an urgent need to ensure the retention of medical professionals in community health care settings. In response, the COME program was introduced to address this issue through targeted human resource development. The program aims to cultivate medical students' understanding of and commitment to practicing in underserved areas, thereby contributing to sustainable health care solutions in the region.

Facilities and Roles of Community Hospital Attending

In 2022, the COME program was implemented across 4 facilities, each staffed with attending physicians dedicated to community-based medical education. These community hospital attendings play a critical role in educating medical students and creating an optimal learning environment. Additionally, they participate in weekly collaborative faculty development sessions at Chiba University, integrating practical experience in community health care with theoretical frameworks, ensuring a comprehensive approach to medical education in underserved areas.

IPE Step 1 Course

Before participating in the COME program, all students had completed the IPE Step 1 course, which establishes the necessary professional attitudes for health care practitioners and develops their ability to communicate effectively with patients, service users, and students from other faculties. At the end of the course, students are expected to reflect on the requirements

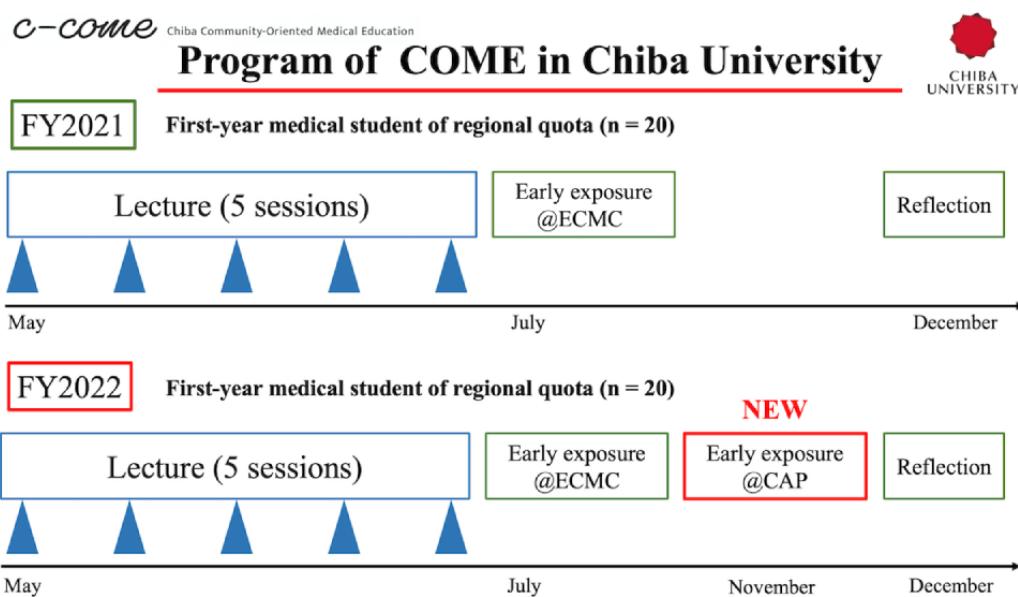
for their professional growth, contribute effectively to team efforts, and respect interprofessional collaboration. These competencies likely influenced students' perceptions and learning outcomes in the subsequent COME program.

Enhancements to the COME Program

In 2021, the COME program consisted of five lecture sessions, one day of early clinical exposure, and subsequent reflection sessions (Figure 1). During this day, students mainly observed outpatient and inpatient activities at Eastern Chiba Medical Center, attended multidisciplinary team meetings, and learned about hospital functions and community health care delivery. Their participation was primarily observational. In 2022, the program was enhanced with an additional day of clinical exposure, orchestrated by COME-trained attending physicians serving as community preceptors. This additional day was implemented at community hospitals where students shadowed COME-trained attendings, interacted directly with patients, and discussed diagnostic reasoning and care planning. Students also engaged in reflective discussions with their mentors, receiving formative feedback on communication and professionalism. Early clinical exposure allowed students to learn directly from physicians who served as supervisors and educational leaders. Before clinical exposure, attending physicians engaged in targeted mentorship through personalized communication via email or online conferences to understand each student's individual needs and learning objectives. The faculty development program, consisting of 40 sessions annually, equipped these physicians with pedagogical skills to provide personalized guidance throughout the medical education process. These attending physicians, strategically positioned in regions with physician shortages, underwent comprehensive faculty development. This development entailed 40 sessions annually, each lasting 2 hours, and covered various pedagogical topics, including educational technology, simulation teaching, and assessment methodologies. These sessions were part of a broader faculty development program to enhance the physicians' overall educational capabilities and equip them to provide personalized guidance throughout the students' medical education. Overall, the 2022 revision represented a shift from passive observation to active engagement and mentorship-based learning, which deepened students' understanding of community health care and enhanced their sense of professional identity. These enhancements aimed to provide students with practical experiences and mentorship that deepen their understanding of community health care and address the challenges faced in underserved areas [18].

In Figure 1, it can be found that, in 2021, the program comprised 5 lecture sessions, one early clinical exposure at Eastern Chiba Medical Center (ECMC), and reflection activities. In 2022, the revised program added a new early clinical exposure component at hospitals staffed by COME-trained attending physicians (CAP), integrating faculty development initiatives to enhance community-based mentorship.

Figure 1. Structure and timeline of the community-oriented medical education (COME) program at Chiba University for first-year regional quota medical students in 2021 and 2022. CAP: community-oriented medical education-trained attending physicians; ECMC: Eastern Chiba Medical Center; FY: final year.



Data Collection

Students' Reports

A qualitative inquiry was conducted. To measure the educational impact, we analyzed students' reports from 2021-2022 as qualitative data. These reports addressed 4 key questions regarding students' understanding of community medicine, health care challenges, their envisioned role in local health care, and the competencies required for these roles (Multimedia Appendix 1).

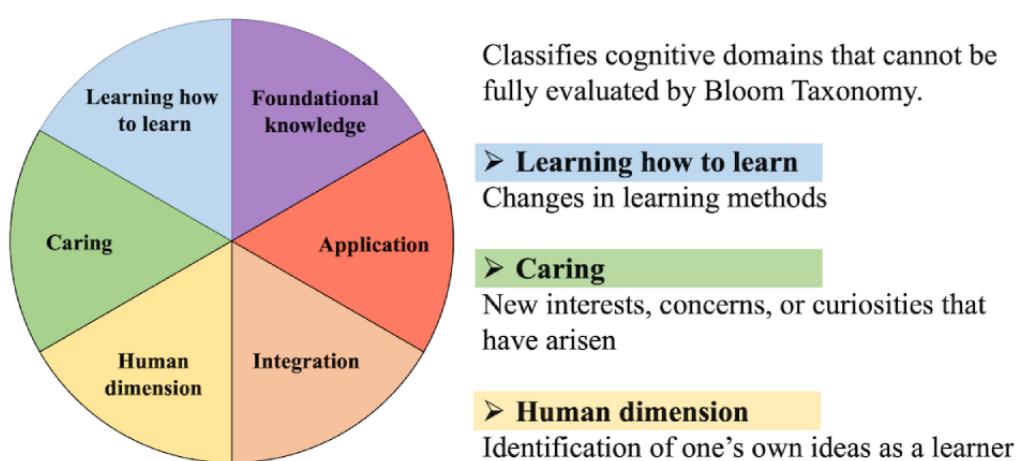
The Fink Taxonomy of Significant Learning

The Fink Taxonomy [19] is distinct in its categorization of cognitive domains, offering a broader scope than the Bloom

Taxonomy [12,20-22]. It emphasizes the “Human dimension”—students' self-awareness as learners; “Caring” the development of new interests and concerns; and “Learning how to learn” the adaptation and improvement of students' learning approaches (Figure 2 and Multimedia Appendix 2) [19].

In Figure 2, the taxonomy classifies learning outcomes into 6 interrelated domains, foundational knowledge, application, integration, human dimension, caring, and learning how to learn. This framework guided the categorization of students' written reflections to evaluate changes in cognitive, affective, and interpersonal learning domains beyond what is captured by Bloom Taxonomy.

Figure 2. Framework of the Fink Taxonomy of significant learning used for qualitative content analysis.



Data Analysis

Content Analysis

Content analysis was conducted using the Fink Taxonomy of significant learning to evaluate students' perceptions of community health care. This taxonomy categorizes cognitive process dimensions into various categories and subcategories [23,24]. KS independently read and coded all transcripts. Subsequently, the descriptors were discussed, identified, and agreed upon by KS, K Yamauchi, and NA. Similar codes were then grouped into categories and subcategories as they emerged from the data. These categories and subcategories were regularly reviewed and discussed by authors experienced in qualitative research to ensure the credibility of the findings [25]. Data saturation was assessed through an iterative process of constant comparison. After each round of coding, the research team reviewed whether new concepts or categories emerged. When no additional themes or codes were identified across 3 consecutive transcripts, we considered both thematic and code saturation to have been achieved [26,27]. The authors discussed and reached a consensus on all further changes, achieving code saturation. We adopted an inductive and deductive content analysis approach to balance theory-driven and data-driven coding. Initially, inductive coding allowed themes to emerge naturally from the students' reflections, ensuring that novel insights were not constrained by pre-existing frameworks. Subsequently, deductive categorization was applied using the Fink Taxonomy to systematically organize the emergent themes into recognized dimensions of significant learning. This combined approach enhanced the analytical rigor and minimized the risk of forcing data into preset categories [22].

Concepts for each cognitive process dimension in the Fink Taxonomy [20] were analyzed, and the number of units of analysis for each concept was counted. The researchers then grouped similar codes into themes and checked to see which dimension of the cognitive process they corresponded to.

AI-Assisted Analytic Support (Generative AI)

Additionally, to enhance the rigor and transparency of our qualitative analysis, we complemented the human coding process with a generative artificial intelligence (AI) tool (ChatGPT 5; OpenAI) [25,26]. The AI tool was prompted to identify potential thematic clusters and semantic relationships within anonymized textual data, following best practices for AI-assisted qualitative research as outlined in recent literature [28-31]. Specifically, the AI was used to (1) propose candidate thematic clusters and relationships across codes/excerpts, (2) suggest alternative categorizations and labels, and (3) identify potential negative cases or boundary instances that did not fit preliminary themes. Only deidentified materials (eg, anonymized excerpts and/or code lists) were entered into the tool; no personal identifiers were provided. The AI outputs were critically reviewed, refined, and confirmed through team discussion to prevent model-induced bias. AI outputs were treated as hypotheses for discussion rather than definitive findings, and the research team documented whether each suggestion was adopted, refined, or rejected after checking against the original

data. This triangulated approach improved coding consistency while ensuring that all final interpretations and coding decisions were made exclusively by human researchers [29,31,32]. Although formal interrater reliability statistics were not calculated, coding credibility was ensured through iterative peer review, team consensus discussions, and triangulation using AI-assisted content verification. Because we did not systematically archive full prompt-response transcripts during the earlier analytic iterations, we provide representative example prompts (and abbreviated outputs) in [Multimedia Appendix 3](#) to enhance transparency and reproducibility.

Reflexivity

We conducted a qualitative study using content analysis to determine the effectiveness of the newly developed program in 2022 on students' perceptions of community health care. In qualitative research, researchers' backgrounds can significantly influence data interpretation and the construction of findings and conclusions, necessitating careful reflexivity. Reflexivity was used to account for researchers' potential biases.

In this study, all authors were clinical educators with experience in COME, providing the necessary expertise to understand and interpret the data accurately. The authors were trained in educational theory (KS, K Yamauchi, and SI) and primary care (SA, Y Kamata, DS, MM, and TO), enabling them to approach the analysis from a well-rounded perspective. This dual expertise ensured a comprehensive understanding of the educational and clinical aspects of community-oriented medical education programs. This reflective approach aimed to ensure the credibility and reliability of our findings and provide valuable insights into the factors contributing to successful training programs.

Ethical Considerations

This study was approved by the Ethics Review Committee of the Graduate School of Medicine, Chiba University (Approval No. 3425, approved on October 4, 2019). All work involving research on human participants complied with the Declaration of Helsinki. The procedures for obtaining informed consent were explained to the medical students, who were also informed that this study would not affect their grades. The researchers explained to the participants and obtained their informed and voluntary consent. No financial or material compensation was provided to participants for their involvement in this study.

Results

Content analysis of the reports from first-year regional quota medical students revealed distinct patterns across the academic years 2021 and 2022, categorized according to the Fink Taxonomy of significant learning ([Tables 1](#) and [2](#)). The demographic characteristics of participants were similar between the 2021 and 2022 cohorts. The median age was 19 years in both groups (2021: 18-24 years; 2022: 18-21 years), and the distribution of sex assigned at birth did not differ significantly (2021: 75% male, 25% female; 2022: 70% male, 30% female; $P>.99$).

Table 1. Results of qualitative content analysis of first-year regional quota medical students' reflection reports in the 2021 community-oriented medical education (COME) program at Chiba University (N=20), crosswalked to the Fink Taxonomy of significant learning.

Theme, key synthesis, and crosswalk to the Fink domains	Subcategories (code counts)
Theme 1: community health care as an interconnected, resource-constrained system	
Community health care is understood as a local system shaped by needs, constraints, and intersector collaboration.	
Primary domains:	
• Foundational knowledge	• Specific initiatives (n=48), challenges/current state (n=38), definition/role (n=26), aging (n=16)
Secondary domains:	
• Integration	• Challenges/responses (n=13), improvement of knowledge and skills (n=11), communication/cooperation (n=9), collaboration with community (n=7)
Theme 2: patient-centered relationships and trust through communication and teamwork	
Communication is positioned as the practical foundation of patient-centered care and trust, supported by collaboration.	
Primary domains:	
• Application	• Communication and trust building (n=15)
Secondary domains:	
• Human dimension	• Patient-centered health care and communication (n=6)
• Caring	• Patient-centered health care (n=7), communication and collaboration (n=5)
Theme 3: emerging professional identity and responsibility toward community service	
Students begin to articulate who they should become as physicians and how they should contribute to the community.	
Primary domains:	
• Human dimension	• Patient-centered health care and communication (n=6), understanding and importance of community health care (n=5), general practice and broad knowledge (n=3), realization of the community health care field (n=2)
• Caring	• Contribution and involvement in community health care (n=9), personal growth and learning (n=8), patient-centered health care (n=7), communication and collaboration (n=5), personal motivation and awareness (n=1)
Secondary domains:	
• Application	• Improve knowledge/skills (n=20), practice/problem-solving (n=18)
Theme 4: developing a self-directed learning orientation for community practice	
Students emphasize experiential learning and continuous improvement as preparation for community practice.	
Primary domains:	
• Learning how to learn	• Practical experience and activity participation (n=2), improvement of skills and knowledge (n=2), improvement of communication skills (n=1), continuous learning and understanding of community (n=1)
Secondary domains:	
• Application	• Improvement of knowledge and skills (n=20), practice and problem-solving in community health care (n=18), communication and trust building (n=15), necessity and outlook of community health care (n=13), improvement of prevention and health awareness (n=9), use of information and communication technology (n=7), environmental preparation and collaboration in community health care (n=5), general practitioners and multidisciplinary knowledge (n=4)
• Foundational knowledge	• Specific initiatives in community health care (n=48), challenges and current state of community health care (n=38), definition and role of community health care (n=26), aging population and community health care (n=16)

Table 2. Results of qualitative content analysis of first-year regional quota medical students' reflection reports in the revised 2022 community-oriented medical education (COME) program at Chiba University (n=20), crosswalked to the Fink Taxonomy of significant learning.

Theme, key synthesis, and crosswalk to the Fink domains	Subcategories (code counts)
Theme 1: community health care as an interconnected, resource-constrained system	
Community health care is understood as a local system shaped by needs, constraints, and intersector collaboration.	
Primary domains:	
• Foundational knowledge	• Quality and supply of health care (n=51), shortage of physicians and medical resources (n=41), importance of communication and consultation (n=23), aging population and the community health care (n=13)
Secondary domain:	
• Integration	• Medical resources and collaboration (n=15), sustainability and technology in health care (n=12), patient-centered care method (n=10)
Theme 2: patient-centered relationships and trust through communication and teamwork	
Communication is positioned as the practical foundation of patient-centered care and trust, supported by collaboration.	
Primary domains:	
• Human dimension	• Patient-centered health care and communication (n=26)
• Application	• Communication and trust building (n=15)
Secondary domain:	
• Caring	• Communication and collaboration (n=15), patient-centered health care (n=11)
Theme 3: emerging professional identity and responsibility toward community service	
Students begin to articulate who they should become as physicians and how they should contribute to the community.	
Primary domains:	
• Human dimension	• Challenges and improvements in community health care (n=21), importance and current state of community health care (n=10)
• Caring	• Contribution and involvement in community health care (n=12), personal motivation and awareness (n=10), personal growth and learning (n=10)
Secondary domains:	
• Integration	• Ethics and communication (n=8)
Theme 4: developing a self-directed learning orientation for community practice	
Students emphasize experiential learning and continuous improvement as preparation for community practice.	
Primary domains:	
• Learning how to learn	• Acquisition of broad knowledge (n=5), activities to broaden knowledge (n=4), interest and training (n=3), mandatory community health care training (n=2), improvement of communication skills (n=1)
Secondary domains:	
• Application	• Improvement of medical knowledge and skills (n=20), practice and problem-solving in community health care (n=20), community healthcare policy and support (n=8), executing ability and motivation (n=8), importance of community health care (n=6), use of information and communication technology (n=5)

Across both cohorts, we identified four higher-order themes that synthesize students' reflections and crosswalk to multiple domains of the Fink Taxonomy: (1) community health care as an interconnected, resource-constrained system; (2) patient-centered relationships and trust through communication and teamwork; (3) emerging professional identity and responsibility toward community service; and (4) developing a self-directed learning orientation for community practice (Tables 1 and 2; representative quotes are provided in [Multimedia Appendices 4](#) and [5](#)). While foundational understanding of community health care was evident in both

years, the 2022 cohort more frequently expressed reflections aligned with affective and interpersonal learning (caring and human dimension) and reflective learning (learning how to learn), consistent with a more integrated pattern of "significant learning" in the Fink framework.

In [Table 1](#), reports were analyzed using the Fink Taxonomy of significant learning, which categorizes learning into six domains. Subcategories and code counts are presented to support the theme-based synthesis and the crosswalk to the Fink domains.

Representative quotes supporting each theme and subcategory are provided in [Multimedia Appendix 4](#).

In [Table 2](#), reports were analyzed using the Fink Taxonomy of significant learning, which categorizes learning into 6 domains. Subcategories and code counts are presented to support the theme-based synthesis and the crosswalk to the Fink domains. Representative quotes supporting each theme and subcategory are provided in [Multimedia Appendix 5](#).

For the academic year 2021, 311 codes were identified. In thematic terms, the 2021 reflections predominantly emphasized building a baseline understanding of community health care systems and challenges (theme 1; foundational knowledge and integration), accompanied by early attention to communication as a core clinical practice (theme 2; application), and emerging awareness of professional roles and values (theme 3; human dimension and caring). In the learning how to learn category, 6 codes were noted, including practical experience and participation in activities (n=2), improvement of skills and knowledge (n=2), improvement of communication skills (n=1), and continuous learning and understanding of the community (n=1). Students frequently reflected on the value of experiential learning, with one noting that “Practical experience enhances learning,” emphasizing how direct participation reinforced their understanding of classroom content. The caring category included 30 codes, highlighting areas such as contribution to and involvement in community health care (n=9), personal growth and learning (n=8), patient-centered health care (n=7), communication and collaboration (n=5), and personal motivation and awareness (n=1). The human dimension category had 16 codes, emphasizing patient-centered health care and communication (n=6), understanding and importance of community health care (n=5), general practice and broad knowledge (n=3), and realization of the community health care field (n=2). The integration category comprised 40 codes, addressing challenges and responses in community health care (n=13), improvement of knowledge and skills (n=11), communication and cooperation (n=9), and collaboration with the community (n=7). The application category contained 91 codes, focusing on improvement of knowledge and skills (n=20), practice and problem-solving in community health care (n=18), communication and trust-building (n=15), necessity and outlook of community health care (n=13), improvement of prevention and health awareness (n=9), use of information and communication technology (n=7), environmental preparation and collaboration in community health care (n=5), and general practitioners and multidisciplinary knowledge (n=4). Finally, the foundational knowledge category included 128 codes, covering specific initiatives in community health care (n=48), challenges and the current state of community health care (n=38), the definition and role of community health care (n=26), and the aging population and community health care (n=16). One student reflected, “Community healthcare is necessary for supporting aging populations,” illustrating growing awareness of the social challenges faced by Japan’s communities.

In comparison, the academic year 2022 had 385 codes. The learning how to learn category expanded to 15 codes, covering the acquisition of broad knowledge (n=5), activities to broaden knowledge (n=4), interest and training (n=3), mandatory

community health care training (n=2), and improvement of communication skills (n=1). The caring category rose significantly to 58 codes, with emphasis on communication and collaboration (n=15), contribution and involvement in community health care (n=12), patient-centered health care (n=11), personal growth and learning (n=10), and personal motivation and awareness (n=10). Students’ reflections often highlighted a stronger sense of teamwork and responsibility, exemplified by one participant who stated, “Collaboration among healthcare providers enhances communication and patient care.” Another remarked, “Personal motivation and awareness are key drivers for effective healthcare delivery,” underscoring the affective learning fostered by the revised program. The human dimension category also increased to 57 codes, focusing on patient-centered health care and communication (n=26), challenges and improvements in community health care (n=21), and the importance and current state of community health care (n=10). The integration category contained 45 codes, addressing medical resources and collaboration (n=15), sustainability and technology in health care (n=12), patient-centered care methods (n=10), and ethics and communication (n=8). The application category contained 82 codes, emphasizing improvement of medical knowledge and skills (n=20), practice and problem-solving in community health care (n=20), communication and trust-building (n=15), community health care policy and support (n=8), execution ability and motivation (n=8), importance of community health care (n=6), and use of information and communication technology (n=5). Finally, the foundational knowledge category remained consistent with 128 codes, focusing on the quality and supply of health care (n=51), shortage of physicians and medical resources (n=41), importance of communication and consultation (n=23), and aging population and community health care (n=13).

Comparing the data from 2 years, a notable increase was observed in the total number of codes, particularly in the caring and human dimension categories in 2022. Rather than relying solely on descriptive counts, the theme-based synthesis suggests a shift in emphasis from predominantly system- and knowledge-oriented descriptions (theme 1) toward reflections that more frequently articulated professional identity, values, and relationships in care (themes 2-3), alongside clearer intentions for continued learning (theme 4). These patterns are consistent with a more integrated profile across the Fink domains in 2022, while maintaining a stable baseline of foundational knowledge.

Beyond the individual code counts, several higher-level patterns emerged across both cohorts. First, the number of foundational knowledge codes remained stable across the 2 years, indicating a consistent baseline of understanding of community health care concepts. Second, the revised 2022 program demonstrated a marked increase in codes within the caring and human dimension categories. Third, learning how to learn codes also increased in 2022. Taken together, the thematic synthesis ([Tables 1](#) and [2](#)) indicates that the 2022 cohort’s reflections more often combined system understanding with interpersonal, value-oriented, and reflective learning elements, whereas the

2021 cohort's reflections were more concentrated in foundational and application-oriented descriptions.

Discussion

Summary of Key Findings

This study aimed to evaluate the educational impact of a COME program on first-year regional quota medical students at Chiba University. Through qualitative content analysis based on the Fink Taxonomy of significant learning, we identified substantial changes in students' perceptions and reflections following the program revision. The revised COME course, which incorporated additional early clinical exposure and faculty development for community hospital attendings, emphasized affective and interpersonal learning, as well as reflective learning—particularly within the caring, human dimension, and learning how to learn domains—while maintaining a stable baseline of foundational knowledge. These shifts were most apparent in themes 2-4, which foreground patient-centered relationships, emerging professional identity, and self-directed learning orientation. This aligns with the study's objective—to explore how early community-based experiences influence students' understanding of and commitment to community health care—and indicates that experiential learning opportunities in real clinical contexts contribute meaningfully to professional identity formation among first-year medical students.

Interpretation and Comparison With Previous Studies

Across both cohorts, several higher-level patterns emerged, offering broader insight into the educational impact of the revised COME program. While the number of foundational knowledge codes remained stable, indicating a consistent baseline understanding of community health care concepts, the revised 2022 program produced substantial increases in the caring, human dimension, and learning how to learn categories. Importantly, beyond these category-level changes, the theme-based synthesis (themes 1-4) indicates that students' reflections in 2022 more often integrated system understanding with interpersonal, value-oriented, and reflective elements of learning (themes 2-4). These findings suggest that the experiential and interpersonal elements of the program contributed to deeper affective, interpersonal, and reflective learning, extending beyond the acquisition of foundational knowledge.

Learning How to Learn and Lifelong Learning

The expansion of the learning how to learn category from 6 codes in 2021 to 15 in 2022 highlights the development of students' reflective and self-directed learning capacities.

This pattern aligns with theme 4 (developing a self-directed learning orientation for community practice), in which students increasingly recognized the importance of continuous learning and professional growth, as reflected in comments such as "Acquiring a broad range of knowledge is important for comprehensive healthcare" and "Ongoing interest and training are essential for professional development." This finding resonates with previous studies emphasizing the importance of reflective learning and self-regulation in developing lifelong

learning competencies [32-34]. The structured faculty mentorship and feedback embedded in the revised program may have contributed to this growth by encouraging students to articulate their learning needs and apply theoretical knowledge to real-world contexts.

Caring and Human Dimension: Growth in Empathy and Collaboration

A major finding was the substantial increase in the caring (from 30 to 58 codes) and human dimension (from 16 to 57 codes) categories. These domains were central to themes 2 and 3, which foregrounded patient-centered relationships, teamwork, and emerging professional identity and responsibility toward community service. Students' reflections increasingly addressed interpersonal communication, collaboration, and patient-centered care. One student stated, "Collaboration among healthcare providers enhances communication and patient care," while another reflected, "Patient-centered healthcare ensures that the care provided meets the needs of the patient." These reflections demonstrate an affective shift toward empathy and teamwork, suggesting that early exposure to real patient encounters and mentorship in community hospitals fostered students' understanding of the relational aspects of medicine. This aligns with prior research demonstrating that community-based and interprofessional learning environments enhance empathy and communication among medical students [35-37]. Similarly, our findings echo studies emphasizing that authentic, longitudinal clinical experiences—particularly those in underserved settings—help shape professional identity and strengthen social accountability [38-40]. One student further noted, "Addressing the challenges in community healthcare requires continuous improvement efforts," capturing a sense of responsibility and ongoing self-improvement that mirrors the iterative nature of clinical reasoning and patient care.

Integration and Application: Connecting Theory to Practice

The integration category increased modestly (40 to 45 codes), while the application category slightly decreased (91 to 82 codes). In the theme-based synthesis, integration was the most visible in theme 1 (community health care as a system) and theme 3 (professional identity and responsibility), where students linked collaboration, sustainability/technology, patient-centered care methods, and ethics. For instance, a student commented, "Collaboration between medical institutions is essential for optimizing resources," illustrating recognition of systemic and organizational aspects of health care delivery. However, the slight decline in application codes suggests that opportunities for practical problem-solving and leadership in community health planning could be expanded. Despite this, students continued to value applied knowledge, as shown in reflections such as, "Healthcare providers must be adept at problem-solving to address diverse community health issues." This pattern implies that while the program successfully deepened conceptual and interpersonal understanding, future iterations could further reinforce practical application and leadership in community contexts. These insights are consistent with previous reports that emphasize the need for a balance

between affective learning and skill-based experiential education [41-43].

Foundational Knowledge: Consolidating Core Understanding

The foundational knowledge category remained stable with 128 codes in both years, reflecting consistent comprehension of fundamental community health care principles. This stable foundation corresponded to theme 1 in both cohorts, where students continued to demonstrate awareness of the structural challenges facing Japan's health care system, particularly physician shortages and the aging population. Statements such as "Ensuring high-quality healthcare is critical for meeting community needs" and "Many rural areas face significant shortages of doctors and medical resources" highlight this awareness. This stability suggests that students entered the program with a basic understanding of community health needs and retained this foundation while expanding their capacity for empathy, reflection, and teamwork through experiential learning [44,45].

Comparison With Existing Educational Models

The differences observed between the COME program and early clinical exposure within university hospitals further highlight the program's unique contribution. Unlike hospital-based experiences that often emphasize specialized and controlled environments, the COME program situates students in authentic community settings where resource limitations and multidisciplinary collaboration are inherent. Students reflected that learning from community physicians provided unique insights into the realities of rural health care. As one student expressed,

Learning from physicians embedded in the community provided valuable insights into the challenges and rewards of rural healthcare.

This finding aligns with prior literature suggesting that early, community-facing experiences cultivate adaptability, resource awareness, and social accountability [46,47]. Furthermore, recent evidence indicates that such community-facing clinical exposure not only enhances engagement and learning outcomes but also fosters the development of professional identity, social accountability, and collaborative competencies essential for clinical reasoning [46-48]. In addition, context-specific curriculum design is crucial for optimizing the educational impact of such programs. Understanding regional variations in students' study habits, learning preferences, and access to educational resources allows educators to adapt teaching methods—such as incorporating imaging-based learning modules, interprofessional collaboration, or problem-based approaches—to enhance engagement and relevance in local settings [49]. These insights highlight that while COME provides authentic, real-world exposure to health care challenges, tailoring curricular content and pedagogy to local educational contexts can further enhance learning outcomes and promote sustained motivation among medical students. By engaging with real patients and health care teams early in their education, students not only improved their communication skills but also

developed a sense of mission toward addressing regional health disparities.

Limitations

This study has some limitations. First, the sample size was limited to first-year regional quota medical students at a single institution (n=40), which restricts the generalizability of the findings. The participants were from a single admission track; hence, demographic and baseline characteristics between the 2021 and 2022 cohorts were not formally analyzed, which may introduce selection bias. However, both cohorts had similar demographic profiles, showing comparable age ranges and distributions of sex assigned at birth. In addition, the results were derived from students' written reflections rather than objective performance-based assessments. Therefore, while these reflections provide valuable insights into students' perceptions and learning processes, they do not directly capture behavioral or performance-level outcomes. The study also relied on self-reported data, which may be subject to bias. The qualitative nature of the content analysis may also limit the reproducibility of the results, as it involves subjective data interpretation. Furthermore, the study used a cross-sectional design comparing 2 adjacent cohorts without a control group; thus, the observed improvements might reflect cohort effects, prior IPE exposure, or faculty-related factors rather than the revised COME program itself. The analysis also relied primarily on one coder, without interrater reliability testing, audit trails, or member checking, which could affect analytical rigor. In addition, applying the Fink Taxonomy a priori may have constrained data interpretation by fitting responses into predefined categories. However, previous qualitative studies have suggested that increases in code frequency and diversity—when analyzed within a validated theoretical framework—can indicate broader and deeper levels of learning and reflection [23,24]. Recent evidence also supports this interpretation, showing that analyses grounded in theoretically informed frameworks such as the Fink Taxonomy can meaningfully capture students' cognitive and affective development through code expansion and thematic diversification [50]. Therefore, in this study, the observed growth in the number and variety of codes (eg, "Caring," "Human dimension") was interpreted as a qualitative indicator of students' multidimensional learning development rather than a purely quantitative change. Nonetheless, no objective or longitudinal outcomes (eg, rural practice retention or sustained career engagement) were assessed. Another limitation is the short duration of the study, which spanned only 2 academic years; longer-term studies are needed to evaluate the sustained impact of the program on students' career choices and retention in rural health care settings. Finally, the study did not account for external factors that may have influenced students' perceptions, such as their personal experiences, socioeconomic backgrounds, or individual learning preferences. Tailoring educational strategies to align with students' learning styles and real-world experiential opportunities may foster more effective and personalized medical education, ultimately enhancing both theoretical understanding and practical competency [51]. Future research should incorporate such individualized approaches and

explore how these factors interact with community-based learning outcomes.

Conclusions

Enhancements to the COME program at Chiba University, including early clinical exposure and the involvement of faculty-development-trained community hospital attendings, significantly improved students' understanding of patient-centered and community-oriented health care. Our theme-based synthesis suggests that the revised program was associated with greater emphasis on patient-centered relationships, emerging professional identity, and self-directed learning (themes 2-4), while maintaining a stable baseline of

foundational understanding (theme 1). These experiences cultivated empathy, reflective learning, and communication skills, which are essential for generalist physicians serving in underserved areas. Beyond these educational benefits, the findings have broader implications for medical education policy and workforce distribution. Structured community-based programs such as COME may contribute to addressing Japan's rural physician shortage by fostering students' identification with community health care roles early in their training. Continual evaluation and iterative refinement of COME programs are necessary to ensure that they remain responsive to regional health care needs and effectively cultivate physicians who are competent, compassionate, and socially accountable.

Acknowledgments

Generative AI tools were used only to support the qualitative data analysis in this study, specifically to assist in identifying potential patterns and alternative categorizations within the dataset. All final coding decisions and interpretations were made exclusively by the human researchers. No generative AI tools were used in the preparation, writing, or editing of the manuscript text.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Data Availability

The dataset generated or analyzed during this study is available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: KS, K Yamauchi, Y Kamata, SI

Methodology: KS, K Yamauchi, Y Kamata, SI

Investigation: KS, K Yamauchi, Y Kamata, SI

Data Curation: KS, K Yamauchi, NA, Y Kamata, IS

Formal Analysis: KS, K Yamauchi, NA, SI

Qualitative Analysis: KS, K Yamauchi, NA, SI

Writing – Original Draft: KS

Writing – Review & Editing: K Yamauchi, NA, Y Kamata, SA, DS, MM, Y Katsuyama, TO, K Yoshimura, SI

Supervision: SI

Project Administration: KS, K Yamauchi, SI

Resources: KS, K Yamauchi, Y Kamata, DS, MM, YY, TO, K Yoshimura, SI.

Visualization: KS

Conflicts of Interest

None declared.

Multimedia Appendix 1

Students' reports.

[[DOCX File, 27 KB - mededu_v12i1e84406_app1.docx](#)]

Multimedia Appendix 2

The Fink Taxonomy of significant learning.

[[DOCX File, 27 KB - mededu_v12i1e84406_app2.docx](#)]

Multimedia Appendix 3

Representative prompts and outputs for artificial intelligence–assisted qualitative content analysis.

[[DOCX File, 32 KB - mededu_v12i1e84406_app3.docx](#)]

Multimedia Appendix 4

Results of qualitative content analysis of first-year regional quota medical students' reflection reports in the 2021 community-oriented medical education program at Chiba University (n=20).

[[DOCX File , 37 KB - mededu_v12i1e84406_app4.docx](#)]

Multimedia Appendix 5

Results of qualitative content analysis of first-year regional quota medical students' reflection reports in the revised 2022 community-oriented medical education program at Chiba University (N=20).

[[DOCX File , 33 KB - mededu_v12i1e84406_app5.docx](#)]

References

1. 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;9(4):e025979 [[FREE Full text](#)] [doi: [10.1136/bmjjopen-2018-025979](https://doi.org/10.1136/bmjjopen-2018-025979)] [Medline: [31133622](#)]
2. Morris S, Sutton M, Gravelle H. Inequity and inequality in the use of health care in England: an empirical investigation. *Soc Sci Med* 2005;60(6):1251-1266. [doi: [10.1016/j.socscimed.2004.07.016](https://doi.org/10.1016/j.socscimed.2004.07.016)] [Medline: [15626522](#)]
3. Guttmann A, Shipman S, Lam K, Goodman D, Stukel T. Primary care physician supply and children's health care use, access, and outcomes: findings from Canada. *Pediatrics* 2010;125(6):1119-1126. [doi: [10.1542/peds.2009-2821](https://doi.org/10.1542/peds.2009-2821)] [Medline: [20498170](#)]
4. 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;50(2):185-202. [doi: [10.1016/s0277-9536\(99\)00183-5](https://doi.org/10.1016/s0277-9536(99)00183-5)] [Medline: [10619689](#)]
5. Winkelmann J, Muench U, Maier CB. Time trends in the regional distribution of physicians, nurses and midwives in Europe. *BMC Health Serv Res* 2020;20(1):937 [[FREE Full text](#)] [doi: [10.1186/s12913-020-05760-y](https://doi.org/10.1186/s12913-020-05760-y)] [Medline: [33046077](#)]
6. Wu J. Measuring inequalities in the demographical and geographical distribution of physicians in China: generalist versus specialist. *Int J Health Plann Manage* 2018;33(4):860-879. [doi: [10.1002/hpm.2539](https://doi.org/10.1002/hpm.2539)] [Medline: [29781216](#)]
7. Ishikawa T, Fujiwara K, Ohba H, Suzuki T, Ogasawara K. Forecasting the regional distribution and sufficiency of physicians in Japan with a coupled system dynamics-geographic information system model. *Hum Resour Health* 2017;15(1):64 [[FREE Full text](#)] [doi: [10.1186/s12960-017-0238-8](https://doi.org/10.1186/s12960-017-0238-8)] [Medline: [28899406](#)]
8. Koike S, Okazaki K, Tokinobu A, Matsumoto M, Kotani K, Kataoka H. Factors associated with regional retention of physicians: a cross-sectional online survey of medical students and graduates in Japan. *Hum Resour Health* 2023;21(1):85 [[FREE Full text](#)] [doi: [10.1186/s12960-023-00871-z](https://doi.org/10.1186/s12960-023-00871-z)] [Medline: [37885012](#)]
9. Kobayashi Y, Takaki H. Geographic distribution of physicians in Japan. *Lancet* 1992;340(8832):1391-1393. [doi: [10.1016/0140-6736\(92\)92569-2](https://doi.org/10.1016/0140-6736(92)92569-2)] [Medline: [1360099](#)]
10. Toyabe S. Trend in geographic distribution of physicians in Japan. *Int J Equity Health* 2009;8(1):5 [[FREE Full text](#)] [doi: [10.1186/1475-9276-8-5](https://doi.org/10.1186/1475-9276-8-5)] [Medline: [19257879](#)]
11. Tanihara S, Kobayashi Y, Une H, Kawachi I. Urbanization and physician maldistribution: a longitudinal study in Japan. *BMC Health Serv Res* 2011;11:260 [[FREE Full text](#)] [doi: [10.1186/1472-6963-11-260](https://doi.org/10.1186/1472-6963-11-260)] [Medline: [21982582](#)]
12. Yahata S, Takeshima T, Kenzaka T, Okayama M. Long-term impact of undergraduate community-based clinical training on community healthcare practice in Japan: a cross-sectional study. *BMC Med Educ* 2020;20(1):337 [[FREE Full text](#)] [doi: [10.1186/s12909-020-02258-3](https://doi.org/10.1186/s12909-020-02258-3)] [Medline: [33040409](#)]
13. Fukui S, Shikino K, Nishizaki Y, Shimizu T, Yamamoto Y, Kobayashi H, et al. Association between regional quota program in medical schools and practical clinical competency based on general medicine in-training examination score: a nationwide cross-sectional study of resident physicians in Japan. *Postgrad Med J* 2023;99(1177):1197-1204. [doi: [10.1093/postmj/qgad059](https://doi.org/10.1093/postmj/qgad059)] [Medline: [37474744](#)]
14. Shikino K, Tago M, Hirata R, Watari T, Sasaki Y, Takahashi H, et al. Educational strategies for general medicine education in accordance with the model core curriculum for medical education in Japan. *J Gen Fam Med* 2023;24(4):272-273 [[FREE Full text](#)] [doi: [10.1002/jgf2.619](https://doi.org/10.1002/jgf2.619)] [Medline: [37484122](#)]
15. Medical Education Model Core Curriculum Expert Research Committee. The model core curriculum for medical education in Japan 2022. Ministry of Education, Culture, Sports, Science and Technology. 2023. URL: https://www.mext.go.jp/content/20250411-mxt_igaku-000028108_00003-2.pdf [accessed 2023-11-23]
16. Ohta R, Ryu Y, Sano C. The contribution of citizens to community-based medical education in Japan: a systematic review. *Int J Environ Res Public Health* 2021;18(4):1575 [[FREE Full text](#)] [doi: [10.3390/ijerph18041575](https://doi.org/10.3390/ijerph18041575)] [Medline: [33562329](#)]
17. 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](https://doi.org/10.1186/s42466-020-00059-z)] [Medline: [33324920](#)]
18. Sandhu VK, Jose DM, Feldman CH. Underserved communities: enhancing care with graduate medical education. *Rheum Dis Clin North Am* 2020;46(1):167-178 [[FREE Full text](#)] [doi: [10.1016/j.rdc.2019.09.009](https://doi.org/10.1016/j.rdc.2019.09.009)] [Medline: [31757283](#)]

19. Taxonomy of Significant Learning. The Peak Performance Center. URL: <https://thepeakperformancecenter.com/educational-learning/thinking/blooms-taxonomy/taxonomy-of-significant-learning/> [accessed 2025-09-13]
20. Anderson LW, David R. A Taxonomy for Learning, Teaching and Assessing: a Revision of Bloom's Taxonomy. New York, NY: Longman; 2001.
21. Bloom BS. Taxonomy of educational objectives, the classification of educational goals. In: Handbook I: Cognitive Domain. New York, NY: Longmans, Green, and Co; 1956.
22. Morton DA, Colbert-Getz JM. Measuring the impact of the flipped anatomy classroom: the importance of categorizing an assessment by bloom's taxonomy. *Anat Sci Educ* 2017;10(2):170-175. [doi: [10.1002/ase.1635](https://doi.org/10.1002/ase.1635)] [Medline: [27427860](https://pubmed.ncbi.nlm.nih.gov/27427860/)]
23. 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](https://doi.org/10.1111/j.1365-2648.2007.04569.x)] [Medline: [18352969](https://pubmed.ncbi.nlm.nih.gov/18352969/)]
24. Graneheim UH, Lindgren B, Lundman B. Methodological challenges in qualitative content analysis: a discussion paper. *Nurse Educ Today* 2017;56:29-34. [doi: [10.1016/j.nedt.2017.06.002](https://doi.org/10.1016/j.nedt.2017.06.002)] [Medline: [28651100](https://pubmed.ncbi.nlm.nih.gov/28651100/)]
25. Granero Lucchetti AL, Ezequiel ODS, Oliveira IND, Moreira-Almeida A, Lucchetti G. Using traditional or flipped classrooms to teach "Geriatrics and Gerontology"? Investigating the impact of active learning on medical students' competences. *Med Teach* 2018;40(12):1248-1256. [doi: [10.1080/0142159X.2018.1426837](https://doi.org/10.1080/0142159X.2018.1426837)] [Medline: [29355063](https://pubmed.ncbi.nlm.nih.gov/29355063/)]
26. Guest G, Namey E, Chen M. A simple method to assess and report thematic saturation in qualitative research. *PLoS One* 2020;15(5):e0232076 [FREE Full text] [doi: [10.1371/journal.pone.0232076](https://doi.org/10.1371/journal.pone.0232076)] [Medline: [32369511](https://pubmed.ncbi.nlm.nih.gov/32369511/)]
27. Kadam PD, Chuan HH. Erratum to: rectocutaneous fistula with transmigration of the suture: a rare delayed complication of vault fixation with the sacrospinous ligament. *Int Urogynecol J* 2016;27(3):505 [FREE Full text] [doi: [10.1007/s00192-016-2952-5](https://doi.org/10.1007/s00192-016-2952-5)] [Medline: [26811110](https://pubmed.ncbi.nlm.nih.gov/26811110/)]
28. Rossi M, Rehman S. Integrating artificial intelligence into telemedicine: evidence, challenges, and future directions. *Cureus* 2025;17(8):e90829. [doi: [10.7759/cureus.90829](https://doi.org/10.7759/cureus.90829)] [Medline: [40988810](https://pubmed.ncbi.nlm.nih.gov/40988810/)]
29. Bennis I, Mouwafaq S. Advancing AI-driven thematic analysis in qualitative research: a comparative study of nine generative models on Cutaneous Leishmaniasis data. *BMC Med Inform Decis Mak* 2025;25(1):124 [FREE Full text] [doi: [10.1186/s12911-025-02961-5](https://doi.org/10.1186/s12911-025-02961-5)] [Medline: [40065373](https://pubmed.ncbi.nlm.nih.gov/40065373/)]
30. Yue Y, Liu D, Lv Y, Hao J, Cui P. A practical guide and assessment on using ChatGPT to conduct grounded theory: tutorial. *J Med Internet Res* 2025;27:e70122 [FREE Full text] [doi: [10.2196/70122](https://doi.org/10.2196/70122)] [Medline: [40367510](https://pubmed.ncbi.nlm.nih.gov/40367510/)]
31. Bijk R, Merkouris SS, Dowling NA, Rodda SN. ChatGPT for automated qualitative research: content analysis. *J Med Internet Res* 2024;26:e59050 [FREE Full text] [doi: [10.2196/59050](https://doi.org/10.2196/59050)] [Medline: [39052327](https://pubmed.ncbi.nlm.nih.gov/39052327/)]
32. Marceau M, Vachon Lachiver ?, Lambert D, Daoust J, Dion V, Langlois M, et al. Assessment practices in continuing professional development activities in health professions: a scoping review. *J Contin Educ Health Prof* 2024;44(2):81-89. [doi: [10.1097/CEH.0000000000000507](https://doi.org/10.1097/CEH.0000000000000507)] [Medline: [37490015](https://pubmed.ncbi.nlm.nih.gov/37490015/)]
33. Main PAE, Anderson S. Evidence for continuing professional development standards for regulated health practitioners in Australia: a systematic review. *Hum Resour Health* 2023;21(1):23 [FREE Full text] [doi: [10.1186/s12960-023-00803-x](https://doi.org/10.1186/s12960-023-00803-x)] [Medline: [36941655](https://pubmed.ncbi.nlm.nih.gov/36941655/)]
34. Institute of Medicine (US) Committee on Planning a Continuing Health Professional Education Institute. Redesigning Continuing Education in the Healthcare Profession. Washington (DC): National Academies Press (US); 2010.
35. Steinmair D, Zervos K, Wong G, Löffler-Stastka H. Importance of communication in medical practice and medical education: an emphasis on empathy and attitudes and their possible influences. *World J Psychiatry* 2022;12(2):323-337 [FREE Full text] [doi: [10.5498/wjp.v12.i2.323](https://doi.org/10.5498/wjp.v12.i2.323)] [Medline: [35317334](https://pubmed.ncbi.nlm.nih.gov/35317334/)]
36. Kerasidou A, Bærøe K, Berger Z, Caruso Brown AE. The need for empathetic healthcare systems. *J Med Ethics* 2020;47(12):e27 [FREE Full text] [doi: [10.1136/medethics-2019-105921](https://doi.org/10.1136/medethics-2019-105921)] [Medline: [32709754](https://pubmed.ncbi.nlm.nih.gov/32709754/)]
37. Moudatsou M, Stavropoulou A, Philalithis A, Koukouli S. The role of empathy in health and social care professionals. *Healthcare (Basel)* 2020;8(1):26 [FREE Full text] [doi: [10.3390/healthcare8010026](https://doi.org/10.3390/healthcare8010026)] [Medline: [32019104](https://pubmed.ncbi.nlm.nih.gov/32019104/)]
38. Kwame A, Petrucci PM. A literature-based study of patient-centered care and communication in nurse-patient interactions: barriers, facilitators, and the way forward. *BMC Nurs* 2021;20(1):158 [FREE Full text] [doi: [10.1186/s12912-021-00684-2](https://doi.org/10.1186/s12912-021-00684-2)] [Medline: [34479560](https://pubmed.ncbi.nlm.nih.gov/34479560/)]
39. Gilligan C, Powell M, Lynagh M, Ward BM, Lonsdale C, Harvey P, et al. Interventions for improving medical students' interpersonal communication in medical consultations. *Cochrane Database Syst Rev* 2021;2(2):CD012418 [FREE Full text] [doi: [10.1002/14651858.CD012418.pub2](https://doi.org/10.1002/14651858.CD012418.pub2)] [Medline: [33559127](https://pubmed.ncbi.nlm.nih.gov/33559127/)]
40. King A, Hoppe RB. "Best practice" for patient-centered communication: a narrative review. *J Grad Med Educ* 2013;5(3):385-393. [doi: [10.4300/JGME-D-13-00072.1](https://doi.org/10.4300/JGME-D-13-00072.1)] [Medline: [24404300](https://pubmed.ncbi.nlm.nih.gov/24404300/)]
41. Munthe C, Fumagalli D, Malmqvist E. Sustainability principle for the ethics of healthcare resource allocation. *J Med Ethics* 2021;47(2):90-97 [FREE Full text] [doi: [10.1136/medethics-2020-106644](https://doi.org/10.1136/medethics-2020-106644)] [Medline: [33154090](https://pubmed.ncbi.nlm.nih.gov/33154090/)]
42. Lewis S, Damarell R, Tieman J, Trenerry C. Finding the integrated care evidence base in pubmed and beyond: a bibliometric study of the challenges. *Int J Integr Care* 2018;18(3):11 [FREE Full text] [doi: [10.5334/ijic.3975](https://doi.org/10.5334/ijic.3975)] [Medline: [30220894](https://pubmed.ncbi.nlm.nih.gov/30220894/)]
43. Institute of Medicine (US) Committee on Using Performance Monitoring to Improve Community Health. Improving Health in the Community: A Role for Performance Monitoring. Washington, DC: National Academies Press; 1997.

44. Mitchell P, Cribb A, Entwistle V. Vagueness and variety in person-centred care. *Wellcome Open Res* 2022;7:170 [FREE Full text] [doi: [10.12688/wellcomeopenres.17970.1](https://doi.org/10.12688/wellcomeopenres.17970.1)] [Medline: [35865218](#)]
45. Haruta J, Ando T, Fujishima S. How do medical students learn in an online community diagnostics program? *BMC Med Educ* 2023;23(1):15 [FREE Full text] [doi: [10.1186/s12909-023-04007-8](https://doi.org/10.1186/s12909-023-04007-8)] [Medline: [36627667](#)]
46. Ewnte B, Yigzaw T. Early clinical exposure in medical education: the experience from debre tabor university. *BMC Med Educ* 2023;23(1):252 [FREE Full text] [doi: [10.1186/s12909-023-04221-4](https://doi.org/10.1186/s12909-023-04221-4)] [Medline: [37069522](#)]
47. Ghanbari-Afra L, Adib-Hajbaghery M, Dianati M. Human caring: a concept analysis. *J Caring Sci* 2022;11(4):246-254 [FREE Full text] [doi: [10.34172/jcs.2022.21](https://doi.org/10.34172/jcs.2022.21)] [Medline: [36483689](#)]
48. Ogut E, Yildirim FB, Senol Y, Senol AU. Comprehensive evaluation of the educational impact and effectiveness of specialized study modules in cross-sectional anatomy: a study on student engagement and learning outcomes. *BMC Med Educ* 2025;25(1):514 [FREE Full text] [doi: [10.1186/s12909-025-07050-9](https://doi.org/10.1186/s12909-025-07050-9)] [Medline: [40211255](#)]
49. Barut C, Ogut E, Karaer E, Yavuz M. Anatomy study preferences of medical students in relation to gender, academic year and geographical distribution: considerations for anatomy education approaches. *Bratisl Med J* 2025 Mar 13;126(4):482-498. [doi: [10.1007/s44411-025-00078-9](https://doi.org/10.1007/s44411-025-00078-9)]
50. Ishizuka K, Shikino K, Takada N, Sakai Y, Ototake Y, Kobayashi T, et al. Enhancing clinical reasoning skills in medical students through team-based learning: a mixed-methods study. *BMC Med Educ* 2025;25(1):221 [FREE Full text] [doi: [10.1186/s12909-025-06784-w](https://doi.org/10.1186/s12909-025-06784-w)] [Medline: [39934738](#)]
51. Ogut E, Senol Y, Yildirim F. Do learning styles affect study duration and academic success? *Eur J Anat* 2017;21(3):235-240.

Abbreviations

AI: artificial intelligence

CBME: community-based medical education

COME: community-oriented medical education

IPE: interprofessional education

Edited by A Stone, T Leung; submitted 19.Sep.2025; peer-reviewed by RS Gomaa Mahmoud, E Ogut, Aminuddin; comments to author 21.Oct.2025; revised version received 23.Dec.2025; accepted 30.Dec.2025; published 19.Jan.2026.

Please cite as:

Shikino K, Yamauchi K, Araki N, Ozaki N, Kamata Y, Aoki S, Katsuyama Y, Sogai D, Miyamoto M, Yoshimura K, Oki T, Ito S. Impact of Community-Oriented Medical Education on Medical Students' Perceptions of Community Health Care: Qualitative Study. JMIR Med Educ 2026;12:e84406

URL: <https://mededu.jmir.org/2026/1/e84406>

doi: [10.2196/84406](https://doi.org/10.2196/84406)

PMID:

©Kiyoshi Shikino, Kazuyo Yamauchi, Nobuyuki Araki, Naoto Ozaki, Yu Kamata, Shinya Aoki, Yota Katsuyama, Daichi Sogai, Mai Miyamoto, Kensuke Yoshimura, Takeshi Oki, Shoichi Ito. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 19.Jan.2026. 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.

Original Paper

Comparative Efficacy of Simulation-Based and Traditional Training in Ultrasound-Assisted Regional Anesthesia for Medical Students: Randomized Controlled Trial

David Sánchez-Poveda^{1*}, MD; Juan Vicente-Mampel^{2*}, PT, PhD; Belén Curto^{3*}, PhD; Vidal Moreno^{3*}, PhD; Juan A García-Estebar^{3*}, PhD; Felipe Hernández-Zaballos^{1*}, MD, PhD; Pablo Alonso-Hernández^{1*}, PhD

¹Pain Unit, Anaesthesiology Service, Complejo Asistencial Universitario de Salamanca (CAUSA), Salamanca, Spain

²Department of Physiotherapy, School of Medicine and Health Science, Valencia Catholic University Saint Vincent Martyr, Valencia, Spain

³Department of Computer Science and Automation, University of Salamanca, University of Salamanca, Salamanca, Spain

* all authors contributed equally

Corresponding Author:

Juan Vicente-Mampel, PT, PhD

Department of Physiotherapy

School of Medicine and Health Science

Valencia Catholic University Saint Vincent Martyr

Calle Quevedo, número 2

Valencia, 46001

Spain

Phone: 1 963637412

Email: juan.vicente@ucv.es

Abstract

Background: Ultrasound is very important in medicine and teaching, but there are not many formal training programs. We also do not know much about what students think. To be good at using ultrasound, one needs to learn technical, thinking, and seeing skills. This is especially true in regional anesthesia (RA), where mistakes in reading images can cause problems. Training with simulations is a safe and good way to learn these skills. Some models are helpful for teaching how to perform procedures using ultrasound.

Objective: This study aimed to evaluate the effectiveness, localization time, and success rate of traditional teaching versus a new simulation-based teaching method for RA designed by the investigators among undergraduate medical students.

Methods: A prospective, randomized controlled trial was conducted at the University of Salamanca from April 2022 to January 2023. A total of 34 medical students in their fourth to sixth academic years were randomly allocated to either a simulation-based training group using the Haptic Ultrasound Probe or a traditional teaching group. The simulation approach used a realistic probe replica and a software-based ultrasound environment, whereas the traditional method comprised a theoretical lecture and curated audiovisual materials. Two days after training, participants underwent a blinded assessment requiring the identification of peripheral nerve plexuses using an ultrasound device. The primary outcome measured was the successful identification of nerves, and the secondary outcome was the time taken to complete each procedure. Data were analyzed using an intention-to-treat approach.

Results: A total of 34 medical students (fourth to sixth years) were recruited to compare traditional teaching with simulation-based training in ultrasound-guided nerve localization. No statistically significant differences were found in the success rates between the groups. For the interscalene approach, the traditional teaching group achieved a 100% (17/17) success rate compared to 82% (14/17) in the simulation group ($P=.07$). The time to task completion was similar across most procedures. In the sciatic nerve division, the traditional teaching group was significantly faster, with a mean time of 42.4 (SD 39.5) seconds ($P=.02$). The regression models showed no significant interaction between the intervention type and academic year. Both teaching methods had positive educational impacts.

Conclusions: Simulation-based learning effectively supports competency acquisition in RA and offers a safe, scalable alternative to traditional methods. Its integration into medical curricula may standardize training, improve skill consistency, and enhance patient safety. Further multicenter studies with larger, diverse cohorts are needed to validate these benefits and guide implementation in medical education.

KEYWORDS

education; ultrasound; anesthesia; nerve block; simulation

Introduction

Medical specialists are increasingly using ultrasound technology for diagnostic purposes and to guide therapeutic procedures owing to its numerous clinical advantages [1]. Some researchers have characterized portable ultrasound as the “visual stethoscope” of the 21st century [2], highlighting its growing relevance in modern medicine. Point-of-care ultrasound is a portable and versatile imaging modality that allows clinicians to perform rapid bedside assessments in various clinical scenarios. It is accessible and efficient [3,4]. Point-of-care ultrasound has become an essential tool in emergency medicine, critical care, and other specialties [5,6]. Consequently, ultrasound training has been incorporated into the undergraduate medical curriculum, encompassing both theoretical instruction and practical hands-on experience. However, the American Institute of Ultrasound in Medicine reports that only approximately one-third of medical schools in the United States have implemented specialized ultrasound training programs. In addition, there is a noticeable gap in the literature regarding medical students’ views on optimal practices for ultrasound education [7]. Furthermore, there is limited understanding of whether its use enhances the acquisition of medical knowledge independent of technical skill development [8].

To effectively acquire ultrasound competencies, medical students must master a combination of technical, cognitive, and perceptual skills. Key areas include proper probe handling, such as the ability to maneuver the probe across anatomical surfaces, and the development of a 3D orientation to accurately visualize and interpret internal structures. Additionally, students must integrate real-time imaging with their anatomical knowledge to make precise adjustments during the examination, which is critical for both diagnostic accuracy and therapeutic efficacy [9]. Equally important is visual training, which enables students to recognize the anatomical structures within ultrasound images. This process requires repeated exposure and active problem-solving to enhance spatial awareness and the ability to discern subtle differences between tissues and organs [10]. These visual and interpretive skills must be grounded in a strong understanding of anatomy to ensure accurate interpretations.

Motor coordination also plays a pivotal role, particularly in interventional procedures that involve inserting needles into targeted soft tissues, where precision is essential to avoid complications. In the field of regional anesthesia (RA), the use of ultrasound has increased significantly over the past few decades because of its ability to enhance the safety and accuracy of procedures that were traditionally guided by anatomical landmarks and neurostimulation [11]. However, to fully realize these safety benefits, trainees must receive comprehensive education on the use of ultrasound technology as the equipment alone cannot guarantee a safe outcome. The number and complexity of RA techniques have also expanded, with some procedures requiring advanced sonoanatomical knowledge for

successful performance [9]. In response to this evolution, Regional Anaesthesia UK (the UK division of the European Society of Regional Anaesthesia and Pain Therapy) proposed a classification system that groups RA techniques according to the level of knowledge and expertise required for their performance. Notably, most malpractice incidents associated with RA result from misinterpretation of ultrasound images or inadequate anatomical exploration [12].

Simulation-based learning provides valuable opportunities to develop these competencies in controlled settings. It allows both undergraduate and postgraduate students to practice without the ethical and logistical challenges of performing procedures on live patients or cadavers while still preserving the tactile feedback necessary for motor skill development [13]. Simulation training also enables repetition, feedback, and safe failure, which are key components of effective procedural training. As such, it has emerged as a highly promising approach for training in ultrasound-guided procedures. Several simulation-based teaching models have demonstrated their effectiveness in developing RA skills [14,15]. In this context, our research group developed a novel ultrasound simulator, the Haptic Ultrasound Probe (HUSP), which was officially registered on April 23, 2018, with the Spanish Trademark and Patent Office (*Oficina Española de Marcas y Patentes*) [16]. This study aimed to evaluate the effectiveness of a simulation-based teaching method using HUSP compared to traditional instruction for teaching RA to undergraduate medical students [1].

Methods

Study Design

This study was designed as a prospective, longitudinal, randomized controlled trial and conducted in accordance with the CONSORT (Consolidated Standards of Reporting Trials) guidelines (Multimedia Appendix 1) [17]. Participants were randomly and equally assigned to either a simulation-based education group or a traditional educational method group. Data were collected between April 2022 and January 2023.

Ethical Considerations

Participation was voluntary, and all students provided informed consent before their inclusion in this study. The study adhered to the ethical principles of the Declaration of Helsinki [18] and was approved by the institutional review board of the Department of Surgery, Faculty of Medicine and Health Sciences, University of Salamanca (protocol code 001-2018; July 13, 2018). Measures were implemented to protect participant privacy and confidentiality, including the anonymization of all collected data and secure storage of records. No compensation or incentives were provided to students.

Sample Size Calculation

This study used convenience sampling by recruiting participants who were easily accessible in a university. While this approach allows for efficient data gathering, it is crucial to recognize that it may limit the generalizability of the study due to potential selection biases. Therefore, an intention-to-treat analysis was performed to address this concern.

Participants

A total of 34 students from the Faculty of Medicine at the University of Salamanca (medical degree program) participated in this study. A sample was selected including those from the fourth, fifth, and sixth years of study. Both data collectors and analysts were unaware of the assigned interventions. The sample consisted of 32% (n=11) male and 68% (n=23) female individuals distributed in 2 independent groups. The inclusion criteria were as follows: (1) students enrolled in the medical degree program at the University of Salamanca; (2) students in the fourth, fifth, or sixth year; (3) students who voluntarily provided informed consent; (4) students willing to participate in all activities and assessments of the study; and (5) students with sufficient cognitive and physical capabilities to participate in the teaching interventions (simulation based).

Randomization and Blinding

All educational sessions were conducted by 2 instructors with extensive experience (>10 years) in ultrasound-guided RA following standardized teaching protocols. One of the researchers delivered both the traditional and simulation-based interventions, whereas the second researcher acted as an independent evaluator and remained blinded to group allocation. To minimize bias, group assignment was randomized using a table generated by an independent researcher in Microsoft Excel. A double-blind methodology was implemented; neither the participants nor the evaluator had knowledge of the group allocation or performance outcomes until the end of the study.

Procedures

This study investigated the simulation of interscalene, supraclavicular, and sciatic nerve blocks, including the external and internal popliteal divisions of the nerves, as they appear as Plan A and Plan B blocks in the Regional Anaesthesia UK guidelines. An Esaote MyLab Alpha ultrasound device, identical to that used to generate simulation ecosystem images, was used in the experiment. Two individuals with plexuses that were confirmed to have adequate ultrasound visibility before testing took part in the study. All participants were scheduled for an assessment 2 days after their simulation training with the HUSP

simulator. The assessment required participants to identify plexuses in 3 live individuals. To ensure impartiality in student group allocations, a blind investigator who was not involved in providing the theoretical lesson or guiding the simulation ecosystem practice administered the tests. The time taken by each student to locate the plexus was recorded, and successful identification was noted (maximum identification time: 120 seconds). Any attempt exceeding 120 seconds was considered unsuccessful in locating the nerve plexus.

Educational Material

Traditional Educational Method

After a theoretical presentation by an expert, the students were exposed to instructor-curated audiovisual content to augment their comprehension. These materials included videos addressing the 3 principal topics covered in the theoretical lectures. The videos were carefully selected to complement the theoretical content by providing practical demonstrations and visual explanations. The instructor had the option of using YouTube videos, which are often freely accessible and exhibit a range of real-world examples, or consulting a textbook on RA or nerve blocks, which typically offers a more structured and academic perspective with detailed and evidence-based information. The choice between these resources depended on the teaching methodology and the desired depth of the material. While YouTube videos can present practical, real-life applications and demonstrations, textbooks generally provide a more comprehensive and authoritative source of knowledge supported by scholarly references.

Simulation Educational Method

Following a concise theoretical introduction by the same expert, the simulation group engaged in a 2-hour practical session at the Faculty of Science, University of Salamanca. During this session, the participants practiced identifying the brachial plexus using the simulation system.

The ultrasound imaging system used comprised theoretical and practical components.

The Ultrasound Desktop Educational Simulator helped students in acquiring ultrasound images, comprehending the spatial relationship between anatomical structures in the ultrasound image and the overall 3D view of the patient's anatomy, and coordinating hand movements with structure visualization. A significant feature of the Ultrasound Desktop Educational Simulator was a realistic ultrasound probe replica that trainees could operate as if it were an actual ultrasound transducer (Figure 1).

Figure 1. Illustration of the probe's movements as recorded by the software, which are used to enhance the precision of the ultrasound-guided approach.



The simulation system screen displayed a 2D image of the corresponding anatomical pathways. The movement of the probe

induced changes in both the skin surface and ultrasound images (Figure 2).

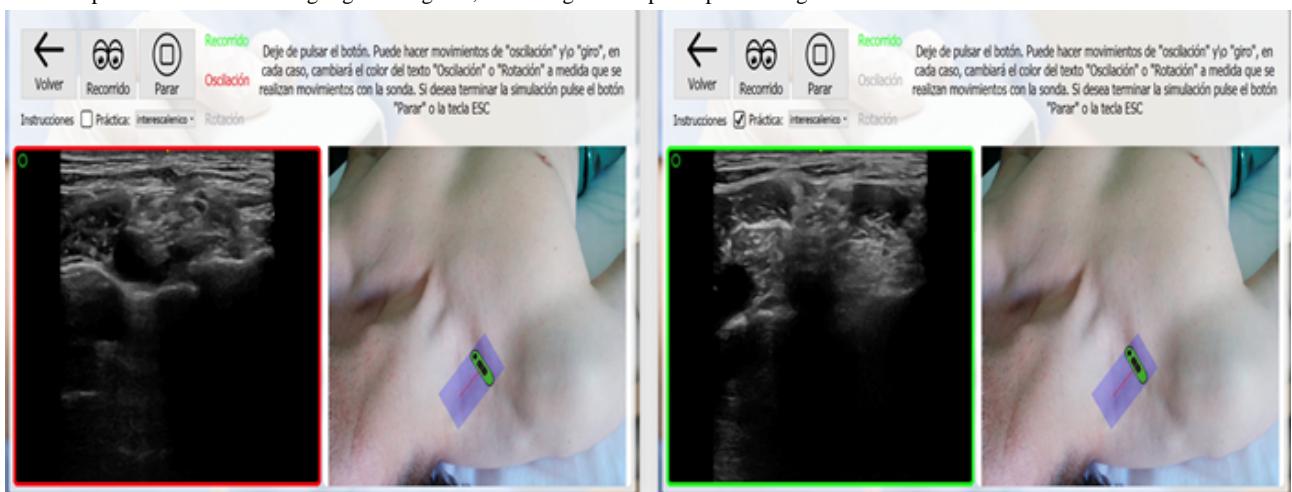
Figure 2. Simulated work environment designed for students, allowing them to perform tasks collaboratively in small groups.



A guide informed the students when they captured correct images (Figure 3). The simulation tool used in this study comprised 2 components: a probe replica that emulated a real ultrasound transducer and a multiprogram desktop application. The probe replica, designated as HUSP by its developers, contained sensors that monitored the trainee's hand movements. The software application generated ultrasound images that were

prerecorded using real equipment based on the trainee's probe manipulations. It also enabled students to capture ultrasound images, anatomical surface images, and probe paths. Screen borders appeared when the student obtained correct ultrasound images. On April 23, 2018, the Spanish Trademark and Patent Office granted utility model number 201700521 to protect the HUSP device [16].

Figure 3. Depiction of the student highlighted in green, indicating correct probe positioning for the intervention.



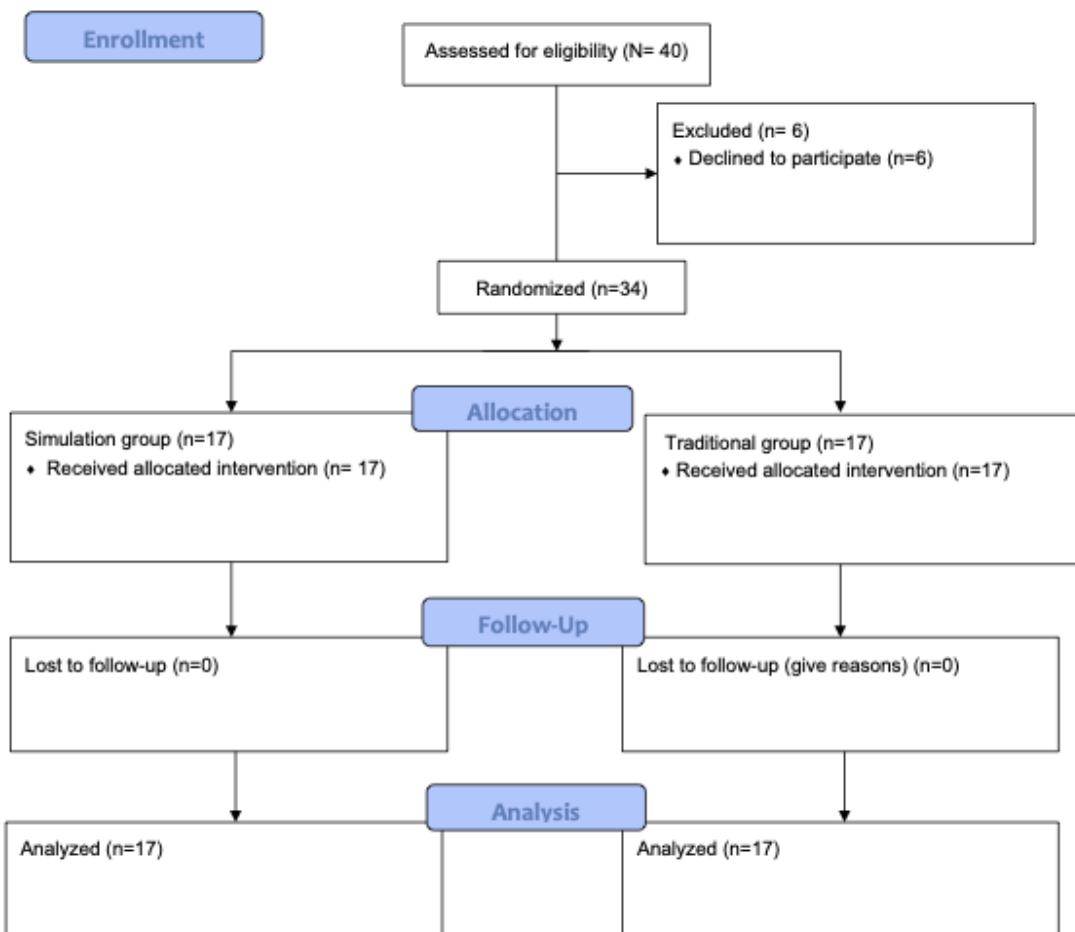
Outcome Measures

The primary outcome variable was the success rate of ultrasound-guided peripheral nerve identification, recorded as a binary outcome (success or failure). The secondary outcome was the time (in seconds) required to complete each nerve localization procedure. The procedures were analyzed across 4 anatomical approaches: interscalene, supraclavicular, popliteal, and sciatic nerve division. The academic year of the students (fourth, fifth, or sixth year) was also included.

Statistical Analysis

All data are expressed as the mean and SD. Descriptive statistics were used to summarize the participants' demographic characteristics (age, sex, and academic year) and procedure times (in seconds). Categorical variables were reported as absolute frequencies and percentages. Data normality was assessed by visually inspecting histograms. Additionally, the Shapiro-Wilk test was performed to formally evaluate the normality of the data. Given the pilot nature of this study and the limited number of participants per group, an intention-to-treat analysis was performed [19]. The primary outcome, success in nerve identification, was evaluated using the chi-square or Fisher exact tests to compare the proportions

Figure 4. Flowchart of the participants in this study.



The cohort comprised students in the fourth, fifth, and sixth years of the medical degree program. Comprehensive demographic data, including course year, sex, and age, are

between the traditional and simulation-based teaching groups for each anatomical approach. For the secondary outcome (procedure time), mixed linear models were applied to assess the mean differences between groups, including the variable “academic year” as a covariate. This allowed for the evaluation of potential confounding and interaction effects on the results of the study. Interaction terms (group × time) were tested to explore whether the effectiveness of the intervention varied over time [19]. Unstandardized coefficients, 95% CIs, and *P* values were reported for all regression models. A significance level of *P*<.05 was considered statistically significant. All statistical analyses were conducted using the JASP software (version 0.16.3).

Results

Recruitment, Participant Flow, and Sample Characteristics

A cohort of 34 students from the Faculty of Medicine at the University of Salamanca (medical degree program) was recruited for this study. All students provided voluntary consent to participate (Figure 4).

presented in Table 1. Furthermore, detailed information stratified by course year and duration of each ultrasound procedure is provided in Table 2.

Table 1. Demographic characteristics of undergraduate medical students participating in a randomized educational intervention study comparing a teaching simulator with traditional education for regional anesthesia training.

Variable	All participants (N=34)	Teaching simulator (n=17)	Traditional education (n=17)
Sex, n (%)			
Male	11 (32)	5 (29)	6 (35)
Female	23 (68)	12 (71)	11 (65)
Age (years), mean (SD)	21.8 (3.3)	22.41 (1.3)	22.4 (0.9)
Year of study, n (%)			
Fourth	11 (32)	6 (35.3)	5 (29)
Fifth	15 (44)	7 (41)	8 (47)
Sixth	8 (24)	4 (23.5)	4 (24)

Table 2. Comparison of procedure duration across different ultrasound-guided nerve block approaches between undergraduate medical students trained using a simulator-based method and traditional education.

Approach	Educational intervention (s), mean (SD)	
	Teaching simulator	Traditional education
Interscalene	38.6 (46.2)	31.41 (28.5)
Supraclavicular	51.2 (41.2)	55.7 (41.2)
Popliteal	37.8 (36.3)	42.4 (39.5)
Sciatic nerve division	10.8 (14.5)	24.1 (36.8)

Success Rate of the Intervention

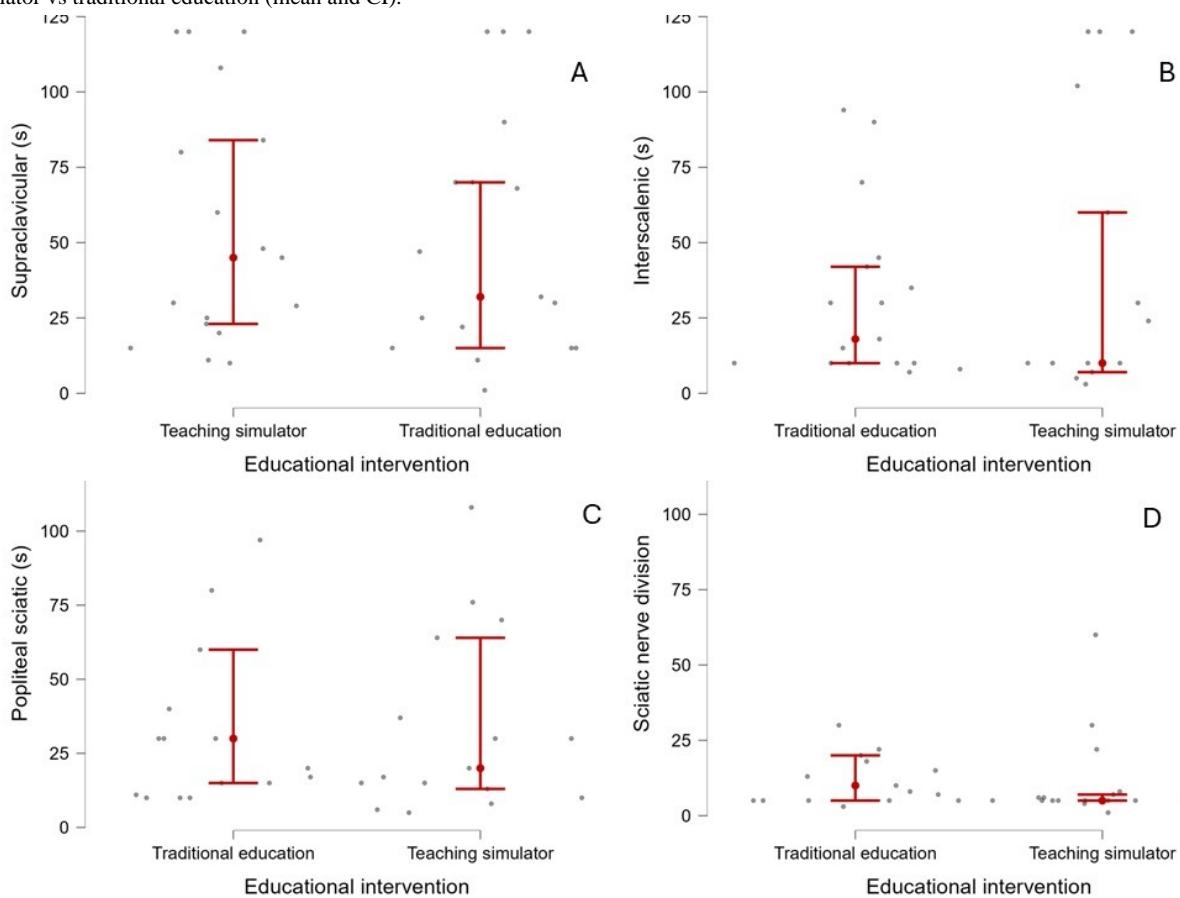
The results revealed no significant differences in successful nerve localization across the various approaches between the traditional teaching and simulation groups. Specifically, in the interscalene approach, 100% (17/17) of the students in the traditional teaching group successfully located the brachial plexus compared to 82% (14/17) in the simulation group, where 18% (3/17) of the students did not succeed. However, no statistically significant difference was observed ($P=.07$). Similarly, in the supraclavicular approach, both groups demonstrated comparable success, with 18% (3/17) of the students in each group unable to locate the relevant structure. In popliteal sciatic nerve localization, slightly fewer students in the simulation group (1/17, 6%) failed than in the traditional teaching group (2/17, 12%); however, this difference was not statistically significant ($P=.07$). Finally, for sciatic nerve division, all students in the simulation group successfully

identified the division, whereas 12% (2/17) of the students from the traditional group, one from the fourth year and one from the fifth year, were unable to do so.

Supraclavicular Approach

Specifically, there were no statistically significant differences based on the educational interventions received. The reported values correspond to unstandardized coefficients (β), which represent the effect of each intervention in the original units of the outcome variable along with their P values and 95% CIs. Both educational interventions had positive impacts: the simulator group achieved β of 106.72 (95% CI 10.06-203.37; $P=.03$), whereas the traditional teaching group achieved β of 102.80 (95% CI 5.03-200.57; $P=.40$; [Figure 5A](#)). The academic year covariate in the regression model did not interact with the effect of the educational interventions ($\beta=-10.43$, 95% CI -29.79 to 8.91; $P=.28$).

Figure 5. Scatterplots showing the time (in seconds) required to perform 4 types of nerve blocks according to the educational intervention: teaching simulator vs traditional education (mean and CI).



Interscalenic Approach

The analysis revealed no statistically significant differences based on the educational methods used. Both intervention groups (traditional education and simulator-based education) exhibited positive effects. The simulator-based group achieved β of 79.38 (95% CI 24.78-164.73; $P=.01$), whereas the traditional teaching group achieved β of 74.36 (95% CI 9.03-145.23; $P=.03$; **Figure 5B**). In the regression model, the academic year covariate did not interact with the effect of the educational interventions ($\beta=-8.69$, 95% CI -22.39 to 4.52; $P=.14$).

Popliteal Sciatic Nerve Approach

The analysis demonstrated no statistically significant differences based on the educational methods. The simulator-based group attained β of 98.93 (95% CI 32.54-188.50; $P=.07$), whereas the traditional teaching group achieved β of 105.46 (95% CI 27.59-195.32; $P=.01$; **Figure 5C**). In the regression model, the academic year covariate did not interact with the effect of the educational interventions ($\beta=-12.63$, 95% CI -28.91 to 1.33; $P=.08$).

Sciatic Nerve Division Approach

The analysis demonstrated statistically significant differences based on the educational method used. The simulator-based group attained β of 60.72 (95% CI -3.63 to 125.07; $P=.06$), whereas the traditional teaching group achieved β of 74.67 (95% CI 9.57-139.77; $P=.02$; **Figure 5D**). Consequently, the traditional educational intervention demonstrated a statistically significant

effect, whereas the simulator-based intervention did not reach statistical significance. In the regression model, the academic year covariate did not interact with the effect of the educational interventions ($\beta=-10.22$, 95% CI -23.11 to 2.66; $P=.08$).

Discussion

Principal Findings

This study evaluated the effectiveness of a simulation-based educational tool (HUSP) compared to traditional teaching for the acquisition of RA competencies in undergraduate medical students. Overall, no statistically significant differences were observed between the 2 groups regarding successful nerve localization across most approaches. In the interscalene approach, 100% (17/17) of the students in the traditional group successfully identified the brachial plexus compared with 82% (14/17) in the simulation group. Similar results were observed for the supraclavicular and popliteal approaches. In the sciatic nerve division approach, all students in the simulation group succeeded, whereas 12% (2/17) of the students in the traditional group failed, reflecting a statistically significant advantage of simulation-based training over traditional teaching in this specific case. No significant interaction was observed between the academic year and intervention type.

These findings align with those of previous studies supporting the use of simulation technologies for teaching ultrasound-guided procedures [20-22]. Simulation allows for safe and standardized practice of complex procedures and has

been shown to improve knowledge retention and practical skill acquisition among medical students [23]. In this study, the HUSP simulator enabled students to reach level 2 of the Kirkpatrick evaluation model, focusing on the acquisition of knowledge and skills following educational intervention. Specifically, students learned to identify the relevant anatomical landmarks for nerve block procedures using ultrasound. Achieving higher levels of clinical competence requires complementary training involving needle use, as found in cadaver laboratories or high-fidelity simulators. During this study, an additional software module integrating needle simulation was developed, enhancing the potential of the HUSP simulator to support advanced skill acquisition.

Chen et al [2] demonstrated the effectiveness of simulation-based learning and emphasized the need for clarification of optimal simulation modalities. Other studies have reported that using live models or fresh cadaver limbs in elective courses can improve anatomical understanding and provide a valuable clinical context [24]. The HUSP simulator offers an effective alternative, especially when access to live models or cadavers is limited in the clinical setting. Its customizable design allows educators to create tailored simulation exercises, which is a pedagogical advantage. The recent integration of needle simulation hardware has further improved clinical applicability. In terms of cost-effectiveness, the simulator is affordable; the hardware is low cost and compatible with any PC, with the software license being the main investment [25]. These characteristics make it a viable educational solution, particularly in resource-constrained settings [26].

A key strength of this study is its contribution to improving ultrasound education during the preclinical stage by addressing

gaps identified in previous research [27]. Despite the growing interest in integrating ultrasound into undergraduate curricula, previous studies have faced limitations such as restricted access to resources, faculty availability, and methodological rigor [28]. The development of a simulator incorporating needle simulation represents a significant step toward comprehensive interventional training for the HUSP system. This study has several limitations. First, the relatively small sample size, dictated by the fixed academic cohort, may limit the generalizability of the findings, a common challenge in similar educational studies [15]. Second, the lack of blinding and the dual role of some authors as both instructors and researchers could have introduced bias. Furthermore, an intention-to-treat analysis was performed, ensuring that all participants were analyzed in their originally assigned groups regardless of protocol adherence, thereby minimizing bias and better reflecting real-world educational conditions.

Conclusions

Simulation-based learning is an effective tool for supporting the acquisition of competencies in RA and offers a practical, safe, and scalable alternative to traditional educational methods. Beyond individual skill acquisition, integrating simulators such as HUSP into medical curricula could standardize training, reduce variability in learner performance, and enhance patient safety by preparing students for clinical practice in a controlled environment. To strengthen the evidence and broaden its applicability, further multicenter studies with larger and more diverse cohorts are warranted, which may inform best practices and guide the implementation of simulation-based approaches in medical education programs.

Acknowledgments

The authors would like to thank all participants of this study. The authors verify and take full responsibility for the use of generative artificial intelligence in the preparation of this manuscript. The authors used Grammarly (Superhuman Platform Inc) to improve the grammar and clarity of the manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Funding

No specific funding was received for this work.

Authors' Contributions

Conceptualization: DS-P, JV-M

Data curation: JV-M

Formal analysis: JV-M

Methodology: FH-Z, PA-H, VM

Software and hardware design: VM, BC, JAG-E

Validation: FH-Z, PA-H, VM, BC

Writing – original draft: DS-P

Writing – review & editing: JV-M, FH-Z, PA-H, DS-P

All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT checklist.

[[PDF File \(Adobe PDF File\), 222 KB - mededu_v12i1e77702_app1.pdf](#)]

References

1. Wang S, Hossack JA, Klibanov AL. From anatomy to functional and molecular biomarker imaging and therapy: ultrasound is safe, ultrafast, portable, and inexpensive. *Invest Radiol* 2020 Sep 17;55(9):559-572 [[FREE Full text](#)] [doi: [10.1097/RLI.0000000000000675](https://doi.org/10.1097/RLI.0000000000000675)] [Medline: [32776766](#)]
2. Chen WT, Kang YN, Wang TC, Lin CW, Cheng CY, Suk FM, et al. Does ultrasound education improve anatomy learning? Effects of the Parallel Ultrasound Hands-on (PUSH) undergraduate medicine course. *BMC Med Educ* 2022 Mar 27;22(1):207 [[FREE Full text](#)] [doi: [10.1186/s12909-022-03255-4](https://doi.org/10.1186/s12909-022-03255-4)] [Medline: [35346161](#)]
3. Spampinato MD, Luppi F, Cristofaro E, Benedetto M, Cianci A, Bachechi T, et al. Diagnostic accuracy of Point Of Care UltraSound (POCUS) in clinical practice: a retrospective, emergency department based study. *J Clin Ultrasound* 2024;52(3):255-264. [doi: [10.1002/jcu.23619](https://doi.org/10.1002/jcu.23619)] [Medline: [38059395](#)]
4. Popat A, Harikrishnan S, Seby N, Sen U, Patel SK, Mittal L, et al. Utilization of point-of-care ultrasound as an imaging modality in the emergency department: a systematic review and meta-analysis. *Cureus* 2024 Jan;16(1):e52371 [[FREE Full text](#)] [doi: [10.7759/cureus.52371](https://doi.org/10.7759/cureus.52371)] [Medline: [38694948](#)]
5. Szabó GV, Szigetváry C, Szabó L, Dembrovszky F, Rottler M, Ocskay K, et al. Point-of-care ultrasound improves clinical outcomes in patients with acute onset dyspnea: a systematic review and meta-analysis. *Intern Emerg Med* 2023 Mar;18(2):639-653 [[FREE Full text](#)] [doi: [10.1007/s11739-022-03126-2](https://doi.org/10.1007/s11739-022-03126-2)] [Medline: [36310302](#)]
6. Cid-Serra X, Hoang W, El-Ansary D, Carty D, Royse A, Royse C. Clinical impact of point-of-care ultrasound in internal medicine inpatients: a systematic review. *Ultrasound Med Biol* 2022 Feb;48(2):170-179. [doi: [10.1016/j.ultrasmedbio.2021.09.013](https://doi.org/10.1016/j.ultrasmedbio.2021.09.013)] [Medline: [34740496](#)]
7. Allen AJ, White AB, Bacon DR, Beck Dallaghan GL, Jordan SG. Commentary on ultrasound instruction in undergraduate medical education: perspective from two students. *Adv Med Educ Pract* 2023 Jan;Volume 14:1-7. [doi: [10.2147/amep.s388044](https://doi.org/10.2147/amep.s388044)]
8. Shanks A, Darwish A, Cook M, Asencio I, Rouse C. Integration of ultrasound simulation to improve medical student knowledge and satisfaction on the obstetrics and gynecology clerkship. *AJOG Glob Rep* 2023 Aug;3(3):100228 [[FREE Full text](#)] [doi: [10.1016/j.xagr.2023.100228](https://doi.org/10.1016/j.xagr.2023.100228)] [Medline: [37645647](#)]
9. Woodworth GE, Carney PA, Cohen JM, Kopp SL, Vokach-Brodsky LE, Horn JE, et al. Development and validation of an assessment of regional anesthesia ultrasound interpretation skills. *Reg Anesth Pain Med* 2015;40(4):306-314. [doi: [10.1097/AAP.0000000000000236](https://doi.org/10.1097/AAP.0000000000000236)] [Medline: [26017720](#)]
10. Shafqat A, Ferguson E, Thanawala V, Bedford NM, Hardman JG, McCahon RA. Visuospatial ability as a predictor of novice performance in ultrasound-guided regional anesthesia. *Anesthesiology* 2015 Nov;123(5):1188-1197. [doi: [10.1097/ALN.0000000000000870](https://doi.org/10.1097/ALN.0000000000000870)] [Medline: [26397142](#)]
11. Albrecht E, Chin KJ. Advances in regional anaesthesia and acute pain management: a narrative review. *Anaesthesia* 2020 Jan;75 Suppl 1:e101-e110 [[FREE Full text](#)] [doi: [10.1111/anae.14868](https://doi.org/10.1111/anae.14868)] [Medline: [31903582](#)]
12. Turbitt LR, Mariano ER, El-Boghdadly K. Future directions in regional anaesthesia: not just for the cognoscenti. *Anaesthesia* 2020 Mar 03;75(3):293-297 [[FREE Full text](#)] [doi: [10.1111/anae.14768](https://doi.org/10.1111/anae.14768)] [Medline: [31268173](#)]
13. Sekhavati P, R Logan R, Bailey JG, Busse JW, Boet S, Gu Y. Simulation-based ultrasound-guided regional anesthesia education: a national survey of Canadian anesthesiology residency training programs. *Can J Anaesth* 2025 Jan 06;72(1):173-180. [doi: [10.1007/s12630-024-02818-x](https://doi.org/10.1007/s12630-024-02818-x)] [Medline: [39107539](#)]
14. Satapathy AR, Khalid IB, Fathil SM. Simulation in regional anaesthesia: a narrative review of its history, evolution and future prospects. *J Clin Med* 2024 Dec 26;14(1):67 [[FREE Full text](#)] [doi: [10.3390/jcm14010067](https://doi.org/10.3390/jcm14010067)] [Medline: [39797150](#)]
15. Ashokka B, Law LS, Areti A, Burckett-St Laurent D, Zuercher RO, Chin KJ, et al. Educational outcomes of simulation-based training in regional anaesthesia: a scoping review. *Br J Anaesth* 2025 Feb;134(2):523-534. [doi: [10.1016/j.bja.2024.07.037](https://doi.org/10.1016/j.bja.2024.07.037)] [Medline: [39358185](#)]
16. Moreno V, Curto B, Garcia-Estebar JA, Hernández Zaballos F, Alonso Hernández P, Serrano FJ. HUSP: a smart haptic probe for reliable training in musculoskeletal evaluation using motion sensors. *Sensors (Basel)* 2018 Dec 29;19(1):101 [[FREE Full text](#)] [doi: [10.3390/s19010101](https://doi.org/10.3390/s19010101)] [Medline: [30597949](#)]
17. Hopewell S, Chan AW, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 statement: updated guideline for reporting randomized trials. *Nat Med* 2025 Jun;31(6):1776-1783. [doi: [10.1038/s41591-025-03635-5](https://doi.org/10.1038/s41591-025-03635-5)] [Medline: [40229553](#)]
18. WMA Declaration of Helsinki – ethical principles for medical research involving human participants. World Medical Association. URL: <https://www.wma.net/policies-post/wma-declaration-of-helsinki/> [accessed 2026-01-29]

19. Twisk JWR. Applied longitudinal data analysis for epidemiology: a practical guide. 2nd ed. Cambridge, UK: Cambridge University Press; 2013.
20. Recker F, Neubauer R, Dong Y, Gschmack AM, Jenssen C, Möller K, et al. Exploring the dynamics of ultrasound training in medical education: current trends, debates, and approaches to didactics and hands-on learning. *BMC Med Educ* 2024 Nov 14;24(1):1311 [FREE Full text] [doi: [10.1186/s12909-024-06092-9](https://doi.org/10.1186/s12909-024-06092-9)] [Medline: [39543551](#)]
21. Haji-Hassan M, Capra RD, Bolboacă SD. Efficacy of handheld ultrasound in medical education: a comprehensive systematic review and narrative analysis. *Diagnostics (Basel)* 2023 Dec 14;13(24):3665 [FREE Full text] [doi: [10.3390/diagnostics13243665](https://doi.org/10.3390/diagnostics13243665)] [Medline: [38132248](#)]
22. Kaddoura R, Faraji H, Otaki F, Radhakrishnan R, Stanley A, Paulus A, et al. High-fidelity simulation versus case-based tutorial sessions for teaching pharmacology: convergent mixed methods research investigating undergraduate medical students' performance and perception. *PLoS One* 2024 Aug 16;19(8):e0302609 [FREE Full text] [doi: [10.1371/journal.pone.0302609](https://doi.org/10.1371/journal.pone.0302609)] [Medline: [39150900](#)]
23. Huang Q, Yan SY, Huang J, Guo Y, Zeng XT, Jin YH. Effectiveness of simulation-based clinical research curriculum for undergraduate medical students - a pre-post intervention study with external control. *BMC Med Educ* 2024 May 15;24(1):542 [FREE Full text] [doi: [10.1186/s12909-024-05455-6](https://doi.org/10.1186/s12909-024-05455-6)] [Medline: [38750452](#)]
24. Huettner DP, Hahn LK, Noonan Haase MA, Jung FA, Orebaugh SL. Medical student instruction in peripheral nerve blockade utilizing fresh cadaver limbs in a simulation center. *J Educ Perioper Med* 2023;25(2):E703 [FREE Full text] [doi: [10.46374/volxxv_issue2_Noonan_Haase](https://doi.org/10.46374/volxxv_issue2_Noonan_Haase)] [Medline: [37377509](#)]
25. Ylönen M, Paajanen P, Kukkonen T, Torkki P, Paloneva J, Rosqvist E. From emergency room to operating room: multidisciplinary simulation training in emergency laparotomy for ruptured abdominal aortic aneurysm - learning outcomes and costs. *Ann Vasc Surg* 2025 Feb;111:341-350. [doi: [10.1016/j.avsg.2024.11.017](https://doi.org/10.1016/j.avsg.2024.11.017)] [Medline: [39617299](#)]
26. Gasco J, Holbrook TJ, Patel A, Smith A, Paulson D, Muns A, et al. Neurosurgery simulation in residency training: feasibility, cost, and educational benefit. *Neurosurgery* 2013 Oct;73 Suppl 1:39-45. [doi: [10.1227/NEU.0000000000000102](https://doi.org/10.1227/NEU.0000000000000102)] [Medline: [24051881](#)]
27. Draper R, Ellinger D, Trottier M, Armstrong K, Ma W. Ultrasound education in an elective course in the undergraduate medical curriculum improves students' subjective learning and clinical rotation and residency preparedness. *Med Sci Educ* 2023 Jun 11;33(3):737-746 [FREE Full text] [doi: [10.1007/s40670-023-01800-4](https://doi.org/10.1007/s40670-023-01800-4)] [Medline: [37501805](#)]
28. Neubauer R, Bauer CJ, Dietrich CF, Strizek B, Schäfer VS, Recker F. Evidence-based ultrasound education? - A systematic literature review of undergraduate ultrasound training studies. *Ultrasound Int Open* 2024 Mar 26;10(CP):a22750702 [FREE Full text] [doi: [10.1055/a-2275-0702](https://doi.org/10.1055/a-2275-0702)] [Medline: [39403546](#)]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials

HUSP: Haptic Ultrasound Probe

RA: regional anesthesia

Edited by T Leung; submitted 18.May.2025; peer-reviewed by J Wilkinson, G McLeod; comments to author 20.Sep.2025; revised version received 09.Dec.2025; accepted 31.Dec.2025; published 03.Feb.2026.

Please cite as:

Sánchez-Poveda D, Vicente-Mampel J, Curto B, Moreno V, García-Esteban JA, Hernández-Zaballos F, Alonso-Hernández P. Comparative Efficacy of Simulation-Based and Traditional Training in Ultrasound-Assisted Regional Anesthesia for Medical Students: Randomized Controlled Trial. *JMIR Med Educ* 2026;12:e77702

URL: <https://mededu.jmir.org/2026/1/e77702>

doi: [10.2196/77702](https://doi.org/10.2196/77702)

PMID:

©David Sánchez-Poveda, Juan Vicente-Mampel, Belén Curto, Vidal Moreno, Juan A García-Esteban, Felipe Hernández-Zaballos, Pablo Alonso-Hernández. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 03.Feb.2026. 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.

Influence of Admission Pathways on Learning Strategies, Assessment Engagement, and Academic Performance Among First-Year Medical Students: Mixed Methods Retrospective Observational and Cross-Sectional Survey Study

Issarawan Keadkraichaiwat, MSc; Chantacha Sitticharoon, MD, PhD; Punyapat Maprapho, MD; Nisa Jangboon, BEd, MSc; Nadda Wannarat, BA, MEd

Faculty of Medicine Siriraj Hospital, Mahidol University, 2 Wanglang Rd, Siriraj, Bangkoknoi, Thailand

Corresponding Author:

Chantacha Sitticharoon, MD, PhD

Faculty of Medicine Siriraj Hospital, Mahidol University, 2 Wanglang Rd, Siriraj, Bangkoknoi, Thailand

Abstract

Background: Medical school admission pathways are designed to select suitable applicants, with different approaches potentially impacting students' learning behaviors and performance.

Objective: This study aimed to compare students' self-regulated learning (SRL) strategies, assessment engagement statistics (AES), nongrading evaluation (Outstanding ["O"]/Satisfactory ["S"]/Unsatisfactory ["U"]) preferences, and academic performance across admission pathways, and analyze correlations and linear regression models among summative scores, AES, and course learning outcome (CLO) scores.

Methods: This mixed methods retrospective observational and cross-sectional survey study used census sampling with selection criteria of all enrolled first-year medical students in 2021 (N=319) across 4 admission pathways: academic (n=23), quota (n=6), test (n=261), and rural (n=29). Demographics included age (19 - 24 years) and sex (167/319, 52.4% male). AES, CLO scores, and summative scores were obtained from institutional databases. Two system-embedded institutional questionnaires assessed SRL strategies (316/319, 99.1% response rate) and "O"/"S"/"U" preferences (299/319, 93.7% response rate). Outcome measures included SRL strategies, AES, "O"/"S"/"U" preferences, CLO scores, and summative scores. Statistical significance was set at $P<.05$.

Results: When compared among pathways, using one-way ANOVA with Fisher least significant difference post hoc tests, the academic group reported significantly higher mean (with 95% CI) goal setting (4.35, 4.07 - 4.63), enthusiasm (4.43, 4.18 - 4.69), and lower stress during study (2.64, 2.15 - 3.12), while the rural group showed higher pre-examination stress (4.38, 4.10 - 4.66) (all $P<.05$). Most academic (14/22, 63.6%), quota (5/6, 83.3%), and test students (132/243, 54.3%) preferred "O"/ "S"/ "U," while the rural students preferred "S"/"U" (13/28, 46.4%). The academic group showed significantly higher CLO and summative scores but fewer total and intentional attempts and instances of first-pass and highest scoring attempts (all $P<.05$), whereas the rural group showed significantly lower CLO and summative scores and higher instances of first-pass and highest scoring attempts (all $P<.05$). For correlation analyses, using Pearson correlation coefficient, summative scores were positively correlated with CLO scores and number of passings and negatively with first-pass attempts. For multiple linear regression analyses, summative scores were positively influenced by number of passings for each CLO and CLO scores and negatively influenced by instances of first-pass attempts and highest scoring attempts. Overall, the academic group demonstrated higher academic performance and fewer attempts and instances of first-pass and highest scoring attempts, while the rural group showed lower academic performance, requiring more attempts for first-passing CLOs.

Conclusions: Admission pathways significantly influence students' SRL strategies, AES, evaluation preferences, and academic performance. This study is innovative in analyzing these interconnected components within a single cohort, unlike prior research that examined them separately. By integrating assessment-engagement analytics with SRL data, it offers equity-oriented evidence on how admission systems shape learning behaviors and academic trajectories. These findings provide actionable insights for inclusive curriculum design and early identification of at-risk students. Real-world implications include targeted mentoring, SRL-focused interventions, and assessment reforms balancing academic rigor with psychological safety.

(*JMIR Med Educ* 2026;12:e68636) doi:[10.2196/68636](https://doi.org/10.2196/68636)

KEYWORDS

admission systems; self-regulated learning strategies; assessment engagement statistics; academic outcome; course learning outcome; nongrading evaluation

Introduction

To address the diversity of medical curricula, institutions use various admission pathways, which may impact students' academic outcomes, self-regulated learning (SRL) strategies, assessment engagement statistics, and evaluation preferences. In comparisons among admission pathways, there were differences in academic outcomes and the likelihood of on-time graduation [1]. In terms of academic performance, students admitted through the Open National Unified Admission track, characterized by the highest competitiveness and a reserved quota for underprivileged students, predominantly comprising children of military personnel and government employees within the Ministry of Education in Jordan, had significantly higher graduate grade point average (GPA) than those in other tracks, whereas students from the other pathway track and an international track exhibited the lowest graduating GPA [1]. Furthermore, students entering through the GENERAL admission pathway in New Zealand demonstrated significantly higher GPAs in years 2 - 3 than students admitted through the Māori and Pacific Admission Scheme, which was designed to address shortages in the Māori and Pacific health workforce, reduce significant disparities in health outcomes, and uphold the rights of Māori within the New Zealand pathway [2]. In addition, multiple regression analysis demonstrated that the undergraduate entry pathway positively contributed to GPA in years 2 - 3 in New Zealand [2].

For students from various academic backgrounds, health-related undergraduate degree students consistently surpassed those from BioMed, Science, Humanities, and Business undergraduate degree programs in performance on medical science, clinical practice, objective structured clinical examinations, and overall performance when compared to other pathways in Australia [3].

For graduation rates, students admitted through the competitive Open National Unified Admission track and the reserved quota for underprivileged student track achieved the highest percentage of on-time graduation, followed by the children of university staff track in Jordan [1]. Similarly, students who gained admission through the GENERAL pathway in New Zealand exhibited a greater likelihood of completing their intended program than those from the Māori and Pacific Admission Scheme pathway [2].

Admission to the Faculty of Medicine Siriraj Hospital in the 2021 academic year followed the Thai University Central Admission System (TCAS) framework and aligned with the broader educational policy goals of Thailand's national medical school admission framework, which emphasizes not only academic merit but also equity, inclusion, and workforce distribution [4].

Importantly, at the time of the 2021 admission cycle, TCAS is structured as 4 sequential rounds and operates as a binding system, that is, once students are accepted and confirm their

place in an earlier round, they cannot participate in subsequent rounds. This structure ensures efficient allocation, reduces redundancy, and facilitates early selection of high-priority candidates. Round 1 emphasizes portfolio submission and interviews, making it suitable for applicants with strong academic records and extracurricular achievements, while round 2 follows a quota system designed for students from specific regions or with special talents. Round 3 relies on standardized examinations such as the Thai General Aptitude Test, the Thai Professional Aptitude Test (TPAT), and core subjects, favoring applicants with strong test performance, whereas round 4 enables institutions with unfilled seats to admit students through their own criteria, offering opportunities for applicants not accepted in earlier rounds [5].

At our institution, via 3 TCAS rounds, 4 pathways (groups) were established: round 1—the academic group targeting high-potential students with demonstrated cognitive strengths and scientific aptitude, particularly those participating in national science Olympiads; round 2—the quota group, promoting diversity by admitting students with exceptional talents in music or sports and individuals holding a bachelor's degree; and round 3—the central admission test pathway (test group) and the pathway for students required to serve as rural doctors (rural group), prioritizing academic achievement through competitive entrance examinations [5,6]. Notably, students in the rural group were officially enrolled under the Faculty of Medicine, Praboromarajchanok Institute, but undertook their preclinical study together with students of the Faculty of Medicine Siriraj Hospital.

Differences in admission pathways may be associated with variations in students' SRL behaviors. According to Zimmerman's cyclical model, SRL encompasses 3 interrelated phases: the forethought phase, which involves goal setting, task analysis, and motivational beliefs; the performance phase, which includes self-control, monitoring, and strategy implementation; and the self-reflection phase, which focuses on self-evaluation and adaptive reactions after learning tasks [7]. SRL may influence students across various admission systems, since it influences how they approach learning tasks, manage challenges, and engage with evaluation, and plays a crucial role in their academic success, particularly among students with different backgrounds [8].

In addition, nongrading or mastery-based systems, such as pass/fail or pass/satisfactory/outstanding formats, have gained attention for their potential to reduce stress [9], promote intrinsic motivation [10-12], and foster deeper, self-directed learning [10]. These approaches are especially relevant in high-stakes disciplines such as medicine, where long-term competence and well-being are critical [13], probably reducing competitiveness and anxiety while promoting equity and fostering collaborative learning environments [10,14,15]. Despite the growing adoption of nongrading evaluation systems in medical education as part of curricular reform, limited evidence exists on how students

from different admission tracks perceive these systems or how such reforms influence their SRL strategies and academic performance.

Student feedback is widely recognized as a crucial mechanism for curriculum evaluation and improvement, serving purposes such as quality assurance, course and program revision, and the evaluation of teaching quality and assessment tasks, while also being critical to student learning by influencing motivation, engagement, self-reflective learning, and overall performance, as well as supporting iterative refinement and long-term acceptance [16-19]. Understanding the perceptions of students from different admission pathways is particularly important, as their diverse educational, social, and cultural backgrounds may shape how they interpret and adapt to new evaluation systems, potentially influencing not only their motivation and learning behaviors but also the effectiveness of these systems in supporting diverse student cohorts [20].

Comparisons of academic outcomes among various admission pathways have been partly reported; however, comparisons of preclinical students' SRL strategies, assessment engagement statistics, preferences regarding nongrading evaluation, and course learning outcomes (CLOs) and summative scores have not been studied. This study aimed to (1) compare students' SRL strategies, assessment engagement statistics, preferences regarding nongrading evaluation, and academic performance among the academic, quota, test, and rural groups; (2) determine correlations among summative scores, students' assessment engagement statistics, and CLO scores in each group; and (3) identify factors contributing to summative scores in each group through multiple linear regression analyses. By addressing these gaps, our study provides a comprehensive understanding of how admission pathways may shape not only academic outcomes but also learning behaviors, essential for designing inclusive medical curricula that support all learners [7].

Methods

Study Design

This study used a mixed methods design consisting of a retrospective observational component and a cross-sectional survey component. The retrospective observational component included assessment engagement statistics and CLO scores obtained from the Siriraj E-Learning and Education Community, as well as summative examination scores retrieved from the Siriraj Campus Management System (SiCMs). The cross-sectional component consisted of 2 system-embedded institutional questionnaires: questionnaire 1 assessing students' SRL strategies and questionnaire 2 assessing preferences regarding nongrading evaluation, both administered within the SiCMs platform. The integration of quantitative and qualitatively informed data enabled a comprehensive examination of learning behaviors and academic performance across admission pathways.

Although this study adopted a mixed methods framework, the qualitative component was limited to a qualitatively informed instrument development process based on literature review and expert input rather than primary qualitative data collection.

Therefore, formal qualitative data saturation was not applicable. Instead, content validity was ensured through iterative expert review and consensus to confirm comprehensive coverage of relevant educational constructs.

Study Protocol

Participants included all first-year medical students enrolled in the academic year 2021 at the Faculty of Medicine Siriraj Hospital, Mahidol University, using a census sampling approach. Selection criteria included all officially enrolled students, with no exclusions applied.

Ethical Considerations

This study used routinely collected educational data and system-embedded questionnaires. The research protocol was reviewed and approved by the Siriraj Institutional Review Board. The initial protocol was granted exemption from full review under protocol number 370/2565 (exempt). A subsequent amendment, which included additional demographic data (age, school region, and hometown region), was approved under certificate of approval number 731/2025. For the retrospective observational component, the institutional review board granted a waiver of written informed consent because the data were obtained from existing institutional databases, posed minimal risk, and were analyzed in deidentified, aggregate form. For the cross-sectional component, participation involved completion of system-embedded questionnaires as part of routine institutional processes; therefore, no additional consent was required for secondary analysis of anonymized data, as approved by the institutional review board. All datasets were anonymized prior to export, with direct identifiers removed, and were stored on password-protected computers with restricted access. No financial or academic incentives were provided, and participation had no impact on academic standing. No identifiable images or sensitive personal information was included in the analysis.

Admission to the Faculty of Medicine Siriraj Hospital

For the academic group, applicants had to be high school students who had achieved gold or silver medals at the National Academic Olympiad in the following fields: Physics, Chemistry, Biology, Mathematics, Computer Science (Informatics), Geography, Astronomy, and Astrophysics; or have completed Camp 2 in Earth and Space Science. In addition, applicants could be high school students in the year of the training camp or those selected to attend the Academic Olympiad Training Camp Round 1 in Biology, Chemistry, Physics, Computer Science, or Mathematics. Furthermore, they were required to have a GPA of at least 3.00 in Science, Mathematics, English, Thai, and Social Studies. Selection was primarily based on portfolio assessment, including activities, academic achievements, and other special skills or talents, as well as a statement of purpose. Subsequently, they underwent multiple mini-interviews (MMIs), which were constructed based on the conceptual framework of the objective structured clinical examination [21], and final selection was determined according to MMIs criteria.

For the quota group, applicants must (1) have talent in sports or music at national or international levels; (2) have eligibility

for the Mahidol Medical Scholars Program, with a GPA of at least 3.00 in Science, Mathematics, English, Thai, and Social Studies; or (3) be holding a bachelor's degree with a GPA of at least 3.25 in any field in Thailand, with age not exceeding 35 years. Applicants were primarily selected based on academic examinations organized by the TCAS, accounting for 70%, including 20% Mathematics, 40% Sciences (Physics, Chemistry, and Biology), 20% English, 10% Thai, and 10% Social Studies, as well as 30% Ordinary National Educational Test for music talent, or General Aptitude Test/Professional Aptitude Test for sport talents and Mahidol Medical Scholar Program, or the TPAT organized by the Consortium of Thai Medical Schools for bachelor's graduates. Finally, they underwent MMIs for definitive selection.

For the test group, applicants were high school students selected through academic examinations weighted at 70% TCAS and 30% TPAT, as described for the quota group, along with a general interview, including both a standard personal interview and a psychiatric assessment, as well as a health examination, to screen out individuals unsuitable for medical studies. The eligibility criteria for passing these assessments were based on the national medical school admission standards, which ensure that applicants are free from any physical or mental health conditions that could hinder their medical education, clinical training, or future medical practice. These criteria are defined in the official 2016 regulation on student eligibility and include, for example, severe psychiatric disorders, active communicable diseases, significant physical disabilities, and uncorrectable hearing or visual impairments.

For the rural group, high school students were admitted under Thailand's national "Doctors for Rural Areas" policy, implemented through the Collaborative Project to Increase Production of Rural Doctors, which was established to address physician shortages in underserved areas and to reflect Thailand's national "Doctors for Rural Areas" policy, aiming to promote equity in health care distribution [22]. Students from designated provinces must have their names in the house records of the specified province continuously for not less than 5 years, attend a school in the designated area, and receive government financial support throughout medical training. Upon graduation, graduates are obligated to return to their home province or designated health region to work in public hospitals as part of a binding service requirement [22]. They are required to serve 3 years in Ministry of Public Health hospitals, with a financial penalty of approximately US \$11,300 imposed for noncompliance [22]. They were selected based on a GPA of at least 3.00, a 70% academic examination score through TCAS as the test group, and 30% interview performance. Although they were officially enrolled under the Faculty of Medicine, Praboromarajchanok Institute, they studied their preclinical years alongside the Faculty of Medicine Siriraj Hospital's students, receiving the same curriculum delivery, teaching and assessment processes, access to facilities and equipment, academic and mentoring support, and extracurricular opportunities. After finishing the preclinical curriculum, they continued their clinical training at the Clinical Education Center, Ratchaburi Hospital.

Curriculum

The Doctor of Medicine program at the Faculty of Medicine Siriraj Hospital, Mahidol University, Thailand, spans 6 years, divided into the preclinical years (years 1 - 3) and the clinical years (years 4 - 6). The latest curriculum, implemented in 2021, introduced several changes, including defining specific CLOs, transferring 4 basic biomedical science courses from year 2 of the 2017 curriculum to year 1 of the 2021 curriculum, changing from compensatory to noncompensatory evaluation, and shifting from grading to nongrading evaluation as Outstanding ("O"), Satisfactory ("S"), or Unsatisfactory ("U").

Assessment and Evaluation

For each CLO performance, students were permitted unlimited attempts, except for course 4, which allowed only 1 attempt with an 80% passing threshold. For summative assessment, examinations for 4 courses were conducted simultaneously on the same day with a minimum pass mark of 60% for courses 1 - 3 and 50% for course 4. For evaluation, students who failed the CLO assessment or the initial attempt at a summative assessment received an "X" grade, indicating that they had not yet passed. If these students subsequently passed the CLO within the specified time frame and achieved at least a 60% score on their second attempt at the summative assessment, they were awarded an "S" grade, but they forfeited eligibility for an "O" grade. To qualify for an "O" grade, students had to attain summative scores of at least 85% in courses 1, 2, and 4, and 80% in course 3, without receiving any "X" grades in any CLO, attitudinal assessments, or summative assessments. A "U" grade was given to students who failed the remediation of the CLO examination, the second attempt at the summative examination, or the assessment of their attitude.

Assessment Engagement Statistics

Students' assessment engagement statistics comprised 319 students: 23 in the academic group, 6 in the quota group, 261 in the test group, and 29 in the rural group. These statistics encompassed various aspects, including the number of total, intentional, and unintentional attempts; instances of first-pass attempt and highest scoring attempt; the number of passings for each CLO; and (an) additional attempts after passing each CLO. Across the 4 courses, course 1 included 5 CLOs, course 2 included 4 CLOs, course 3 included 5 CLOs, and course 4 included 5 CLOs.

For each CLO, the number of total attempts is the total number of times the students accessed the assessment platform. The number of intentional attempts is defined as attempts in which $\geq 50\%$ of the items were completed. The number of unintentional attempts is defined as attempts in which $< 50\%$ of the items were completed. The instances of the first-pass attempt are the first instances when the students successfully meet the passing level. The instances of the highest-scoring attempt are the instances when the students achieve the maximum score. The number of passings for each CLO is the count of instances in which the students successfully meet the passing level. The additional attempts after passing each CLO are the number of subsequent instances after students have already passed.

Questionnaires

The questionnaires were developed collaboratively by the authors under the supervision of the deputy dean for undergraduate education at the Faculty of Medicine Siriraj Hospital. The questionnaires regarding students' SRL strategies and preferences regarding nongrading evaluation among the academic, quota, test, and rural groups were administered to first-year medical students during the academic year 2021, following the summative examination.

Questionnaire 1, rated on a Likert scale from 1 (strongly disagree) to 5 (strongly agree), assessed students' perspectives and experiences across four domains: (1) teaching and learning, (2) CLOs and program learning outcomes, (3) nongrading evaluation, and (4) academic obstacles, included to identify barriers perceived by students in adapting to the new structure, learning activities, and nongrading evaluation system.

Questionnaire 1 was initially based on a university-standard template and was subsequently refined using items adapted from 2 well-established instruments: first, the University Students Learning Strategies Questionnaire, which contributed items addressing goal-setting and emotional barriers to learning [23], for example, "I try to set goals before studying," "I feel discouraged from studying," and "I have concerns that interfere with my learning." Second, the Academic Motivation Scale, college version, which informed items related to intrinsic motivation (eg, "I enjoy and feel satisfied when learning new things" and "I feel happy when I achieve each course learning outcome") and amotivation (eg, "I feel like I waste time participating in class activities") [24].

Questionnaire 1 represents 3 phases of SRL strategies. The Forethought Phase (P1), representing motivation and beliefs, included items such as goal setting and enthusiasm toward learning and teaching activities. The Performance Phase (P2), reflecting monitoring and strategy use, encompassed concerns or obstacles that hindered studying such as the perceived impact of unlimited CLO assessment attempts. The Self-Reflection Phase (P3), representing assessment and self-reaction, included items related to satisfaction with nongrading evaluation and overall learning experiences gained from the current curriculum. It also explored students' perceptions of the nongrading "O"/"S"/"U" evaluation, the applicability of knowledge to future clinical and national assessments, and overall satisfaction with the curriculum. Both positively and negatively worded statements were included to support internal consistency testing.

The overall questionnaire 1 response rate was 99.1% (316/319), including 100% (23/23) in the academic group, 100% (6/6) in the quota group, 98.9% (258/261) in the test group, and 100% (29/29) in the rural group. The proportions of respondents were 7.3% (23/316) in the academic group, 1.9% (6/316) in the quota group, 81.6% (258/316) in the test group, and 9.2% (29/316) in the rural group.

Questionnaire 2, which allowed multiple responses, was developed to explore students' preferences regarding the newly implemented "O"/"S"/"U" grading system, with a particular focus on their perspectives toward the inclusion of the "O" grade, in relation to aspects of SRL strategies (Forethought,

Performance, and Self-Reflection phases). While the "O" grade may serve as a form of academic recognition that rewards effort, motivation, and high performance, several student concerns also emerged, particularly regarding its association with extrinsic motivation, heightened anxiety, and increased competitiveness. Although no literature has directly addressed the specific effects of the "O" grade itself, studies on A-F grading suggest that top grades may similarly function as motivators and thus can be reasonably used to approximate the motivational impact of the "O" grade in this context [25]. Based on these recurring themes, the research team identified measurable items that addressed key constructs frequently cited in the literature on grading systems, including stress [9], competition [10,26], study focus [26], effort, and life-goal setting.

Although this questionnaire did not undergo a full validation process and did not adopt an existing standardized instrument, its content was informed by student feedback gathered during monthly online visits by the deputy dean and assistant dean (the corresponding author, CS) of undergraduate education with the first-year cohort, internal faculty discussions, and relevant literature on the psychological and behavioral impacts of grading practices. A qualitatively informed survey design approach was used to generate a focused list of 5 questions designed to capture concrete student perspectives, thereby enabling clearer interpretation of their positions on key issues as part of the validity argument. A simplified yes/no response format was deliberately selected to enhance clarity, improve response process validity, and encourage full participation by prompting students to take a definitive stance on each item.

The overall response rate was 93.7% (299/319), with participation rates of 95.7% (22/23) in the academic group, 100% (6/6) in the quota group, 93.1% (243/261) in the test group, and 96.6% (28/29) in the rural group. The proportions of respondents were 7.4% (22/299) in the academic group, 2.0% (6/299) in the quota group, 81.3% (243/299) in the test group, and 9.4% (28/299) in the rural group.

To ensure content validity, both questionnaires were reviewed by 3 medical education experts. Their feedback was incorporated to improve item clarity, coverage, and logical flow. The finalized questionnaire was administered at the end of the academic year but before grade announcement to minimize potential bias.

Construct validity for questionnaire 1 was examined using exploratory factor analysis (principal component extraction with Varimax rotation). Sampling adequacy was acceptable (Kaiser-Meyer-Olkin Measure of Sampling Adequacy=0.812), and Bartlett's test of sphericity was significant ($\chi^2_{136}=2253.45$; $P<.001$). Four factors were extracted, explaining 62.9% of the total variance after rotation. Factor loadings supported a theoretically consistent structure aligned with SRL domains: (1) teaching and learning, (2) CLOs and program learning outcomes, (3) nongrading evaluation, and (4) academic obstacles. Most communalities were ≥ 0.50 . Internal consistency was acceptable (Cronbach $\alpha=0.729$). Content validity was ensured through expert review and adaptation from validated instruments.

For questionnaire 2, the first item assessed grading system preference and was analyzed descriptively because it represents a categorical choice rather than a latent construct. The remaining 5 dichotomous items assessed perceived consequences of the “O” grade (stress, competition, focus, effort, and goals). When treated as a single composite, internal consistency was low (Cronbach $\alpha=0.349$), indicating that these items did not form a unidimensional scale. Interitem correlations suggested 2 distinct clusters—perceived costs and perceived benefits—with negative correlations between clusters. Given this multidimensional structure and low reliability, items were analyzed individually. Construct validity was not established for questionnaire 2; however, content validity was supported through expert review and literature-informed item development.

Academic Performance

Academic performance was represented by CLO scores and summative scores, both comprising a total of 319 students: 23 in the academic group, 6 in the quota group, 261 in the test group, and 29 in the rural group.

Missing Data Handling

Missing data were assessed using the SPSS (version 30; IBM Corp) Missing Value Analysis module. The dataset for assessment engagement statistics and CLO scores included 319 students, and summative scores were available for all 319 students. No missing data were observed for questionnaire 1 (N=316) and questionnaire 2 (N=299), as well as for summative scores. For assessment engagement statistics and CLO scores, a small proportion of missing data occurred exclusively in the test group due to some students not attempting specific CLO assessments or data unavailability. Specifically, missing data were observed for CLO 1 (3 students), CLO 2 (5 students), CLO 3 (3 students), CLO 4 (6 students), and CLO 5 (4 students) in course 1; for CLO 1 (3 students), CLO 2 (3 students), CLO 3 (6 students), and CLO 4 (7 students) in course 2; and for CLOs 1 - 5 (3 students) in course 3. Across all CLO variables, the proportion of missing data ranged from 0.94% to 2.19%, which is below the commonly accepted threshold of 5%.

Little's Missing Completely at Random (MCAR) test was performed on the combined dataset of assessment engagement statistics and CLO scores using the SPSS Missing Value Analysis module. The test was nonsignificant ($\chi^2_{464}=0.000$; $P=1.000$), indicating that the observed missing data pattern was consistent with MCAR. Because the proportion of missing data was very small and the MCAR assumption was satisfied, complete case analysis was applied. Multiple imputation was not performed.

Reporting Guidelines

To enhance transparency and clarity of reporting, this study adhered to EQUATOR-aligned guidelines appropriate to its methodological structure. The STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guideline

was applied to the retrospective observational components [27], and the GRAMMS (Good Reporting of A Mixed Methods Study) guideline was applied to the mixed methods framework [28]. Completed STROBE and GRAMMS checklists have been uploaded as supplementary files.

Statistics

For statistical analysis, SPSS (version 18) was used. Participant demographics and students' preferences regarding nongrading evaluation are presented as frequencies (N) and percentages. Students' assessment engagement statistics are reported as ranges of means and medians. Comparisons of SRL, assessment engagement statistics, and CLO and summative scores among students in the academic, quota, test, and rural groups were analyzed using one-way ANOVA and are expressed as means with 95% CIs. Post hoc analyses using Fisher least significant difference test were performed, as appropriate. The correlations of summative scores with students' assessment engagement statistics and CLO scores within the academic, quota, test, and rural groups were analyzed using Pearson correlation coefficient (R) with 95% CIs. Multiple linear regression analyses were conducted to identify factors significantly associated with summative scores in these groups, with regression coefficients reported alongside their 95% CIs. Statistical significance was set at $P<.05$.

Results

Participant Demographics by Admission Pathway

The demographic characteristics of participants across the 4 TCAS admission groups—academic, quota, test, and rural—demonstrated distinct patterns in age, sex, school region, and hometown region, as shown in Table 1. Most students were aged between 18 and 20 years, except in the quota group, where one-third were aged 21 - 25 years and another one-third 26 - 30 years. Accordingly, the mean age was highest in the quota group (23.7 years), compared with 19.4 - 19.5 years in the other groups. Male students predominated in the academic (15/23, 65.2%), test (134/261, 51.3%), and rural (17/29, 58.6%) groups, whereas the quota group was primarily female (5/6, 83.3%). By school region, Bangkok was most common in the test group (147/261, 56.3%), the central region predominated in the academic group (15/23, 65.2%), and the western region was most frequent in the rural group (22/29, 75.9%). The quota group showed a more even distribution, with 50.0% (3/6) from the central region and 16.7% (1/6) each from the west, south, and Bangkok. By hometown region, Bangkok accounted for the largest share in the test group (122/261, 46.7%); the academic group was distributed across Bangkok (9/23, 39.1%), central (5/23, 21.7%), and northeast (5/23, 21.7%); and the rural group primarily originated from the western region (23/29, 79.3%). The quota group consisted of one-third each from Bangkok and the central region, with smaller proportions from the west and south (1/6, 16.7% each).

Table . Participant demographics by admission pathway.

Factors	Academic group (N=23), n (%)	Quota group (N=6), n (%)	Test group (N=261), n (%)	Rural group (N=29), n (%)
Age (years)				
18 - 20	23 (100.0)	2 (33.4)	246 (94.2)	29 (100.0)
21 - 25	N/A ^a	2 (33.3)	15 (5.8)	N/A
26 - 30	N/A	2 (33.4)	N/A	N/A
Mean	19.4	23.7	19.5	19.5
Median	19.4	24.0	19.4	19.5
Range	18.7 - 20.0	19.4 - 29.7	18.0 - 22.7	18.8 - 20.0
Sex				
Male	15 (65.2)	1 (16.7)	134 (51.3)	17 (58.6)
Female	8 (34.8)	5 (83.3)	127 (48.7)	12 (41.4)
School region				
Bangkok	3 (13.0)	1 (16.7)	147 (56.3)	N/A
Central	15 (65.2)	3 (50.0)	90 (34.5)	7 (24.1)
North	1 (4.4)	N/A	2 (0.8)	N/A
Northeast	3 (13.0)	N/A	9 (3.5)	N/A
East	1 (4.4)	N/A	3 (1.1)	N/A
West	N/A	1 (16.7)	N/A	22 (75.9)
South	N/A	1 (16.7)	10 (3.8)	N/A
Hometown region				
Bangkok	9 (39.1)	2 (33.3)	122 (46.7)	N/A
Central	5 (21.7)	2 (33.3)	59 (22.6)	6 (20.7)
North	1 (4.4)	N/A	3 (1.2)	N/A
Northeast	5 (21.7)	N/A	24 (9.2)	N/A
East	2 (8.7)	N/A	11 (4.2)	N/A
West	N/A	1 (16.7)	8 (3.1)	23 (79.3)
South	1 (4.4)	1 (16.7)	34 (13.0)	N/A

^aN/A: not applicable.

Comparisons of SRL Strategies Among Academic, Quota, Test, and Rural Groups

Comparisons of SRL strategies among academic, quota, test, and rural groups are shown in [Table 2](#). In the aspect of teaching and learning, students in the academic group reported significantly higher levels of actively setting goals before studying than those in the quota group ($P=.044$; [Table 2](#)). The 95% CI for the academic group was narrow (4.07 - 4.63), indicating high precision. In contrast, the quota group showed

a much wider 95% CI (2.58 - 4.75), reflecting lower precision due to both the small sample size and the high variability in students' responses within this group ([Table 2](#)). Students in the academic group also reported significantly higher levels of feeling enthusiastic about learning and teaching activities than students in the rural group ($P=.042$; [Table 2](#)). The 95% CI for the academic group (4.18 - 4.69) was narrow, suggesting stable estimates, whereas the rural group's moderately wide 95% CI (3.65 - 4.28) indicates greater variability in enthusiasm levels among students in this pathway ([Table 2](#)).

Table . Comparisons of self-regulated learning strategies among academic, quota, test, and rural groups.

Questions	Academic group (N=23)	Quota group (N=6)	Test group (N=258)	Rural group (N=29)	
	Mean	Mean	Mean	Mean	
	(95% CI)	(95% CI)	(95% CI)	(95% CI)	
Teaching and learning					
1	I actively set goals for myself before studying. (P1 ^a)	4.35 (4.07 - 4.63)	3.67 (2.58-4.75) ^b	4.22 (4.13 - 4.31)	4.10 (3.79 - 4.41)
2	I am enthusiastic about learning and teaching activities. (P1 ^a)	4.43 (4.18 - 4.69)	3.83 (2.80 - 4.87)	4.12 (4.02 - 4.22)	3.97 (3.65-4.28) ^b
3	I am happy and satisfied when I learn something new. (P3 ^c)	4.39 (4.08 - 4.70)	4.50 (3.93 - 5.07)	4.41 (4.32 - 4.50)	4.34 (4.11 - 4.58)
4	The teaching and learning management of this curriculum helps me experience learning in an excellent manner. (P3 ^c)	4.39 (4.05 - 4.73)	4.67 (4.12 - 5.21)	4.38 (4.29 - 4.46)	4.31 (4.10 - 4.52)
5	Overall, I am satisfied with the learning experience gained from the current curriculum. (P3 ^c)	4.35 (4.14 - 4.56)	4.67 (4.12 - 5.21)	4.35 (4.27 - 4.43)	4.38 (4.12 - 4.64)
CLOs^d and PLOs^e					
6	I feel happy when I achieve the CLOs each time. (P3 ^c)	4.83 (4.66 - 4.99)	5.00 (5.00 - 5.00)	4.62 (4.55 - 4.70)	4.55 (4.36 - 4.74)
7	I like the opportunity to have unlimited attempts at achieving the CLOs. (P2 ^f)	4.74 (4.54 - 4.93)	4.83 (4.40 - 5.26)	4.85 (4.80 - 4.90)	4.76 (4.56 - 4.95)
8	I think that the knowledge and experience gained from studying in the first year can be applied in future medical studies. (P3 ^c)	4.61 (4.36 - 4.86)	5.00 (5.00 - 5.00)	4.67 (4.61 - 4.74)	4.72 (4.55 - 4.90)
9	I think that the knowledge and experience gained from studying will be applicable to the preclinical comprehensive examination and the National License Examination step 1. (P3 ^c)	4.74 (4.54 - 4.93)	4.83 (4.40 - 5.26)	4.64 (4.57 - 4.71)	4.62 (4.41 - 4.83)
10	The current examination method (once/semester) allows me to study happily. (P3 ^c)	3.17 (2.64 - 3.71)	3.33 (2.48 - 4.19)	3.40 (3.24 - 3.56)	3.24 (2.81 - 3.67)
Nongrading evaluation					

Questions	Academic group (N=23)	Quota group (N=6)	Test group (N=258)	Rural group (N=29)	
	Mean	Mean	Mean	Mean	
	(95% CI)	(95% CI)	(95% CI)	(95% CI)	
11	I am happy to learn through the nongrading evaluation ("O"/"S"/"U"). (P3 ^c)	4.35 (4.07 - 4.63)	5.00 (5.00-5.00) ^b	4.77 (4.70 - 4.84) ^g	4.86 (4.73 - 5.00) ^g
12	I agree with changing from the grading to the nongrading evaluation ("O"/"S"/"U") for all preclinical courses. (P3 ^c)	4.35 (4.07 - 4.63)	5.00 (5.00-5.00) ^b	4.79 (4.72 - 4.86) ^h	4.90 (4.78 - 5.01) ^h
Academic obstacles					
13	I feel like I'm wasting my time participating in the teaching and learning activities. (P2 ^f)	1.83 (1.47 - 2.19)	1.67 (0.81 - 2.52)	2.07 (1.91 - 2.22)	2.00 (1.55 - 2.45)
14	I have concerns or problems that prevent me from studying. (P2 ^f)	2.13 (1.60 - 2.66)	1.83 (1.04 - 2.62)	2.60 (2.45 - 2.76)	2.76 (2.38 - 3.13)
15	I don't feel like studying. (P1 ^a)	1.87 (1.43 - 2.31)	2.00 (1.06 - 2.94)	2.37 (2.22 - 2.52)	2.38 (1.98 - 2.78)
16	I have stress when reading books or following lessons during studies. (P2 ^f)	2.64 (2.15 - 3.12)	3.00 (1.67 - 4.33)	3.27 (3.14 - 3.40) ^g	3.52 (3.20 - 3.83) ^g
17	I have stress when reading books before exams. (P2 ^f)	3.70 (3.11 - 4.29) ⁱ	3.50 (1.78 - 5.22)	3.97 (3.85 - 4.10) ⁱ	4.38 (4.10 - 4.66)

^aP1 represents Motivation and Beliefs (motivation and beliefs (Forethought Phase).

^b $P<.05$ compared with the academic group.

^cP3 represents assessment and self-reaction (Self-Reflection Phase).

^dCLOs: course learning outcomes.

^ePLOs: program learning outcomes.

^fP² represents Monitoring and Strategy Use (Performance Phase).

^g $P<.01$ compared with the academic group.

^h $P<.001$ compared with the academic group.

ⁱ $P<.05$ compared with the rural group.

In the aspect of nongrading evaluation, students in the academic group reported significantly lower levels of agreement with learning via nongrading evaluation ("O"/"S"/"U") than the quota group ($P=.012$), test group ($P=.001$), and rural group ($P=.001$) (Table 2). The 95% CI for the academic group (4.07 - 4.63) was relatively narrow, indicating high precision and consistent responses, whereas the quota group showed a ceiling effect (5.00 - 5.00) suggesting no variability, and the 95% CIs for the test (4.70 - 4.84) and rural groups (4.73 - 5.00) were very narrow, reflecting precise estimates (Table 2). Similarly, students in the academic group reported significantly lower agreement with changing from grading to nongrading evaluation ("O"/"S"/"U") for all preclinical courses than the quota group

($P=.011$), test group ($P<.001$), and rural group ($P<.001$; Table 2). The 95% CI for the academic group (4.07 - 4.63) was narrow, suggesting stable estimates, whereas the quota group again showed a ceiling effect (5.00 - 5.00), and the test (4.72 - 4.86) and rural groups (4.78 - 5.01) had narrow 95% CIs, indicating high precision (Table 2).

In the aspect of academic obstacles, students in the academic group rated significantly lower levels of stress when reading books or following lessons during studies than students in the test ($P=.007$) and rural groups ($P=.003$; Table 2). The 95% CI for the academic group (2.15 - 3.12) was moderately wide, indicating greater variability in perceived stress, whereas the test (3.14 - 3.40) and rural groups (3.20 - 3.83) showed

narrower 95% CIs reflecting more consistent responses, and the quota group demonstrated a wide 95% CI (1.67 - 4.33), indicating low precision due to both high variability and a small sample size (Table 2).

Furthermore, students in the rural group rated significantly higher levels of stress when reading books before examinations than students in the academic ($P=.02$) and test groups ($P=.046$; Table 2). The 95% CI for the rural group (4.10 - 4.66) was relatively narrow, indicating high precision, while the 95% CI for the academic group (3.11 - 4.29) was moderately wide, suggesting moderate precision, and the 95% CI for the test group (3.85 - 4.10) was narrow, reflecting precise estimates; the quota group had an extremely wide 95% CI (1.78 - 5.22), indicating high variability and small sample size (Table 2).

Table . Students' preferences regarding nongrading evaluation.

	Academic group (N ^a =22), n (%)	Quota group (N=6), n (%)	Test group (N=243), n (%)	Rural group (N=28), n (%)
Choosing the evaluation				
“S” ^b /“U” ^b	7 (31.8)	1 (16.7)	90 (37.0)	13 (46.4)
“O” ^c /“S” ^c /“U” ^c	14 (63.6)	5 (83.3)	132 (54.3)	12 (42.9)
Both “S” ^c /“U” ^c and “O” ^c /“S” ^c /“U” ^c	1 (4.6)	N/A ^d	21 (8.7)	3 (10.7)
Topics				
“O” makes students feel stressed. (P3 ^e)	10 (45.5)	N/A	83 (34.2)	11 (39.3)
“O” leads to increased competition. (P2 ^f)	13 (59.1)	1 (16.7)	98 (40.3)	11 (39.3)
“O” makes students more focused on studying. (P2)	15 (68.2)	3 (50.0)	116 (47.7)	12 (42.9)
“O” leads to increased effort. (P2)	15 (68.2)	3 (50.0)	161 (66.3)	18 (64.3)
“O” leads to having more goals in life. (P1 ^g)	13 (59.5)	1 (16.7)	104 (42.8)	13 (46.4)

^a“S”: satisfactory.

^b“U”: unsatisfactory.

^c“O”: outstanding.

^dN/A: not applicable.

^eP3 represents assessment and self-reaction (Self-Reflection Phase).

^fP2 represents monitoring and strategy use (Performance Phase).

^gP1 represents motivation and beliefs (Forethought Phase).

Students' Preferences Regarding Nongrading Evaluation

Students' preferences regarding nongrading evaluation and the “O” grade in academic, quota, test, and rural groups are shown in Table 3. Students in the academic, quota, and test groups predominately selected “O”/“S”/“U” for 63.6% (14/22), 83.3% (5/6), and 54.3% (132/243), respectively, while students in the rural group mainly selected “S”/“U” at 46.4% (13/28; Table 3). Furthermore, students in the academic and quota groups reported that “O” mainly made them more focused on studying and led to increased effort with equal percentage (15/22, 68.2% and 3/6, 50%, respectively; Table 3). In addition, students in the test and rural groups reported that “O” mostly led to increased effort (161/243, 66.3% and 18/28, 64.3%, respectively; Table 3).

Students' Assessment Engagement Statistics in Performing Each CLO

Students' assessment engagement statistics in performing each CLO of students in academic, quota, test, and rural groups are shown as ranges of means and medians in **Table 4**. These

assessment engagement statistics include multiple aspects of assessment engagement, namely, total attempts, intentional and unintentional attempts, first-pass attempts, highest scoring attempts, passing instances, and additional attempts after passing for each CLO.

Table 4. Students' assessment engagement statistics in performing each course learning outcome.

Factors	Academic group (N=23)		Quota group (N=6)		Test group (N=258)		Rural group (N=29)	
	Mean (range)	Median (range)	Mean (range)	Median (range)	Mean (range)	Median (range)	Mean (range)	Median (range)
Number of total attempts of each CLO ^a .	2.0 - 9.1	1 - 7	3.5 - 29.0	2.5 - 19	2.6 - 15.3	2 - 8	2.4 - 12.3	2 - 10
Number of intentional attempts of each CLO.	1.9 - 6.3	1 - 5	2.0 - 6.3	2 - 5.5	2.0 - 7.5	2 - 6	2.3 - 9.8	2 - 9
Number of unintentional attempts of each CLO.	0 - 5.1	0	0 - 22.8	0 - 8.5	0 - 10.9	0	0 - 6.5	0
Instances of first-pass attempt of each CLO.	1.4 - 3.6	1 - 3	1.3 - 4.0	1 - 3	1.4 - 4.5	1 - 3	1.9 - 6.5	1 - 5
Instances of highest scoring attempt of each CLO.	1.7 - 5.7	1 - 4	1.7 - 6.0	1.5 - 5	1.9 - 6.7	2 - 5	2.2 - 9.5	2 - 9
Number of passings for each CLO.	1.5 - 3.0	1 - 2	1.5 - 4.0	1 - 3.5	1.4 - 3.3	1 - 2	1.4 - 2.9	1 - 2
Additional attempts after passing each CLO.	0.5 - 2.7	0 - 2	0.5 - 4.0	0 - 3	0.6 - 3.0	0 - 1	0.4 - 3.4	0 - 1

^aCLO: course learning outcome.

Comparisons of Assessment Engagement Statistics Among Students in Academic, Quota, Test, and Rural Groups

Comparisons of assessment engagement statistics among students in academic, quota, test, and rural groups are shown in **Figure 1**. For each CLO, the number of total attempts was significantly higher in the test group compared with the academic group in course 1 CLO 1 ($P=.044$) and course 2 CLO 2 ($P=.049$) and compared with the rural group in course 1 CLO 3 ($P=.03$), as well as significantly higher in the quota group compared with the academic ($P=.02$) and rural groups ($P=.041$) in course 3 CLO 2 (**Figure 1A**). The number of intentional attempts was significantly lower in the academic group in course 3 CLO 1 compared with the rural group ($P=.027$; **Figure 1B**). The number of unintentional attempts was significantly higher in the quota group compared with the rural group in course 1 CLO 3 ($P=.045$), course 3 CLO 2 ($P=.043$), and course 4 ($P=.048$), and compared with the academic group in course 3

CLO 2 ($P=.03$), as well as significantly higher in the test group compared with the rural group in course 1 CLO 3 ($P=.042$; **Figure 1C**). The instances of first-pass attempt were significantly lower in the academic group compared with the rural group in course 1 CLO 1 ($P=.04$), course 2 CLO 1 ($P=.008$) and 2 ($P=.004$), course 3 CLO 1 ($P=.03$), CLO 4 ($P=.044$), and CLO 5 ($P=.03$), and compared with the quota group in course 3 CLO 2 ($P=.02$), as well as significantly lower in the test group compared with the rural group in course 2 CLO 1 ($P=.02$) and CLO 2 ($P=.03$) and course 3 CLO 1 ($P=.04$) and 4 ($P=.005$) (**Figure 1D**). The instances of highest scoring attempts were significantly higher in the rural group compared with the academic group in course 2 CLO 2 ($P=.02$) and compared with the academic group ($P=.01$), quota group ($P=.03$), and test group ($P=.02$) in course 3 CLO 1 (**Figure 1E**). The number of passings in each CLO was significantly lower in the rural group compared with the test group in course 1 CLO 1 ($P=.03$), as well as significantly higher in the quota group compared with the academic group ($P=.004$), test group ($P<.001$), and rural group

($P<.001$) in course 2 CLO 4 (Figure 1F). The additional attempts after passing each CLO were significantly higher in the quota group compared with the academic group ($P=.01$), test group ($P=.004$), and rural group ($P=.001$) in course 2 CLO 4 (Figure 1G).

Across total attempts (Figure 1A) and unintentional attempts (Figure 1C), the test group consistently demonstrated narrow 95% CIs indicating high precision, the academic and rural groups showed 95% CIs of narrow to moderate width indicating high to moderate precision, and the quota group exhibited

extremely wide 95% CIs reflecting low precision due to substantial variability and a small sample size.

Across intentional attempts (Figure 1B), first-pass attempts (Figure 1D), highest scoring attempts (Figure 1E), number of passings (Figure 1F), and additional attempts after passing (Figure 1G), a similar precision pattern was observed, in which the test group consistently showed narrow 95% CIs indicating high precision, the academic and rural groups demonstrated moderate width 95% CIs reflecting moderate precision, and the quota group exhibited the widest 95% CIs indicating low precision due to high variability and its small sample size.

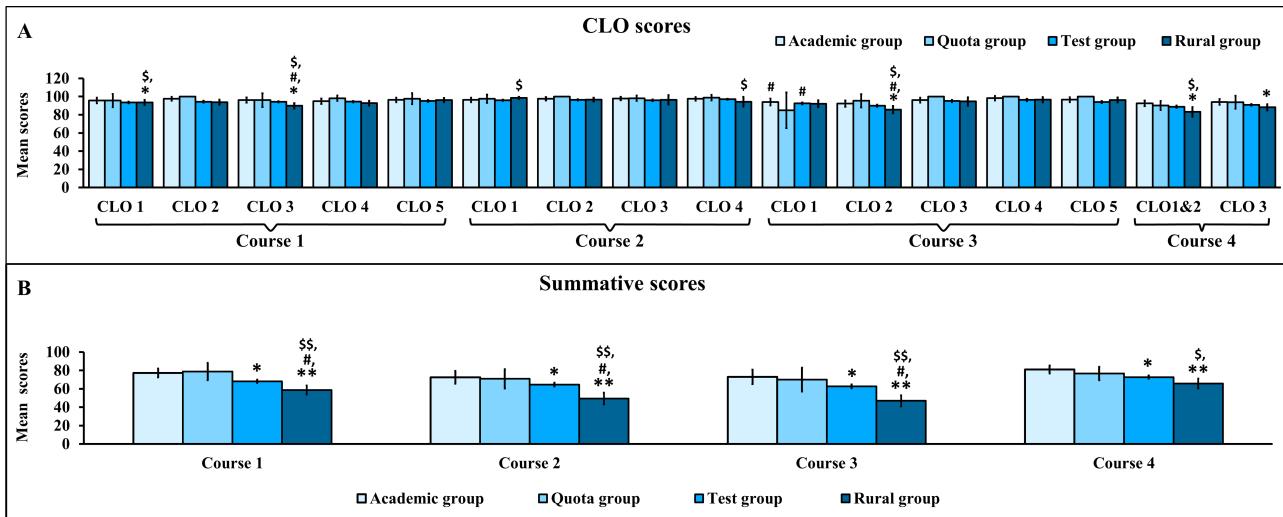
Figure 1. Comparisons of assessment engagement statistics among students in academic, quota, test, and rural groups. Data are presented as mean with 95% CIs. $*P<.05$ compared with students in the academic group; $^{\#}P<.05$, $^{##}P<.001$ compared with students in the quota group; and $^{\$}P<.05$ compared with students in the test group. (A) Number of total attempts of each CLO. (B) Number of intentional attempts of each CLO. (C) Number of unintentional attempts of each CLO. (D) The instance(s) of first-pass attempt of each CLO. (E) The instance(s) of highest scoring attempt of each CLO. (F) Number of passing(s) each CLO. (G) Additional attempt(s) after passing each CLO. CLO: course learning outcome.

Comparisons of CLO and Summative Scores Among Students in Academic, Quota, Test, and Rural Groups

Comparisons of CLO and summative scores among students in academic, quota, test, and rural groups are shown in **Figure 2**.

Some CLO scores (**Figure 2A**) and all summative scores (**Figure 2B**) in courses 1 - 4 were significantly lower in the rural group when compared with the academic, quota, and test groups

Figure 2. Comparisons of CLO and summative scores among students in academic, quota, test, and rural groups. Data are presented as mean with 95% CIs. * $P<.05$, ** $P<.001$ compared with students in the academic group; # $P<.05$ compared with students in the quota group; and \$ $P<.05$, \$\$ $P<.001$ compared with students in the test group. (A) CLO scores. (B) Summative scores. CLO: course learning outcome.



Correlations of Summative Scores With Students' Assessment Engagement Statistics and CLO Scores Within the Course

Correlations of summative scores with students' assessment engagement statistics and CLO scores within the course of students in academic, quota, test, and rural groups are shown in [Multimedia Appendix 1](#).

In the academic group, summative scores of each course were positively correlated with number of passings each CLO of courses 1 and 3 ($R=0.415 - 0.580$), additional attempts after passing each CLO of course 1 ($R=0.415 - 0.555$), and CLO scores of all courses ($R=0.432 - 0.707$) but were negatively correlated with number of total attempts ($R=-0.450$) and number of unintentional attempts ($R=-0.511$) of course 1 and instances of first-pass attempt of courses 1 and 2 ($R=(-0.437) - (-0.491)$) ($P<.05$ all; [Multimedia Appendix 1](#)).

In the quota group, summative scores of each course exhibited positive correlation with number of total attempts ($R=0.814$) and number of intentional attempts ($R=0.926$) of course 2 and CLO scores of course 3 ($R=0.844$) but had negative correlations with number of total attempts ($R=-0.964$) and number of unintentional attempts ($R=-0.845$) of course 3 ($P<.05$ all; [Multimedia Appendix 1](#)).

In the test group, summative scores of each course were significantly positively correlated with number of passings each CLO ($R=0.164 - 0.233$) and additional attempts after passing each CLO ($R=0.125 - 0.209$) of courses 1, 2, and 3; CLO scores of all courses ($R=0.129 - 0.253$); and number of intentional

($P<.05$ all). Across CLO scores (**Figure 2A**) and summative scores (**Figure 2B**), a similar precision pattern was observed, in which the test group consistently showed narrow 95% CIs indicating high precision, the academic and rural groups demonstrated moderate-width 95% CIs reflecting moderate precision, and the quota group exhibited the widest 95% CIs indicating low precision due to substantial variability and a small sample size.

attempts of course 3 ($R=0.177$) but were significantly negatively correlated with number of total attempts ($R=(-0.128) - (-0.233)$) and number of unintentional attempts ($R=(-0.148) - (-0.208)$) of courses 1 and 3; number of intentional attempts of course 1 ($R=-0.134$); instances of first-pass attempt of courses 1, 2, and 3 ($R=(-0.123) - (-0.344)$); and instances of highest scoring attempt of courses 1 and 2 ($R=(-0.153) - (-0.160)$) ($P<.05$ all; [Multimedia Appendix 1](#)).

In the rural group, summative scores of each course showed positive correlations with CLO scores of courses 1, 3, and 4 ($R=0.420 - 0.725$) and number of unintentional attempts ($R=0.393$) and number of passings each CLO ($R=0.390$) of course 3 but exhibited negative correlations with number of intentional attempts of courses 2 and 3 ($R=(-0.399) - (-0.577)$); instances of first-pass attempt of courses 2 and 3 ($R=(-0.388) - (-0.655)$); and instances of highest scoring attempt of courses 2 and 3 ($R=(-0.441) - (-0.614)$) ($P<.05$ all; [Multimedia Appendix 1](#)).

Notably, the scores for course 1 CLO 2, course 2 CLO 2, and course 3 CLO 3, 4, and 5 in the quota group did not correlate with the summative scores because all students in this group received a score of 100 for these CLOs. Across correlations between summative scores and assessment engagement statistics and CLO scores, a similar precision pattern was observed, in which the test group consistently showed narrow 95% CIs indicating high precision, the academic and rural groups demonstrated moderate-width 95% CIs reflecting moderate precision, and the quota group exhibited the widest 95% CIs indicating low precision due to substantial variability and a small sample size ([Multimedia Appendix 1](#)).

Multiple Linear Regression Analyses of Summative Scores

Multiple linear regression analyses of summative scores in academic, quota, test, and rural groups are shown in [Table 5](#).

Table . Multiple linear regression analyses of summative scores.

Model	R	R ²	P value	Coefficient	SE	t test (df)	P value	95% CI
Academic group								
Summative scores of course 1								
1	0.630	0.397	.001 ^a					
				Constant	−40.301	31.689	−1.272 (21)	.217
				CLO5 ^b scores	1.221	0.329	3.717 (21)	.001 ^a
2	0.722	0.521	<.001 ^c					
				Constant	−17.197	30.682	−0.560 (20)	.581
				CLO5 scores	0.912	0.330	2.769 (20)	.012 ^d
				Number of passings (CLO1)	2.349	1.033	2.273 (20)	.034 ^d
3	0.784	0.614	<.001 ^c					
				Constant	−1.143	29.210	−0.039 (19)	.969
				CLO5 scores	0.764	0.311	2.456 (19)	.024 ^d
				Number of passings (CLO1)	2.112	0.957	2.206 (19)	.040 ^d
				Number of unintentional attempts (CLO4)	−0.466	0.217	−2.148 (19)	.045 ^d
Summative scores of course 2								
1	0.491	0.241	.017 ^d					
				Constant	88.704	6.934	12.793 (21)	<.001 ^c
				Instances of first-pass attempt (CLO3)	−10.648	4.125	−2.582 (21)	.017 ^d
Summative scores of course 3								
1	0.707	0.500	<.001 ^c					
				Constant	−119.198	42.038	−2.835 (21)	.010 ^d
				CLO5 scores	1.991	0.435	4.580 (21)	<.001 ^c
2	0.778	0.606	<.001 ^c					
				Constant	−160.196	42.113	−3.804 (20)	.001 ^a
				CLO5 scores	2.285	0.415	5.504 (20)	<.001 ^c
				Number of passings (CLO4)	8.538	3.678	2.322 (20)	.031 ^d

Model	R	R ²	P value	Coefficient	SE	t test (df)	P value	95% CI
Summative scores of course 4								
1	0.520	0.271	.011 ^d	Constant	16.399	23.234	0.706 (21)	.488
				CLO1 and 2 scores	0.699	0.250	2.793 (21)	.011 ^d
								-31.920 to 64.717
								0.178 to 1.219
Quota group								
Summative scores of course 2								
1	0.926	0.857	.008 ^a	Constant	45.455	5.454	8.335 (4)	.001 ^a
				Number of intentional attempts (CLO1)	7.614	1.553	4.903 (4)	.008 ^a
								30.313 to 60.596
								3.302 to 11.925
Summative scores of course 3								
1	0.964	0.929	.002 ^a	Constant	86.429	2.719	31.782 (4)	<.001 ^c
				Number of total attempts (CLO4)	-4.107	0.567	-7.243 (4)	.002 ^d
								-5.681 to -2.533
Test group								
Summative scores of course 1								
1	0.344	0.118	<.001 ^c	Constant	73.796	1.251	58.999 (256)	<.001 ^c
				Instances of first-pass attempt (CLO3)	-2.653	0.453	-5.856 (256)	<.001 ^c
								-3.545 to -1.761
2	0.399	0.159	<.001 ^c	Constant	70.442	1.546	45.554 (254)	<.001 ^c
				Instances of first-pass attempt (CLO3)	-2.649	0.444	-5.963 (254)	<.001 ^c
				Number of passing (CLO5)	1.765	0.498	3.545 (254)	<.001 ^c
								0.784 to 2.745
3	0.433	0.187	<.001 ^c	Constant	71.500	1.565	45.689 (253)	<.001 ^c
				Instances of first-pass attempt (CLO3)	-2.558	0.439	-5.833 (253)	<.001 ^c
								-3.422 to -1.695

Model	R	R ²	P value	Coefficient	SE	t test (df)	P value	95% CI
				Number of passing (CLO5)	1.638	0.492	3.327 (253)	.001 ^a 0.668 to 2.608
				Number of unintentional attempts (CLO5)	-0.344	0.116	-2.953 (253)	.003 ^a -0.573 to -0.115
Summative scores of course 2								
1	0.260	0.068	<.001 ^c	Constant	71.040	1.797	39.530 (253)	<.001 ^c 67.501 to 74.580
				Instances of first-pass attempt (CLO3)	-3.164	0.739	-4.282 (253)	<.001 ^c -4.620 to -1.709
2	0.370	0.137	<.001 ^c	Constant	16.729	12.185	1.373 (252)	.171 -7.269 to 40.727
				Instances of first-pass attempt (CLO3)	-3.289	0.713	-4.614 (252)	<.001 ^c -4.693 to -1.885
				CLO3 scores	0.568	0.126	4.503 (252)	<.001 ^c 0.320 to 0.817
3	0.402	0.162	<.001 ^c	Constant	21.178	12.145	1.744 (251)	.082 -2.742 to 45.098
				Instances of first-pass attempt (CLO3)	-3.222	0.704	-4.574 (251)	<.001 ^c -4.610 to -1.835
				CLO3 scores	0.472	0.130	3.643 (251)	<.001 ^c 0.217 to 0.727
				Number of passing (CLO2)	2.260	0.833	2.714 (251)	.007 ^a 0.620 to 3.900
4	0.420	0.176	<.001 ^c	Constant	24.308	12.157	2.000 (250)	.047 ^d 0.366 to 48.251
				Instances of first-pass attempt (CLO3)	-2.881	0.719	-4.009 (250)	<.001 ^c -4.296 to -1.465
				CLO3 scores	0.455	0.129	3.525 (250)	.001 ^a 0.201 to 0.709
				Number of passing (CLO2)	3.663	1.065	3.441 (250)	.001 ^a 1.566 to 5.760
				Instances of highest scoring attempts (CLO2)	-1.625	0.777	-2.093 (250)	.037 ^d -3.155 to -0.096
Summative scores of course 3								

Model	R	R ²	P value	Coefficient	SE	t test (df)	P value	95% CI
1	0.211	0.045	<.001 ^c	Constant	59.682	1.385	43.099 (256)	<.001 ^c 56.955 to 62.409
				Number of passing (CLO1)	1.032	0.299	3.456 (256)	.001 ^a 0.444 to 1.621
2	0.310	0.096	<.001 ^c	Constant	63.643	1.701	37.406 (255)	<.001 ^c 60.292 to 66.994
				Number of passing (CLO1)	1.014	0.291	3.483 (255)	.001 ^a 0.441 to 1.587
				Instances of first-pass attempt (CLO3)	-2.142	0.560	-3.822 (255)	<.001 ^c -3.246 to -1.038
3	0.354	0.125	<.001 ^c	Constant	60.276	2.043	29.508 (254)	<.001 ^c 56.253 to 64.299
				Number of passing (CLO1)	0.735	0.303	2.429 (254)	.016 ^d 0.139 to 1.332
				Instances of first-pass attempt (CLO3)	-2.570	0.572	-4.493 (254)	<.001 ^c -3.697 to -1.444
				Number of passing (CLO4)	3.520	1.219	2.888 (254)	.004 ^a 1.120 to 5.921
4	0.405	0.164	<.001 ^c	Constant	61.009	2.012	30.317 (253)	<.001 ^c 57.046 to 64.972
				Number of passing (CLO1)	0.679	0.297	2.288 (253)	.023 ^d 0.095 to 1.264
				Instances of first-pass attempt (CLO3)	-2.201	0.571	-3.857 (253)	<.001 ^c -3.325 to -1.077
				Number of passing (CLO4)	5.381	1.312	4.102 (253)	<.001 ^c 2.797 to 7.964
				Number of total attempts (CLO4)	-1.469	0.429	-3.424 (253)	.001 ^a -2.313 to -0.624
Summative scores of course 4								
1	0.357	0.127	<.001 ^c	Constant	32.361	6.595	4.907 (259)	<.001 ^c 19.374 to 45.347
				CLO1 and 2 scores	0.454	0.074	6.150 (259)	<.001 ^c 0.308 to 0.599

Model	R	R ²	P value	Coefficient	SE	t test (df)	P value	95% CI
2	0.395	0.156	<.001 ^c					
				Constant	13.440	9.095	1.478 (258)	.141
				CLO1 and 2 scores	0.354	0.080	4.433 (258)	<.001 ^c
				CLO3 scores	0.305	0.103	2.973 (258)	.003 ^a
Rural group								
Summative scores of course 1								
1	0.482	0.232	.008 ^a					
				Constant	-20.619	27.819	-0.741 (27)	.465
				CLO2 scores	0.845	0.296	2.857 (27)	.008 ^a
Summative scores of course 2								
1	0.655	0.429	<.001 ^c					
				Constant	75.026	6.184	12.132 (27)	<.001 ^c
				Instances of first-pass attempt (CLO1)	-10.182	2.261	-4.504 (27)	<.001 ^c
Summative scores of course 3								
1	0.441	0.194	.017 ^d					
				Constant	59.697	5.668	10.532 (27)	<.001 ^c
				Instances of highest scoring attempt (CLO4)	-5.853	2.294	-2.551 (27)	.017 ^d
2	0.600	0.360	.003 ^a					
				Constant	57.041	5.250	10.865 (26)	<.001 ^c
				Instances of highest scoring attempt (CLO4)	-6.060	2.086	-2.905 (26)	.007 ^a
				Number of passing (CLO2)	1.325	0.511	2.592 (26)	.015 ^d
3	0.748	0.559	<.001 ^c					
				Constant	54.567	4.502	12.119 (26)	<.001 ^c
				Instances of highest scoring attempt (CLO4)	-6.331	1.767	-3.584 (25)	.001 ^a
				Number of passing (CLO2)	1.444	0.434	3.327 (25)	.003 ^a

Model	R	R ²	P value	Coefficient	SE	t test (df)	P value	95% CI
				Number of unintentional attempts (CLO1)	1.718	0.511	3.364 (25)	.002 ^a 0.666 to 2.769
Summative scores of course 4								
1	0.725	0.526	<.001 ^c	Constant	3.731	11.465	0.325 (27)	.747 -19.793 to 27.256
				CLO1 and 2 scores	0.745	0.136	5.476 (27)	<.001 ^c 0.466 to 1.025

^a $P<.01$.

^bCLO: course learning outcome.

^c $P<.001$.

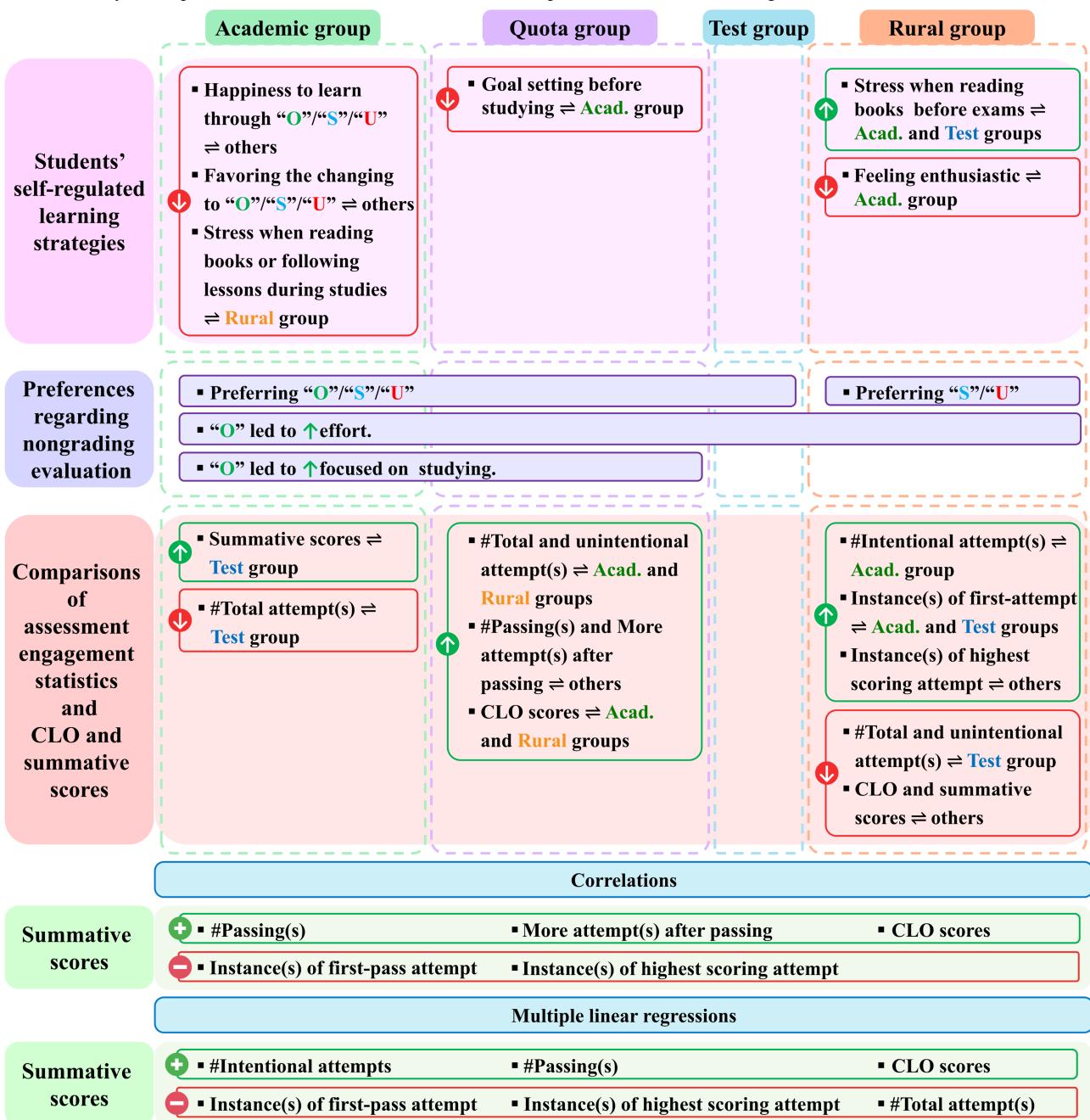
^d $P<.05$.

In the academic group, with summative scores as the dependent variable, positive influences included CLO scores and number of passings for each CLO, while a negative influence comprised instances of first-pass attempts (Table 5). In the quota group, a positive contribution was the number of intentional attempts, while a negative contribution was the number of total attempts (Table 5). In the test group, positive factors comprised number of passings for each CLO and CLO scores, while negative factors included instances of first-pass attempts, instances of highest scoring attempts of CLO, and number of total attempts (Table 5). In the rural group, by setting summative scores as the dependent variable, positive influences included CLO scores, number of passings for each CLO, and number of unintentional

attempts, while negative contributions comprised instances of first-pass attempts and instances of highest scoring attempts of CLO (Table 5).

Across all predictors and constant terms in the multiple linear regression models predicting summative scores, the test group generally showed narrow 95% CIs for predictors and moderately wide 95% CIs for constants, indicating overall moderate to high precision; the quota group exhibited moderate 95% CIs, reflecting moderate precision; and the academic and rural groups showed moderate to wide 95% CIs, especially wide for constants, indicating moderate to low precision (Table 5). A summary of results is shown in Figure 3.

Figure 3. The summary of results. Acad: academic group, CLO: course learning outcome; “O”: Outstanding, others: other groups, “S”: Satisfactory, “U”: Unsatisfactory, : compared with; : increased; : decreased; : positive correlations; : negative correlations.



Discussion

Principal Findings

This study aimed to compare SRL strategies, assessment engagement behaviors, preferences for nongrading evaluation, and academic performance across 4 admission pathways (academic, quota, test, and rural); examine correlations among summative scores, CLO scores, and engagement metrics; and identify predictors of summative scores using multiple linear regression. In line with these objectives, the main findings indicate that students in the academic group demonstrated stronger SRL strategies, fewer but more effective assessment engagement behaviors, and higher CLO and summative scores; and the rural group exhibited lower SRL, greater academic

obstacles, higher numbers of assessment attempts, and the lowest academic performance, while the quota and test groups showed intermediate patterns. Most academic, quota, and test students preferred “O”/“S”/“U,” whereas rural students favored “S”/“U.” Summative scores exhibited positive correlations with CLO scores and number of passings, whereas they showed negative correlations with first-pass attempts. In the multiple linear regression analyses, summative scores were contributed positively by number of passings and CLO scores, whereas they were predicted negatively by instances of first-pass attempts and highest scoring attempts.

When comparing SRL strategies across admission pathways, students in the academic group showed stronger goal setting and greater enthusiasm for learning activities than other groups.

They also reported less stress during routine study, whereas rural students experienced more stress before examinations than their peers.

Basically, the academic group, composed of science Olympiads students who participate in science competitions, generally demonstrates a high interest in science and learning, comes from competitive backgrounds, and possesses higher levels of general cognitive abilities [29]. Furthermore, they typically have a self-concept of ability [30], self-efficacy [29], short-term and long-term goals [29], perseverance [31], and self-discipline [31], all indicative of SRL [29,32]. In addition, they usually have support from peers, family, and teachers [29,33]. Therefore, it is not surprising that they exhibited higher levels of active goal setting and enthusiasm, showed less stress, and were more likely to receive higher scores. This finding aligns with recent research showing that medical graduates from more educationally privileged backgrounds (ie, those who attended fee-paying schools) were significantly more likely to succeed in postgraduate medical examinations [34]. In contrast, the rural group, typically from less competitive backgrounds, rated lower levels of enthusiasm but higher levels of stress when learning alongside students from high academic backgrounds [35]. These students may face unique challenges, such as limited access to educational resources and support systems [35-37].

Interestingly, students in the academic group were less satisfied with learning through nongrading evaluation and less supportive of adopting it for all preclinical courses than other groups. Most students in the academic, quota, and test pathways favored the inclusion of the “O” grade because they felt that it encouraged focus and effort, whereas rural students showed a slight preference for the simpler “S”/“U” system. These results reflect that the academic group might have familiarity with competitive academic environments [29,32] and a preference for evaluation systems that differentiate performance and offer visible recognition of achievement. The “O” grade appears to serve as an external motivating factor by reinforcing effort and focus through a tangible reward for students in most groups [10]. Students in the rural group showed a slight preference for the simpler “S”/“U” system, which was possibly due to a desire to reduce pressure or stress associated with competitive assessments, or differing motivational orientations shaped by prior educational experiences [35-37].

For assessment engagement and academic performance, students in the academic group achieved higher CLO and summative scores and completed fewer attempts overall, including early passes and highest scoring attempts, compared with other groups. These results suggest that the academic group, with a science Olympiads background, had strong SRL abilities that contributed to efficient learning and better outcomes.

In contrast, rural students showed lower CLO and summative scores and needed more attempts to pass each CLO, indicating greater challenges in regulating their learning, which may stem from factors that hinder their ability to effectively manage learning [36]. These results are consistent with a previous study in Jordan, which demonstrated that students from remote areas performed worse than those admitted through the competitive pathway or the academically excellent track [1].

In the correlation analysis of summative scores, the number of passings for each CLO, additional attempts after passing, and CLO scores consistently showed positive associations, whereas instances of first-pass attempts and highest scoring attempts showed negative associations. Similarly, in the multiple linear regression analyses, the number of intentional attempts, number of passings, and CLO scores contributed positively to summative scores, whereas total attempts, first-pass attempts, and highest scoring attempts contributed negatively. Taken together, these findings indicate that formative mastery—rather than the frequency of attempts—is the strongest predictor of summative success. This interpretation aligns with prior research showing that repeated unsuccessful attempts often reflect ineffective regulation or shallow processing strategies, which in turn predict lower achievement [38,39].

These empirical patterns align well with theoretical perspectives from Zimmerman’s cyclical SRL model, in which the Forethought Phase—comprising goal setting and strategic planning—shapes subsequent performance behaviors [7]. Students requiring multiple first-pass attempts or repeated highest scoring attempts may reflect weaker planning or monitoring processes—consistent with evidence that inadequate forethought predicts inefficient learning behaviors and lower achievement [7,39]. Conversely, the positive contributions of CLO scores and purposeful engagement align with findings that according to Zimmerman’s SRL model, self-monitoring enables learners to detect progress and develop the efficacy needed to perform at a high level of skill, while adaptive reactions increase the effectiveness of students’ learning strategies, and self-reflection feeds forward into subsequent forethought processes [7]. Similar relationships between SRL components and academic performance have been reported in medical education, where proactive goal setting and metacognitive regulation strongly predict examination outcomes [40,41].

These interpretations are based on associations and should not be construed as evidence that modifying a single engagement metric would necessarily lead to improved summative scores without concurrent changes in broader learning strategies and support. Nevertheless, addressing these challenges requires equitable access to resources, supportive environments, and targeted instruction [42]. Our findings emphasize how differences in learning behavior and assessment engagement statistics are strongly associated with academic performance. The academic group’s pattern of fewer but more effective attempts, coupled with higher CLO and summative scores, likely reflects their stronger SRL skills. While performance-based distinctions such as the “O” grade may enhance motivation and engagement for some students, they may also exacerbate stress or feelings of disadvantage for others. To promote equity and acceptance of nongrading systems, educational strategies should include clear communication of purpose, tailored feedback, and early orientation—particularly for students less familiar with performance-based academic cultures. In contrast, the rural group required more attempts and demonstrated lower performance, suggesting a need for earlier support and closer monitoring. These insights underscore the importance of using assessment engagement metrics as early indicators for identifying at-risk students. Interventions such as targeted

mentoring, structured remediation, and proactive counseling, along with curriculum designs that promote SRL, may help foster equitable academic success across diverse student populations.

Limitations

This study has several limitations. First, there were disparities in sample sizes across admission groups, particularly in the quota group, which included only 6 students. This small number may have limited the statistical power of comparisons and correlation analyses within and between groups. In addition, a few students in the quota group exhibited an unusually high number of CLO attempts, contributing to outliers that may have affected the robustness and generalizability of the findings, as reflected by the very wide 95% CIs observed in this group, indicating low precision due to high variability and the small sample size. Second, the number of allowed attempts for CLO assessments varied across courses; notably, only 1 attempt was permitted for course 4, which limited the ability to comprehensively analyze assessment engagement behaviors across all courses. Third, although descriptive demographic variables were added, including age, sex, school region, and hometown region, more detailed socioeconomic and family background information (eg, parental education, first-generation college or medical school status, and household income) was not comprehensively collected in the SiCMs. The absence of these variables limits interpretation of whether contextual factors may have contributed to the observed group differences, and future studies should incorporate such measures.

Conclusions

Admission pathways play a crucial role in shaping students' learning strategies, assessment engagement behaviors, and

academic performance. These findings highlight the need for medical schools to adopt equity-oriented curricular approaches that not only diversify selection processes but also actively support students with varying learning profiles throughout their training. Integrating assessment-engagement metrics with SRL data provides a scalable, data-driven framework for early identification of at-risk students and enables timely, targeted interventions—such as personalized mentoring, structured remediation, and SRL-focused instructional support. Moreover, understanding group-specific preferences toward nongrading evaluation underscores the importance of designing assessment reforms that balance academic rigor with psychological safety, ensuring fairness and acceptance across diverse student cohorts, while also recognizing that SRL strategies and engagement behaviors interact with these preferences in shaping learning outcomes. Prior research often examined these factors separately [1-3], so integrating them conceptually strengthens the evidence for equity-focused curriculum design and early support strategies. In this context, our study differs from existing research by concurrently examining SRL strategies, assessment engagement behaviors, preferences regarding nongrading evaluation, and academic performance across 4 distinct admission pathways within a single preclinical cohort. This integrated analytic approach—linking engagement metrics with SRL data and performance outcomes—provides new insights into how student characteristics and learning behaviors interact within diverse admission systems. These findings generate novel, actionable knowledge for designing equity-driven educational policies and leveraging learning analytics to inform curriculum development in medical education.

Acknowledgments

For offering invaluable guidance and assistance that greatly aided in the successful completion of this research, the authors would like to express their thanks to Associate Professor Rungnirand Praditsuwan, Former Deputy Dean for Undergraduate Education at the Faculty of Medicine Siriraj Hospital, Mahidol University. For preparing the tables, the authors would like to thank Nutradee Makphon. For manuscript formatting and data checking, the authors thank Inpreeya Inpaen for her assistance.

Funding

This study was financially supported by the Medical Council of Thailand (Police General Dr Jongjate Aojanepong Foundation) and the Faculty of Medicine Siriraj Hospital Medical Education Research Fund (IO: R016661001).

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

CS and IK made substantial contributions to the conception and design. CS, IK, NJ, and NW were responsible for data acquisition. IK, CS, and PM performed the analysis and interpretation of data. IK, CS, and PM drafted the paper or revised it critically for important intellectual content. Finally, all authors approved the final version for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Correlations of summative scores with students' assessment engagement statistics and course learning outcome scores within the course. Blue shades represent significant positive correlations, red shades represent significant negative correlations, and gray shades represent correlation not computable due to zero variance.

[[DOCX File, 51 KB - mededu_v12i1e68636_app1.docx](#)]

Checklist 1

STROBE checklist.

[[PDF File, 92 KB - mededu_v12i1e68636_app2.pdf](#)]

Checklist 2

GRAMMS checklist.

[[PDF File, 10 KB - mededu_v12i1e68636_app3.pdf](#)]

References

1. Tamimi A, Hassuneh M, Tamimi I, et al. Admission criteria and academic performance in medical school. *BMC Med Educ* 2023 Apr 21;23(1):273. [doi: [10.1186/s12909-023-04251-y](https://doi.org/10.1186/s12909-023-04251-y)] [Medline: [37085824](#)]
2. Curtis E, Wikaire E, Jiang Y, et al. Examining the predictors of academic outcomes for indigenous Māori, Pacific and rural students admitted into medicine via two equity pathways: a retrospective observational study at the University of Auckland, Aotearoa New Zealand. *BMJ Open* 2017 Aug;7(8):e017276. [doi: [10.1136/bmjopen-2017-017276](https://doi.org/10.1136/bmjopen-2017-017276)]
3. Aston-Mourney K, McLeod J, Rivera LR, McNeill BA, Baldi DL. Prior degree and academic performance in medical school: evidence for prioritising health students and moving away from a bio-medical science-focused entry stream. *BMC Med Educ* 2022 Oct 4;22(1):700. [doi: [10.1186/s12909-022-03768-y](https://doi.org/10.1186/s12909-022-03768-y)] [Medline: [36195862](#)]
4. TCAS64 round 3 application guide. version 1. Council of University Presidents of Thailand. 2021 May 6. URL: <https://www.admissionpremium.com/uploads/contents/2021051090735.pdf> [accessed 2025-08-25]
5. Chanmaneewong S, Yasri P. Challenges faced by high school students in Thailand when preparing for medical school admissions: a comparative study of regular Thai programmes and international schools. *Worldte* 2023;2(1):1-9. [doi: [10.33422/worldte.v2i1.51](https://doi.org/10.33422/worldte.v2i1.51)]
6. "Study anywhere to become a doctor, but not all are the same": Siriraj revises medical education and transforms the medical curriculum to meet learners' needs for the 2021 academic year, emphasizing the development of competent and compassionate physicians. Faculty of Medicine Siriraj Hospital, Mahidol University. 2020 Oct 22. URL: https://www.si.mahidol.ac.th/th/hotnewsdetail.asp?hn_id=2584 [accessed 2025-08-25]
7. Zimmerman BJ. Becoming a self-regulated learner: an overview. *Theory Pract* 2002 May;41(2):64-70. [doi: [10.1207/s15430421tip4102_2](https://doi.org/10.1207/s15430421tip4102_2)]
8. Villatoro Moral S, de-Benito Crossetti B. Self-regulation of learning and the co-design of personalized learning pathways in higher education: a theoretical model approach. *J Interactive Media Educ* 2022 Sep 12;2022(1). [doi: [10.5334/jime.749](https://doi.org/10.5334/jime.749)]
9. Ali M, Asim H, Edhi AI, et al. Does academic assessment system type affect levels of academic stress in medical students? A cross-sectional study from Pakistan. *Med Educ Online* 2015;20:27706. [doi: [10.3402/meo.v20.27706](https://doi.org/10.3402/meo.v20.27706)] [Medline: [26112353](#)]
10. White CB, Fantone JC. Pass-fail grading: laying the foundation for self-regulated learning. *Adv Health Sci Educ Theory Pract* 2010 Oct;15(4):469-477. [doi: [10.1007/s10459-009-9211-1](https://doi.org/10.1007/s10459-009-9211-1)] [Medline: [20012686](#)]
11. Wilkinson TJ, Wells JE, Bushnell JA. What is the educational impact of standards-based assessment in a medical degree? *Med Educ* 2007 Jun;41(6):565-572. [doi: [10.1111/j.1365-2923.2007.02766.x](https://doi.org/10.1111/j.1365-2923.2007.02766.x)] [Medline: [17518836](#)]
12. Wilkinson T. Pass/fail grading: not everything that counts can be counted. *Med Educ* 2011 Sep;45(9):860-862. [doi: [10.1111/j.1365-2923.2011.04018.x](https://doi.org/10.1111/j.1365-2923.2011.04018.x)] [Medline: [21848711](#)]
13. Hill MR, Goicochea S, Merlo LJ. In their own words: stressors facing medical students in the millennial generation. *Med Educ Online* 2018 Dec;23(1):1530558. [doi: [10.1080/10872981.2018.1530558](https://doi.org/10.1080/10872981.2018.1530558)] [Medline: [30286698](#)]
14. Moir F, Yielder J, Sanson J, Chen Y. Depression in medical students: current insights. *Adv Med Educ Pract* 2018;9:323-333. [doi: [10.2147/AMEP.S137384](https://doi.org/10.2147/AMEP.S137384)] [Medline: [29765261](#)]
15. Iyer AA, Hayes C, Chang BS, et al. Should medical school grading be tiered or pass/fail? A scoping review of conceptual arguments and empirical data. *Acad Med* 2025 Aug 1;100(8):975-985. [doi: [10.1097/ACM.00000000000006085](https://doi.org/10.1097/ACM.00000000000006085)] [Medline: [40359252](#)]
16. Sosibo L. The views of academics on the use of student feedback for curriculum improvement. *J Educ* 2010 Dec;49:113-132 [FREE Full text]
17. Williams A. Delivering effective student feedback in higher education: an evaluation of the challenges and best practice. *Int J Res Educ Sci* 2024 May;10(2):473-501. [doi: [10.46328/ijres.3404](https://doi.org/10.46328/ijres.3404)]
18. Bendermacher GWG, De Grave WS, Wolfhagen IHAP, Dolmans DHJM, Oude Egbrink MGA. Shaping a Culture for Continuous Quality Improvement in Undergraduate Medical Education. *Acad Med* 2020 Dec;95(12):1913-1920. [doi: [10.1097/ACM.00000000000003406](https://doi.org/10.1097/ACM.00000000000003406)] [Medline: [32287081](#)]

19. Broyles I, Savidge M, Schwalenberg-Leip E, Thompson K, Lee R, Sprafka S. Stages of concern during curriculum change. *J Int Assoc Med Sci Educ* 2007;17(1):14-26 [FREE Full text]
20. Li IW, Jackson D. Influence of entry pathway and equity group status on retention and the student experience in higher education. *High Educ* 2024 May;87(5):1411-1431. [doi: [10.1007/s10734-023-01070-4](https://doi.org/10.1007/s10734-023-01070-4)]
21. Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. *Med Educ* 2004 Mar;38(3):314-326. [doi: [10.1046/j.1365-2923.2004.01776.x](https://doi.org/10.1046/j.1365-2923.2004.01776.x)] [Medline: [14996341](https://pubmed.ncbi.nlm.nih.gov/14996341/)]
22. Nithiapinyasakul A, Arora R, Chamnan P. Impact of a 20-year collaborative approach to increasing the production of rural doctors in Thailand. *Int J Med Educ* 2016 Dec 22;7(414-6):414-416. [doi: [10.5116/ijme.582f.4d3b](https://doi.org/10.5116/ijme.582f.4d3b)] [Medline: [28008137](https://pubmed.ncbi.nlm.nih.gov/28008137/)]
23. Martín Cabrera E, García LA, Torbay Betancor Á. Factorial structure and reliability of a questionnaire on learning strategies in university students: CEA-U. *An Psicol* 2007 Jun;23(1):1-6 [FREE Full text]
24. Vallerand RJ, Pelletier LG, Blais MR, Briere NM, Senecal C, Vallières EF. The Academic Motivation Scale: a measure of intrinsic, extrinsic, and amotivation in education. *Educ Psychol Meas* 1992 Dec;52(4):1003-1017. [doi: [10.1177/0013164492052004025](https://doi.org/10.1177/0013164492052004025)]
25. Schinske J, Tanner K. Teaching more by grading less (or differently). *CBE Life Sci Educ* 2014;13(2):159-166. [doi: [10.1187/cbe.cbe-14-03-0054](https://doi.org/10.1187/cbe.cbe-14-03-0054)] [Medline: [26086649](https://pubmed.ncbi.nlm.nih.gov/26086649/)]
26. Spring L, Robillard D, Gehlbach L, Simas TAM. Impact of pass/fail grading on medical students' well-being and academic outcomes. *Med Educ* 2011 Sep;45(9):867-877. [doi: [10.1111/j.1365-2923.2011.03989.x](https://doi.org/10.1111/j.1365-2923.2011.03989.x)] [Medline: [21848714](https://pubmed.ncbi.nlm.nih.gov/21848714/)]
27. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Ann Intern Med* 2007 Oct 16;147(8):W163-W194. [doi: [10.7326/0003-4819-147-8-20071016-00010-w1](https://doi.org/10.7326/0003-4819-147-8-20071016-00010-w1)] [Medline: [17938389](https://pubmed.ncbi.nlm.nih.gov/17938389/)]
28. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. *J Health Serv Res Policy* 2008 Apr;13(2):92-98. [doi: [10.1258/jhsrp.2007.007074](https://doi.org/10.1258/jhsrp.2007.007074)] [Medline: [18416914](https://pubmed.ncbi.nlm.nih.gov/18416914/)]
29. Tschisgale PL, Steegh A, Petersen S, Kubsch M, Wulff P, Neumann K. Are science competitions meeting their intentions? a case study on affective and cognitive predictors of success in the Physics Olympiad. *Discip Interdiscip Sci Educ Res* 2024 Mar;6(1):10. [doi: [10.1186/s43031-024-00102-y](https://doi.org/10.1186/s43031-024-00102-y)]
30. Campbell JR. Early identification of mathematics talent has long-term positive consequences for career contributions. *Int J Educ Res* 1996;25(6):497-522. [doi: [10.1016/S0883-0355\(97\)86728-6](https://doi.org/10.1016/S0883-0355(97)86728-6)]
31. Verna MA, Feng AX. American Chemistry Olympians achieve the highest level of equity. *J Educ Res* 2002;12(1):101-105 [FREE Full text]
32. Cho KK, Marjadi B, Langendyk V, Hu W. The self-regulated learning of medical students in the clinical environment—a scoping review. *BMC Med Educ* 2017 Jul 10;17(1):112. [doi: [10.1186/s12909-017-0956-6](https://doi.org/10.1186/s12909-017-0956-6)] [Medline: [28693468](https://pubmed.ncbi.nlm.nih.gov/28693468/)]
33. Campbell JR, Feng AX. Comparing adult productivity of American Mathematics, Chemistry, and Physics Olympians With Terman's Longitudinal Study. *Roeper Rev* 2010 Dec 30;33(1):18-25. [doi: [10.1080/02783193.2011.530203](https://doi.org/10.1080/02783193.2011.530203)]
34. Ellis R, Knapton A, Cannon J, Lee AJ, Cleland J. A multivariate analysis examining the relationship between sociodemographic differences and UK graduates' performance on postgraduate medical exams. *Med Teach* 2025 Dec;47(12):2006-2020. [doi: [10.1080/0142159X.2025.2513426](https://doi.org/10.1080/0142159X.2025.2513426)] [Medline: [40512226](https://pubmed.ncbi.nlm.nih.gov/40512226/)]
35. Wilkinson TJ, McKenzie JM, Ali AN, Rudland J, Carter FA, Bell CJ. Identifying medical students at risk of underperformance from significant stressors. *BMC Med Educ* 2016 Feb 2;16(1):43. [doi: [10.1186/s12909-016-0565-9](https://doi.org/10.1186/s12909-016-0565-9)] [Medline: [26837428](https://pubmed.ncbi.nlm.nih.gov/26837428/)]
36. Rodriguez F, Rivas MJ, Matsumura LH, Warschauer M, Sato BK. How do students study in STEM courses? Findings from a light-touch intervention and its relevance for underrepresented students. *PLoS One* 2018;13(7):e0200767. [doi: [10.1371/journal.pone.0200767](https://doi.org/10.1371/journal.pone.0200767)] [Medline: [30063744](https://pubmed.ncbi.nlm.nih.gov/30063744/)]
37. Boateng BA, Thomas B. How can we ease the social isolation of underrepresented minority students? *Acad Med* 2011 Oct;86(10):1190. [doi: [10.1097/ACM.0b013e31822be60a](https://doi.org/10.1097/ACM.0b013e31822be60a)] [Medline: [21955702](https://pubmed.ncbi.nlm.nih.gov/21955702/)]
38. Alvi E, Iqbal Z, Masood F, et al. A qualitative account of the nature and use of self-regulated learning (SRL) strategies employed by university students. *Aust J Teach Educ* 2016;41(8):40-59. [doi: [10.14221/ajte.2016v41n8.3](https://doi.org/10.14221/ajte.2016v41n8.3)]
39. Rovers SFE, Stalmeijer RE, van Merriënboer JJJ, Savelberg HHCM, de Bruin ABH. How and why do students use learning strategies? A mixed methods study on learning strategies and desirable difficulties with effective strategy users. *Front Psychol* 2018;9:2501. [doi: [10.3389/fpsyg.2018.02501](https://doi.org/10.3389/fpsyg.2018.02501)] [Medline: [30618932](https://pubmed.ncbi.nlm.nih.gov/30618932/)]
40. Bursali Boz N, Oz H. The role of goal setting in metacognitive awareness as a self-regulatory behavior in foreign language learning. *Int Online J Educ Teach* 2018 Jul;5:662-671 [FREE Full text]
41. Morisano D, Locke EA. Goal setting and academic achievement. In: Hattie JAC, Anderman EM, editors. *International Guide to Student Achievement*: NY: Routledge; 2013:45-48 URL: <https://psycnet.apa.org/record/2013-13643-014> [accessed 2025-08-25]
42. O'Neal L, Perkins A. Rural exclusion from science and academia. *Trends Microbiol* 2021 Nov;29(11):953-956. [doi: [10.1016/j.tim.2021.06.012](https://doi.org/10.1016/j.tim.2021.06.012)] [Medline: [34274196](https://pubmed.ncbi.nlm.nih.gov/34274196/)]

Abbreviations

AES: assessment engagement statistics

CLO: course learning outcome

GPA: grade point average

GRAMMS: Good Reporting of A Mixed Methods Study

MCAR: Missing Completely at Random

MMI: multiple mini-interview

O: outstanding

S: Satisfactory

SiCMs: Siriraj Campus Management System

SRL: self-regulated learning

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

TCAS: Thai University Central Admission System

TPAT: Thai Professional Aptitude Test

U: Unsatisfactory

Edited by S Brini; submitted 12.Nov.2024; peer-reviewed by C Wallwey, L Harrison; revised version received 15.Dec.2025; accepted 21.Dec.2025; published 02.Feb.2026.

Please cite as:

Keadkraichaiwat I, Sitticharoon C, Maprapho P, Jangboon N, Wannarat N

Influence of Admission Pathways on Learning Strategies, Assessment Engagement, and Academic Performance Among First-Year Medical Students: Mixed Methods Retrospective Observational and Cross-Sectional Survey Study

JMIR Med Educ 2026;12:e68636

URL: <https://mededu.jmir.org/2026/1/e68636>

doi:10.2196/68636

© Issarawan Keadkraichaiwat, Chantacha Sitticharoon, Punyapat Maprapho, Nisa Jangboon, Nadda Wannarat. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 2.Feb.2026. 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.

The Effect of a Traditional Chinese Medicine Course on Western Medicine Students' Attitudes Toward Traditional Chinese Medicine: Self-Controlled Pre-Post Questionnaire Study

Haoyu He^{1,2,3}, MD; Hanjin Cui⁴, MD; Li Guo⁵; Yanhui Liao⁶, MD

¹Department of Psychology, College of Education, Hunan First Normal University, Changsha, Hunan, China

²Research Center for Innovative Development of Teacher Education in the New Era, Changsha, China

³Hunan Key Laboratory for Children's Psychological Development and Brain & Cognition Science, Changsha, China

⁴Institute of Integrative Medicine, Xiangya Hospital, Central South University, Changsha, Hunan, China

⁵Department of Educational Administration, The Second People's Hospital of Hunan Province (Brain Hospital of Hunan Province), Changsha, China

⁶Department of Psychiatry, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, 3 East Qingchun Road, Hangzhou, Zhejiang, China

Corresponding Author:

Yanhui Liao, MD

Department of Psychiatry, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, 3 East Qingchun Road, Hangzhou, Zhejiang, China

Abstract

Background: Traditional Chinese medicine (TCM) has been widely used to treat various diseases in China for thousands of years and has shown satisfactory effectiveness. However, many surveys found that TCM receives little recognition from Western medicine (WM) physicians and students. At present, TCM is offered as a compulsory course for WM students in WM schools.

Objective: This study aimed to investigate whether TCM courses can affect the WM students' attitude toward TCM.

Methods: WM students from Xiangya Medical School were invited to complete a web-based questionnaire before and immediately after a TCM course. Their attitude toward TCM and treatment preferences for different kinds of diseases were tested. The Attitude Scale of TCM (ASTCM) was used. The main part of the ASTCM was designed to measure the attitude of medical students toward TCM. It consisted of 18 items, divided into cognitive dimension (5 terms), emotional dimension (8 terms), and behavioral tendency factor (5 terms).

Results: Finally, the results of 118 five-year program (FYP) and 36 eight-year program (EYP) students were included. For FYP students, there was a significant increase in the total mean score (66.42, SD 7.66 vs 71.43, SD 7.38; $P<.001$) of ASTCM after the TCM course. Significant increases in mean scores of the 3 factors of attitude (cognition: 21.64, SD 2.08 vs 22.90, SD 1.94; affection: 25.21, SD 4.39 vs 27.96, SD 4.4; and behavioral tendency: 19.577, SD 3.02 vs 20.58, SD 2.76; $P<.001$) were also observed. Except for the score of behavioral tendency (17.50, SD 3.54 vs 18.78, SD 3.22; $P=.02$), a significant increase was not detected in total score, cognition, and affection in EYP students (total score: mean 60.36, SD 10.53 vs mean 62.92, SD 10.05; cognition: mean 20.50, SD 2.73 vs mean 20.69, SD 2.73; and affection: mean 22.36, SD 6.32 vs mean 23.44, SD 5.84; all $P>.05$). The treatment preference of FYP students in acute ($P=.02$), chronic ($P=.003$), and physical diseases ($P=.02$) showed remarkable change. A major change was also detected in internal diseases ($P=.02$), surgical diseases (perioperative period; $P=.01$), and mental illnesses ($P=.02$) in EYP students. This change mainly appeared as a decline in WM preference and an increase in TCM and WM preference.

Conclusions: The study showed that earlier exposure to the TCM course increased the positive attitude toward TCM in students majoring in WM. The results provide some suggestions for arranging TCM courses in WM schools.

(*JMIR Med Educ* 2026;12:e55972) doi:[10.2196/55972](https://doi.org/10.2196/55972)

KEYWORDS

traditional Chinese medicine; TCM; Chinese medicine; natural medicine; naturopathy; naturopathic; Western medicine students; Western medicine; modern medicine; medical students; attitude

Introduction

Traditional Chinese medicine (TCM) has been widely used to treat various diseases in China for more than 2500 years, and

it has become an integral part of Chinese culture [1]. On the basis of its holistic theory and rich practical experiences, TCM offers a diversity of therapeutic strategies for complex diseases to balance the Yin and Yang disorders of the human body [2,3].

However, Western medicine (WM) has become the mainstream medicine in modern China [4]. The number of TCM hospitals is only approximately one-fifth of the number of WM hospitals, with the number of TCM practitioners less than a quarter of that of WM physicians [5]. Medical colleges mainly focus on WM teaching, except for several colleges of TCM [6]. The curricula that WM students learn include basic and clinical medical courses, which are similar to European and American programs. Basic medical courses include human anatomy, embryology, biochemistry, physiology, pathology, pharmacology, and so on. These courses provide a theoretical foundation for clinical medicine. Clinical medicine courses focus more on practice, including internal medicine, surgery, obstetrics and gynecology, pediatrics, neurology, infectious diseases, and so on. These courses help students master the diagnostic and treatment skills needed for managing diseases. Thus, TCM receives less recognition from WM physicians and students [4]. A study among Hong Kong WM physicians showed that only 11.9% and 9.4% of them had ever considered or referred their patients to TCM, respectively [7]. Another survey found that medical students appeared to become more negative toward TCM after learning WM [8]. In addition, the positive attitude held by WM students toward TCM was less than half reported by some other surveys [8-10]. The present situation due to the negative attitude of WM students toward TCM greatly hinders the inheritance, development, and innovation of TCM in China [11]. Attitude refers to a constant psychological inclination that individuals hold toward other individuals, ideas, emotions, or events. It encompasses a subjective evaluation and influences the behavior of individuals [12]. Cognition, affection, and behavioral tendency are the 3 critical components of attitude. Among them, the cognitive component is the foundation of the other components of attitude. The emotional component is the core and key of attitude, which affects both cognitive and behavioral components. Behavioral tendency components can affect people's future reactions to attitude objects. The 3 components are interrelated and mutually restrictive [13]. Consequently, cultivating a suitable attitude toward TCM will serve to foster a precise perspective on TCM for WM students.

TCM courses in WM schools are a direct way to rectify the attitude. A survey conducted at the Capital Medical University suggested that after finishing the TCM course, more than 65% and 71.4% of WM students approved the safety and efficacy of TCM, respectively; 94.3% of students believed that TCM was worth learning; and 82.9% of students found that TCM was helpful for clinical practice [14]. It is believed that the course lets WM students understand the basic ideas, theories, and skills of TCM. Meanwhile, the course helps them understand the differences between TCM and WM, broadens their mind, and enriches the strategies of diagnosis and treatment. Students should possess a profound understanding of both TCM and WM and offer better medical service for patients in their future clinical practice [14,15].

Xiangya School of Medicine was founded in 1914 by the Hunan Society of Education and the Yale Association. It was the first WM school in China. At present, at Xiangya School of Medicine of Central South University, TCM is offered as a compulsory course for grade 3 WM students in the 5-year program (FYP)

and grade 5 in the 8-year program (EYP). It was the first time for both of them to access the TCM course. TCM is, however, less recognized by WM physicians and students, posing challenges for its integration into modern medical education as a compulsory course. This study used a self-controlled design with online questionnaires administered before and after the TCM course to WM students. The aim of the study was to explore whether TCM courses can exert an impact on the attitude of WM students toward TCM. The a priori hypothesis is that the TCM course will significantly improve WM students' attitudes toward TCM, particularly among FYP students, because of their relatively less established WM knowledge base compared to EYP students. The results are most likely to be used by medical educators, medical education researchers, WM students, future clinicians, and health care policy makers.

Methods

Description of TCM Course

According to the teaching arrangement of the dean's office at Xiangya Hospital, the TCM course for FYP and EYP students was taught by the same teaching group in the second semester of 2022—from April 15, 2022, to June 10, 2022, once a week for students in FYP, and from June 21, 2022, to August 15, 2022, once a week for students in EYP. Each theoretical part (20 class h in total) was followed by an internship (20 class h in total). The course includes the basic theory of TCM, diagnostics of TCM, Chinese pharmacy, Chinese medical formulas, acupuncture, and moxibustion. All the students were required to study *Approaching TCM*, which is a synchronous online course recorded by teachers from the Department of TCM in Xiangya Hospital. In addition, they were required to finish the self-examination after each chapter. The TCM course was delivered by teachers from the department of TCM in Xiangya Hospital. All the teachers are clinical physicians with medium or senior titles, holding qualification certificates for physicians and teacher qualifications from colleges and universities. *Traditional Chinese Medicine, Ninth Edition*, published by the People's Health Publishing House, was used as the textbook. According to the syllabus, the basic theory of TCM, diagnostics of TCM, traditional Chinese pharmacology, science of prescriptions, and acupuncture and moxibustion were taught.

Participants and Procedure

The participants comprised 145 grade 3 students in an FYP and 37 grade 5 students in an EYP. The students were asked to complete the same questionnaire before and after the TCM course online.

Ethical Considerations

This study was approved by the Ethics Committee of Hunan First Normal University (202202; [Multimedia Appendix 1](#)). Informed consent was obtained from all participants by choosing "yes" at the beginning of the online questionnaire. The information provided by participants will not be disclosed to any third party and will be used solely for the purposes of this research. All participant identifiers have been anonymized to ensure confidentiality. As the intervention involved regular

academic coursework during a natural semester and posed no risk of harm, no compensation was provided to participants on ethical grounds.

Measures

The questionnaire comprised four parts. The first part was the informed consent, including the purpose and risk of the study, the benefits of participation, and confidentiality. Part two was the basic demographic information (eg, name, sex, age, grade, and educational program). Part three was the Attitude Scale of TCM (ASTCM) developed by Wu [16]. This scale consists of 18 items divided into cognitive dimension (5 terms), emotional dimension (8 terms), and behavioral tendency factor (5 terms). The reliability test showed that internal consistency coefficients (ranging from 0.78 to 0.85), split-half reliability (ranging from 0.64 to 0.84), and retest reliability (ranging from 0.62 to 0.85) were significant. The validity test showed that the explanatory rate of the 3 factors from exploratory factor analysis was 53.58%. Confirmatory factor analysis showed that the root mean square error of approximation, comparative fit index, and non-normed fit index were 0.067, 0.903, and 0.995, respectively. The ASTCM was designed to measure the attitude of medical students toward TCM, and the reliability and validity indicators meet the measurement requirements. Physical illness is the opposite of “mental illness.” There are pathological changes in the structure of body tissues (organs) or disruptions in physiological functions. Part 4 was set to assess treatment preferences for different kinds of diseases [17], for example, whether participants would choose TCM, WM, or a combination of the two to treat functional disorders.

Statistical Analysis

Analyses were carried out using SPSS (version 23; IBM Corp). For data with homogeneous variance, the intergroup differences in attitude scores (grouped by demographic information) were analyzed using an independent sample 2-tailed *t* test, and intragroup differences (self-controlled) were analyzed using a paired 2-tailed *t* test. For data with heterogeneous variance, the intergroup differences in attitude scores (grouped by demographic information) were analyzed using the Mann-Whitney *U* test, and the intragroup differences (self-controlled) were analyzed using the Wilcoxon *W* test.

Paired sample 2-tailed *t* tests were used to analyze changes in the total, cognition, affection, and behavioral tendency scores between the pretest and posttest. The McNemar test was used for analyzing changes in the selection preference of different therapeutic strategies. Cohen *d* was used to measure the effect sizes of changes from pre- to postcourse using the following formula: $d = t \text{matched} / \sqrt{n} - \rho$, where *n* is the paired sample size and ρ refers to the correlation coefficient. Cohen suggested that $d=0.2$ be considered a “small” effect size, 0.5 represents a “medium” effect size, and 0.8 represents a “large” effect size [18].

Results

The data were collected from April 15, 2022, to August 15, 2022. In total, 118 FYP and 36 EYP students completed the before and after tests, and the data analysis was completed.

Basic Demographic Information of Participants

There were 118 and 36 questionnaires returned from 145 students in the FYP and 37 students in the EYP, respectively. Among the returned questionnaires, 118 (FYP) and 36 (EYP) were valid with before and after tests completed. Finally, 118 (FYP) and 36 (EYP) questionnaires were included for statistical analysis.

The basic demographic information of participants is presented in Table 1. The sample was 39.8% (*n*=47) male and 60.2% (*n*=71) female participants in FYP, and 38.9% (*n*=14) male and 61.1% (*n*=22) female participants in EYP (question 6 of the questionnaire). The average age was 20.64 (SD 0.70) years and 22.56 (SD 0.56) years for FYP and EYP, respectively. Of 118 FYP students, 18 (15.3%) reported that they had relatives engaged in TCM clinical or scientific research (question 9), whereas this percentage was 16.7% (6/36) among EYP students. Moreover, 74.6% (88/118) of FYP students and 66.7% (24/36) of EYP students reported that they had received TCM treatment (question 10). In addition, 64.4% (76/118) of FYP students and 52.8% (19/36) of EYP students reported that there were celebrities in the field of TCM whom they admired (question 11; Table 1).

Table . Basic demographic information of participants.

Characteristics	FYP ^a students (n=118)	EYP ^b students (n=36)
Sex, n (%)		
Male	47 (39.8)	14 (38.9)
Female	71 (60.2)	22 (61.1)
Age (y), mean (SD)	20.64 (0.70)	22.56 (0.56)
Students whose relatives engaged in TCM ^c clinical or scientific research, n (%)	18 (15.3)	6 (16.7)
Students who had received TCM treatment, n (%)	88 (74.6)	24 (66.7)
Students who admired celebrities in the field of TCM, n (%)	76 (64.4)	19 (52.8)

^aFYP: 5-year program.^bEYP: 8-year program.^cTCM: traditional Chinese medicine.

There was no significant difference in the attitude pretest scores of FYP students on questions 5, 9, and 10, but there was a significant difference on question 11. On question 11, the total attitude score and 3-factor scores of students who answered yes

were significantly higher than those who answered no (Table 2). We did not find significant differences in the pretest scores and 3-factor scores of EYP students in terms of demographic factors.

Table . The difference in the attitude pretest scores of 5-year program students (n=118) on question 11.

Score	Do you have any celebrities in the field of TCM ^a whom you admire?		t test ^b (df)	P value	Cohen d
	Yes (n=76), mean (SD)	No (n=42), mean (SD)			
Total score	68.26 (7.15)	63.07 (7.50)	3.711 (116)	<.001	0.71
Cognition score	21.97 (2.03)	21.02 (2.07)	2.421 (116)	.02	0.46
Affection score	25.95 (4.19)	23.88 (4.46)	2.506 (116)	.01	0.48
Behavior tendency score	20.34 (3.04)	18.17 (2.76)	3.848 (116)	<.001	0.74

^aTCM: traditional Chinese medicine.^bIndependent sample 2-tailed t test.

The Effect of TCM Course on Attitude Toward TCM

For FYP students, there was a significant increase in the total mean score of ASTCM after the TCM course (from 66.42, SD 7.66 to 71.43, SD 7.38; $P<.001$). A significant increase was also observed in the mean scores of the 3 factors: cognition (from 21.64, SD 2.08 to 22.90, SD 1.94; $P<.001$), affection (from 25.21, SD 4.39 to 27.96, SD 4.4; $P<.001$), and behavioral

tendency (from 19.57, SD 3.02 to 20.58, SD 2.76; $P<.001$). Interestingly, except for the behavioral tendency mean score (from 17.50, SD 3.54 to 18.78, SD 3.22; $P=.02$), a significant increase was not detected in the total mean score (from 60.36, SD 10.53 to 62.92, SD 10.05; $P=.19$), cognition (from 20.50, SD 2.73 to 20.69, SD 2.73; $P=.69$), and affection of ASTCM (from 22.36, SD 6.32 to 23.44, SD 5.84; $P=.34$) in EYP students (Table 3). The effect sizes (Cohen d) are presented in Table 3.

Table . Changes in attitude scores of Western medicine students.

Score	FYP ^a (n=118)					EYP ^b (n=36)				
	Precourse, mean (SD)	Postcourse, mean (SD)	t test ^c (df)	P value	Cohen d	Precourse, mean (SD)	Postcourse, mean (SD)	t test (df)	P value	Cohen d
Total score	66.42 (7.66)	71.43 (7.38)	-7.206 (117)	<.001	-0.67	60.36 (10.53)	62.92 (10.05)	-1.353 (35)	.19	-0.25
Cognition score	21.64 (2.08)	22.90 (1.94)	-6.276 (117)	<.001	-0.63	20.50 (2.73)	20.69 (2.73)	-0.397 (35)	.69	-0.07
Affection score	25.21 (4.39)	27.96 (4.45)	-6.912 (117)	<.001	-0.6	22.36 (6.32)	23.44 (5.84)	-0.976 (35)	.34	-0.18
Behavioral tendency score	19.57 (3.02)	20.58 (2.76)	-3.892 (117)	<.001	-0.34	17.50 (3.54)	18.78 (3.22)	-2.354 (35)	.02	-0.38

^aFYP: 5-year program.^bEYP: 8-year program.^cPaired sample 2-tailed *t* test.

In addition, we found that the attitude scores of FYP students significantly increased after the TCM course, regardless of whether they admired a Chinese medicine celebrity. However,

no significant difference was observed between students who admired a TCM celebrity and those who did not in the posttest attitude scores (Table 4).

Table . The difference in the attitude scores between pretest and posttest of 5-year program students on question 11 (n=118).

Score	Do you have any celebrities in the field of TCM ^a whom you admire?									
	Yes (n=76)					No (n=42)				
	Precourse, mean (SD)	Postcourse, mean (SD) ^b	t test ^c (df)	P value	Cohen d	Precourse, mean (SD)	Postcourse, mean (SD) ^b	t test (df)	P value	Cohen d
Total score	68.26 (7.15)	72.54 (5.13)	-5.803 (75)	<.001	-0.60	63.07 (7.50)	69.63 (10.04)	-4.472 (41)	<.001	-0.70
Cognition score	21.97 (2.03)	23.14 (1.46)	-5.221 (75)	<.001	-0.58	21.02 (2.07)	22.45 (2.55)	-3.606 (41)	.001	-0.61
Affection score	25.95 (4.19)	28.42 (3.70)	-5.608 (75)	<.001	-0.55	23.88 (4.46)	27.12 (5.52)	-4.140 (41)	<.001	-0.64
Behavioral tendency score	20.34 (3.04)	20.97 (2.13)	-2.120 (75)	.04	-0.22	18.17 (2.76)	19.86 (3.55)	-3.548 (41)	.001	-0.53

^aTCM: traditional Chinese medicine.

^bDue to the heterogeneity of variance, we compared the posttest attitude scores between groups using the Mann-Whitney *U* test, and the result showed no significant difference.

^cPaired sample 2-tailed *t* test.

The Effect of TCM Course on Selection Preference of Different Therapeutic Strategies

For FYP students, there was a significant difference in the selection preferences of different therapeutic strategies for acute ($P=.025$), chronic ($P=.02$), and physical diseases ($P=.004$). After the TCM course, the percentage of students who preferred WM for acute diseases decreased from 94.1% (n=111) to 83.9% (n=99). The preference for the combination of TCM and WM (TCM and WM) increased from 5.1% (n=6) to 14.4% (n=17), and the preference for TCM increased from 0.8% (n=1) to 1.7% (n=2). For chronic diseases, fewer students preferred WM (from n=10, 8.5% to n=1, 0.8%), whereas more students preferred TCM (from n=53, 44.9% to n=72, 61%). As for physical

diseases, the percentage of students preferring WM declined (from n=56, 47.5% to n=39, 33.1%), whereas the preference for TCM and WM increased (from n=52, 44.1% to n=78, 66.1%; Table 5).

For EYP students, the significant difference in the selection preferences for different therapeutic strategies mainly focused on internal diseases ($P=.02$), surgical diseases in the perioperative period ($P=.01$), and mental illnesses ($P=.02$). More students realized that TCM and WM was a better choice (from n=12, 33.3%-52.8%) for internal diseases. The percentage of students who preferred TCM and WM for perioperative surgical diseases nearly tripled from 8.3% (n=3) to 25% (n=9). There was also an approximately 20% point increase in WM and TCM

preference for mental illnesses (from n=51, 41.7% to n=22, 36.1%; **Table 3**). 61.1%) with a marked decline in WM (n=19, 52.8% to n=13,

Table 3. Change in treatment preference of Western medicine (WM) students (n=118).

Program, disease, treatment preference, and data type	Before course, n (%)	After course, n (%)	P value ^a
FYP^b			
Acute diseases			.02
TCM ^c	1 (0.8)	2 (1.7)	
WM	111 (94.1)	99 (83.9)	
TCM and WM	6 (5.1)	17 (14.4)	
Chronic diseases			.003
TCM	53 (44.9)	72 (61)	
WM	10 (8.5)	1 (0.8)	
TCM and WM	55 (46.6)	45 (38.1)	
Somatic disease			.02
TCM	10 (8.5)	7 (5.9)	
WM	56 (47.5)	39 (33.1)	
TCM and WM	52 (44.1)	78 (61)	
EYP^d			
Internal diseases			.02
TCM	4 (11.1)	2 (5.6)	
WM	20 (55.6)	15 (41.7)	
TCM and WM	12 (33.3)	19 (52.8)	
Surgical diseases (perioperative period)			.01
TCM	1 (2.8)	1 (2.8)	
WM	32 (88.9)	26 (72.2)	
TCM and WM	3 (8.3)	9 (25)	
Mental illnesses			.02
TCM	2 (5.6)	1 (2.8)	
WM	19 (52.8)	13 (36.1)	
TCM and WM	15 (41.7)	22 (61.1)	

^aMcNemar test.

^bFYP: 5-year program.

^cTCM: traditional Chinese medicine.

^dEYP: 8-year program.

Discussion

Principal Findings

This study was conducted in natural classes, based on the real teaching arrangements of the dean's office at Xiangya Hospital. The study examined the effect of TCM courses on the attitude toward TCM in students majoring in WM using the ASTCM. The study found that the course significantly increased the total score and the scores of the 3 factors of ASTCM in FYP students but did not change these scores in EYP students, except for the score of the behavioral factor. The study also found a significant

change in therapeutic strategy preferences before and after TCM learning, with the percentage reduction in WM and increase in TCM or TCM and WM. The alteration in therapeutic approach inclination primarily manifested in acute, chronic, and somatic ailments among FYP scholars and in internal, perioperative surgical, and psychological disorders in EYP learners. The results suggested that TCM courses increased a positive attitude toward TCM for WM students.

Basic Demographic Information of Participants

The demographic results showed that there were more female students in both FYP and EYP groups. The result was similar

to the gender ratio officially reported by Central South University. The age of EYP students was 2 years older than that of FYP students. This was in line with the reality. An interesting result was that FYP students who admired Chinese medicine celebrities had a significantly better attitude toward TCM than those who did not admire Chinese medicine celebrities. This, to some extent, indicates that celebrity worship had a positive impact on attitudes toward TCM [19].

The Effect of TCM Course on Cognitive Component of Attitude in FYP

This study reported a more positive attitude toward TCM in FYP students after completing the TCM course. The results were supported by another study carried out at the Capital Medical University, which found that WM students became more positive toward TCM after learning about it [20]. Attitude is a disposition toward or against a specified phenomenon, person, or thing. An attitude comprises cognitive, affective, and behavioral tendency components [21]. It is believed that attitude can be changed by interventions or messages delivered at a particular time from the perspective of cognition, affection, and behavioral tendency [22].

TCM is generally considered to be mysterious, profound, and lacking a scientific basis [23]. By learning the basic theory of TCM, students gained a preliminary understanding of the basic theoretical framework of TCM. They learned that TCM is not mystical but has a solid foundation to support. They also learned how TCM physicians diagnose diseases with their unique syndrome differentiation system by learning the diagnostics of TCM. They witnessed how to use so-called “flowers and plants” (herbs) to treat diseases by learning Chinese pharmacy and medical formulas. In the acupuncture and moxibustion courses, the mysterious meridians were demystified and presented to students. All the abovementioned courses are believed to improve the students’ attitude toward TCM from the perspective of cognition.

The Effect of TCM Course on Affective Component of Attitude in FYP

Improvement in the affection component of attitude was also observed after the TCM course. It is a form of identification and resonance. First, TCM was acknowledged by students rationally with correct cognition via theoretical lectures. Second, identification and resonance of students were achieved through internships, which are considered the most crucial experiences for undergraduate medical students [24]. The students witnessed the remarkable therapeutic effect of TCM. The cases showed that 2 doses of medicine cured a 3-month persistent cough, and immediate acupuncture for acute knee injuries were often shown. Some students even experienced TCM for dysmenorrhea: the entire diagnosis and treatment process, from pulse feeling to prescription. Finally, during the study, many famous TCM physicians were known to students and became their admired seniors. It greatly increased the affection component of attitude, as shown in the result of question 11.

The Effect of TCM Course on Behavioral Tendency Component of Attitude in FYP

The behavioral tendency component of attitude also improved after the TCM course among FYP students. The clinical internship after the theoretical lecture was also conducted to change the behavioral component of attitude. The internship aimed at integrating TCM theory with practice. In this way, abstract TCM theories and methods would be specified [25]. For example, students visited patients with different syndromes, and they were taught the yin and yang attributes of syndromes according to the Yin-Yang theory. The internship also showed the students a complete diagnosis and treatment process of TCM, which included 4 diagnostic methods (ie, observation, auscultation and olfaction, inquiry, and pulse feeling and palpation), syndrome differentiation, and treatment as well [26]. During the internship, students experienced some unique diagnostic methods of TCM, such as tongue diagnosis, pulse feeling, and palpation. Students even tried the needling sensation (Deqi) themselves in the internship. Hence, the clinical internship afforded the students a hands-on and introspective encounter with TCM, rather than mere hearsay and boastful claims. It was of great importance for students to learn about the unique advantages of TCM, to eliminate prejudices about TCM, and, finally, to make a more objective and comprehensive judgment about TCM.

The Effect of Chinese Medicine Celebrity and TCM Course on Attitude in FYP

There was also an interesting result that before the course, students who admired Chinese medicine celebrity worship had a more positive attitude toward TCM compared with those who did not. However, after the course, there was no intergroup difference in attitudes between these 2 groups of students, although there were intragroup differences. In other words, the course made the attitudes of the 2 groups of students, which were originally different, become consistent and significantly more positive. The hierarchical effect may explain this interesting result. Students who did not admire Chinese medicine celebrities may follow a standard learning hierarchy. In this learning hierarchy, cognition was the foundation, which in turn affects emotions and ultimately behavioral tendencies. Students who admired Chinese medicine celebrities followed an experiential hierarchy, in which emotions were the foundation that influenced behavior and ultimately cognition.

Why the TCM Course Did Not Affect Cognitive and Affective Components of Attitude in EYP

Compared with FYP students, it was very interesting that TCM courses did not significantly change the total score or the scores in the cognition and affection domains in EYP students. Some facts attracted our attention that might be reasons for quite different results. First, 5th-grade EYP students had completed WM diagnostics and accessed almost all the clinical professional courses, including internal medicine, surgery, gynecology, and pediatrics, alike. Moreover, they had determined their future professional major and selected their doctoral supervisors. Therefore, long-term WM clinical training had already established an almost complete WM knowledge system and the mode of thinking for students in EYP. It was difficult to be

changed by TCM learning at a relatively later stage. It was a formidable challenge for EYP students to acknowledge and assimilate TCM, given their preconceived learning of WM, because of the fact that these 2 domains are founded on divergent theories and modalities [27-29]. However, for FYP students, 3-year learning in basic and clinical bridging courses has not yet fully westernized their thinking. They were relatively simple to change through TCM learning. This might be one of the critical reasons why total scores and scores in cognition and affection did not increase in EYP. The findings of this study were in accordance with the report of Hon et al [8] that preclinical students were more positive toward TCM compared to students in clinical years, the latter showing a more negative attitude toward TCM after studying WM. In addition, the small sample size included (36 EYP students) might be another important reason for the absence of a difference between before and after tests. It is difficult to avoid the fact that the college entrance examination enrollment of the EYP is very limited. Future studies should adopt more rigorous experimental designs to address this limitation, for example, a combination of effects of several consecutive years should be feasible.

The Effect of TCM Course on Behavioral Tendency Component of Attitude in EYP

The behavioral tendency component of attitude among EYP students was improved after the course. Behavioral intention is different from actual behavior, but it is an individual's behavioral readiness toward a certain object, representing a possible behavioral tendency. The result suggests that after learning TCM, EYP students gained more in-depth exposure to another medical strategy. It is greatly different from WM yet effective, well founded, and uniquely advantageous, and can solve some problems that WM cannot solve. Therefore, they may be willing to use TCM for health maintenance, to use the knowledge of TCM to safeguard their parents' health, to learn more about TCM, and to consider using TCM to solve patients' illnesses in future clinical work.

The Effect of TCM Course on Selection Preference of Different Therapeutic Strategies in FYP

The purpose of this part of the questionnaire was to test the behavior tendency in specific clinical conditions. The test was designed to simulate a physician's daily work, such as making an appropriate treatment plan for different types of diseases or, at least, giving proper referrals for different patients. After completing the TCM course, more FYP students realized that TCM and WM were very effective in treating acute diseases, which broke the stereotype that TCM was only suitable for chronic diseases. Many real cases they witnessed during their TCM internship may have influenced their selection preference. For example, a TCM physician cured a child with unexplained persistent high fever with a combination of xiaochaihu decoction and gypsum soup. As for chronic diseases, more FYP students tended to choose TCM treatment after the course. These diseases often have a long course and complex condition, requiring long-term management and treatment. The TCM course taught students that TCM relies on its unique theoretical system and rich practical experience and has demonstrated significant advantages and potential [30]. It emphasizes a holistic concept

and treatment with syndrome differentiation for chronic diseases, which helps physicians treat patients using the overall situation, comprehensively considering the patients' physical condition, lifestyle habits, emotional factors, and other aspects. As for organic diseases, more FYP students realized that the strategy of using both TCM and WM is better. This result is reasonable and explainable. It is widely recognized that WM is the mainstream and effective treatment for organic diseases. However, during the internship, students found that participation of TCM in the treatment of organic diseases accelerated recovery, minimized side effects, and increased patient comfort. Thus, they tended to use TCM and WM for organic diseases.

The Effect of TCM Course on Selection Preference of Different Therapeutic Strategies in EYP

After the TCM course, more EYP students chose TCM and WM to treat internal diseases. EYP students had completed many clinical courses, so they were more likely to divide diseases into internal and surgical diseases according to their scope and treatment methods. Generally speaking, internal medicine diseases can be treated with medication. By learning the basic theory of TCM, along with Chinese pharmacy and Chinese medical formulas, they understood how TCM cures diseases. It is greatly different from that of WM but is a good supplement and promotion to WM. However, for surgical diseases, surgical treatment remains the first choice undoubtedly. Nevertheless, EYP students found that the TCM therapies, such as acupuncture and massage, help restore gastrointestinal, cognitive, and limb functions in postoperative patients. Therefore, they preferred to use a combination of TCM and WM to address perioperative disorders. Mental illness is another major category of disease. The therapeutic effect of antipsychotic drugs is limited, and there are many side effects. Mental illness belongs to the category of emotional disorders in TCM. By learning TCM etiology and pathogenesis, students understood that emotional disorders are related to various factors, such as organ dysfunction, yin-yang imbalance, and so on. Common treatment methods include regulating qi and relieving depression, calming the heart and eliminating phlegm, purging the liver and clearing fire, and so on. EYP students also found that TCM and WM treatment showed a satisfactory effect in patients with mental illness. Thus, more EYP students preferred a TCM and WM strategy for mental illness. The results additionally aligned with a previous study, which found that TCM was not practiced in isolation but in conjunction with WM [17]. The increased participation of TCM in treatment was expected to enhance its effectiveness.

Implications of Findings

The findings of the present study highlight the importance of integrating TCM into WM education and practice. It will lead to more holistic and patient-centered care, enhance clinical outcomes, and contribute to the development of integrated health care policies. Future research should continue to explore the long-term effects of TCM education on WM students' clinical practice and the potential benefits of integrating TCM and WM in various health care settings.

Comparison With Prior Literature

The findings of this study align with and extend several key themes observed in the existing literature on the integration of TCM and WM, particularly in the context of medical education and attitude formation. This study builds on existing literature by confirming the positive impact of TCM education on WM students' attitudes and by extending the understanding of how this impact varies by educational stage and specific disease contexts. The detailed analysis of attitude components and the introduction of novel factors, such as celebrity worship, provide new insights that can inform future research and educational practices in the integration of TCM and WM.

Strengths and Limitations

This study has several distinct advantages that contribute to its overall quality and relevance. The study used a self-controlled design, which is a robust method for assessing changes in attitudes over time. By comparing the same group of students before and after the TCM course, the study effectively controls for individual differences and external variables, providing a more accurate assessment of the impact of the TCM course. The study used the ASTCM, a validated and reliable tool designed specifically to measure medical students' attitudes toward TCM through a multidimensional analysis. This ensured that the measurements are accurate, reproducible, and reflective

of the true changes in students' attitudes. Some inevitable limitations in this study should be pointed out. First, a self-administered questionnaire was used instead of an interview. The reliability of the answers may be discounted. The qualitative interview should be included to make the study more persuasive in our future work. Second, the questionnaire is a kind of survey based on text reading and filling in answers. To avoid the feeling of boredom in participants, the questionnaire was designed to be relatively brief. Thus, it was impossible to have an in-depth examination. Moreover, it is an online questionnaire. Researchers possess a limited understanding of the actual identity of the participant, the mood of the participant at that time, and whether the process of finishing the questionnaire was influenced by others. Finally, the sample size of EYP students was relatively small. In the 2-tailed *t* test, the smaller sample size makes it harder to gain differential results. Thus, the results may be influenced. A sufficient sample size or combination of effects of EYP for several consecutive years is expected in the future.

Conclusions

The study showed that offering a TCM course increased a positive attitude toward TCM in students majoring in WM. The results will provide recommendations with regard to the TCM curriculum arrangement in institutions of WM learning.

Acknowledgments

The authors would like to thank Dan Han, School of Medicine, Zhejiang University, for data collection and analysis.

Funding

This study was supported by grants from the Scientific Research Project of Hunan Provincial Department of Education (22B0876).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Ethical approval document.

[[PDF File, 173 KB - mededu_v12i1e55972_app1.pdf](#)]

References

1. Zhang YL, Wang YL, Yan K, et al. Traditional Chinese medicine formulae QY305 reducing cutaneous adverse reaction and diarrhea by its nanostructure. *Adv Sci (Weinh)* 2024 Feb;11(5):e2306140. [doi: [10.1002/advs.202306140](https://doi.org/10.1002/advs.202306140)] [Medline: [38044276](https://pubmed.ncbi.nlm.nih.gov/38044276/)]
2. Liu S, Zhu JJ, Li JC. The interpretation of human body in traditional Chinese medicine and its influence on the characteristics of TCM theory. *Anat Rec (Hoboken)* 2021 Nov;304(11):2559-2565. [doi: [10.1002/ar.24643](https://doi.org/10.1002/ar.24643)] [Medline: [34117702](https://pubmed.ncbi.nlm.nih.gov/34117702/)]
3. Weng H, Chen J, Ou A, Lao Y. Leveraging representation learning for the construction and application of a knowledge graph for traditional Chinese medicine: framework development study. *JMIR Med Inform* 2022 Sep 2;10(9):e38414. [doi: [10.2196/38414](https://doi.org/10.2196/38414)] [Medline: [36053574](https://pubmed.ncbi.nlm.nih.gov/36053574/)]
4. Jin L, Lin L, Liu G, et al. Inconsistency between physicians' attitudes and behaviors toward TCM: a questionnaire survey in Chengdu. *Medicine (Baltimore)* 2024 Mar 22;103(12):e37596. [doi: [10.1097/MD.00000000000037596](https://doi.org/10.1097/MD.00000000000037596)] [Medline: [38518017](https://pubmed.ncbi.nlm.nih.gov/38518017/)]
5. National Health Commission. China Health Statistics Yearbook of 2022: Peking Union Medical College Press; 2022. URL: <https://www.nhc.gov.cn/mohwsbwstjxxzx/tjtjn/202305/49beded3bd984669bfe9089c6f231cf5.shtml> [accessed 2026-01-02]
6. Liu X, Feng J, Liu C, et al. Medical education systems in China: development, status, and evaluation. *Acad Med* 2023 Jan 1;98(1):43-49. [doi: [10.1097/ACM.0000000000004919](https://doi.org/10.1097/ACM.0000000000004919)] [Medline: [35947483](https://pubmed.ncbi.nlm.nih.gov/35947483/)]

7. Chung VCH, Hillier S, Lau CH, Wong SYS, Yeoh EK, Griffiths SM. Referral to and attitude towards traditional Chinese medicine amongst western medical doctors in postcolonial Hong Kong. *Soc Sci Med* 2011 Jan;72(2):247-255. [doi: [10.1016/j.socscimed.2010.10.021](https://doi.org/10.1016/j.socscimed.2010.10.021)] [Medline: [21145150](#)]
8. Hon KLE, Leung TF, Tse HM, et al. A survey of attitudes to traditional Chinese medicine among Chinese medical students. *Am J Chin Med* 2005;33(2):269-279. [doi: [10.1142/S0192415X05002904](https://doi.org/10.1142/S0192415X05002904)] [Medline: [15974486](#)]
9. Hon EKL, Lee K, Tse HM, et al. A survey of attitudes to traditional Chinese medicine in Hong Kong pharmacy students. *Complement Ther Med* 2004 Mar;12(1):51-56. [doi: [10.1016/j.ctim.2003.12.002](https://doi.org/10.1016/j.ctim.2003.12.002)] [Medline: [15130573](#)]
10. Hon KLE, Twinn SF, Leung TF, Thompson DR, Wong Y, Fok TF. Chinese nursing students' attitudes toward traditional Chinese medicine. *J Nurs Educ* 2006 May;45(5):182-185. [doi: [10.3928/01484834-20060501-08](https://doi.org/10.3928/01484834-20060501-08)] [Medline: [16722501](#)]
11. Yu F, Gao J. Challenges that the law of traditional Chinese medicine has to face. *Cap Food Med* 2016;23(9):21-22 [FREE Full text]
12. Ng TKC, Lo MF, Fong BYF, Yee HHL. Predictors of the intention to use traditional Chinese medicine (TCM) using extended theory of planned behavior: a cross-sectional study among TCM users in Hong Kong. *BMC Complement Med Ther* 2022 Apr 22;22(1):113. [doi: [10.1186/s12906-022-03598-x](https://doi.org/10.1186/s12906-022-03598-x)] [Medline: [35459198](#)]
13. Corchado Rodríguez JM, Uden L, Wang LSL, Yang H, editors. The 8th International Conference on Knowledge Management in Organizations: Social and Big Data Computing for Knowledge Management, 1st edition: Springer; 2014. [doi: [10.1007/978-94-007-7287-8](https://doi.org/10.1007/978-94-007-7287-8)]
14. Gan X, Yang J, Yang B, Yuan J. Problems and strategies of traditional Chinese medicine teaching in western medical colleges. *J Pract Intern Med Tradit Chin Med* 2020;34(7):11-14. [doi: [10.13729/j.issn.1671-7813.Z20200253](https://doi.org/10.13729/j.issn.1671-7813.Z20200253)]
15. Xu K. Thoughts on traditional Chinese medicine teaching in western medical colleges. *Chin Med Med Mod Distance Educ China* 2021;4(9):19-21. [doi: [10.3969/j.issn.1672-2779.2021.04.007](https://doi.org/10.3969/j.issn.1672-2779.2021.04.007)]
16. Wu Z. Development of Chinese Medicine Attitude Scale [Master's thesis]. : Beijing University of Traditional Chinese Medicine; 2011.
17. Harmsworth K, Lewith GT. Attitudes to traditional Chinese medicine amongst western trained doctors in the People's Republic of China. *Soc Sci Med* 2001 Jan;52(1):149-153. [doi: [10.1016/s0277-9536\(00\)00124-6](https://doi.org/10.1016/s0277-9536(00)00124-6)] [Medline: [11144912](#)]
18. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol Rev Camb Philos Soc* 2007 Nov;82(4):591-605. [doi: [10.1111/j.1469-185X.2007.00027.x](https://doi.org/10.1111/j.1469-185X.2007.00027.x)] [Medline: [17944619](#)]
19. Lu Y. Study on the phenomenon of celebrity worship among college students in the new era [Master's thesis]. : Anhui University of Engineering; 2023 URL: <https://tinyurl.com/3wy9kh2m> [accessed 2026-01-02]
20. Xuan Y, Haihong L, Tao S, et al. Investigation and analysis of traditional Chinese medicine curriculum in clinical specialty of western medicine. *Int J Tradit Chin Med* 2018;40(6):554-557. [doi: [10.3760/cma.j.issn.1673-4246.2018.06.018](https://doi.org/10.3760/cma.j.issn.1673-4246.2018.06.018)]
21. Altmann TK. Attitude: a concept analysis. *Nurs Forum (Auckl)* 2008 Jul;43(3):144-150. [doi: [10.1111/j.1744-6198.2008.00106.x](https://doi.org/10.1111/j.1744-6198.2008.00106.x)]
22. Albarracín D, Shavitt S. Attitudes and attitude change. *Annu Rev Psychol* 2018 Jan 4;69(69):299-327. [doi: [10.1146/annurev-psych-122216-011911](https://doi.org/10.1146/annurev-psych-122216-011911)] [Medline: [28841390](#)]
23. Huang Z, Miao J, Chen J, et al. A traditional Chinese medicine syndrome classification model based on cross-feature generation by convolution neural network: model development and validation. *JMIR Med Inform* 2022 Apr 6;10(4):e29290. [doi: [10.2196/29290](https://doi.org/10.2196/29290)] [Medline: [35384854](#)]
24. 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](https://doi.org/10.2196/43763)] [Medline: [37728989](#)]
25. Zhou Y, Shen H. Research and exploration on traditional Chinese medicine as an internship in western medical colleges. *Curr Educ Res* 2020(29):122-124.
26. Lu L, Lu T, Tian C, Zhang X. AI: Bridging ancient wisdom and modern innovation in traditional Chinese medicine. *JMIR Med Inform* 2024 Jun 28;12:e58491. [doi: [10.2196/58491](https://doi.org/10.2196/58491)] [Medline: [38941141](#)]
27. Li A, Yajie L, Guoqiang C. Research on teaching methods of traditional Chinese medicine in colleges and universities of western medicine. *Med Inf* 2018;31(12):21-22. [doi: [10.3969/j.issn.1006-1959.2018.12.007](https://doi.org/10.3969/j.issn.1006-1959.2018.12.007)]
28. Hu S. Traditional Chinese medicine teaching material tailored for western hospital school: "Xiehe edition" traditional Chinese medicine. *Med J Peking Union Med Coll Hosp* 2012;3(2):242-243 [FREE Full text]
29. Liu XH, Jin F, Hsu J, Li DN, Chen W. Comparing smartphone apps for traditional Chinese medicine and modern medicine in China: systematic search and content analysis. *JMIR Mhealth Uhealth* 2021 Mar 24;9(3):e27406. [doi: [10.2196/27406](https://doi.org/10.2196/27406)] [Medline: [33759786](#)]
30. Wang J, Chen L, Wang Z, Fang X, Zhen Y. Perceptions, attitudes, beliefs, and practices toward traditional Chinese medicine therapies for herpes zoster: a cross-sectional study of healthcare professionals. *Risk Manag Healthc Policy* 2024;17:1239-1251. [doi: [10.2147/RMHP.S461149](https://doi.org/10.2147/RMHP.S461149)] [Medline: [38765781](#)]

Abbreviations

ASTCM: Attitude Scale of Traditional Chinese Medicine

EYP: 8-year program**FYP:** 5-year program**TCM:** traditional Chinese medicine**WM:** Western medicine

Edited by B Lesselroth; submitted 01.Jan.2024; peer-reviewed by L Yu, LM Chen; revised version received 19.Apr.2025; accepted 13.May.2025; published 16.Jan.2026.

Please cite as:

He H, Cui H, Guo L, Liao Y

The Effect of a Traditional Chinese Medicine Course on Western Medicine Students' Attitudes Toward Traditional Chinese Medicine: Self-Controlled Pre-Post Questionnaire Study

JMIR Med Educ 2026;12:e55972

URL: <https://mededu.jmir.org/2026/1/e55972>

doi:10.2196/55972

© Haoyu He, Hanjin Cui, Li Guo, Yanhui Liao. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 16.Jan.2026. 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.

Original Paper

Interactive, Image-Based Modules as a Complement to Prosection-Based Anatomy Laboratories: Multicohort Evaluation

Caroline Sumner^{1*}, BS; Sami L Case^{2*}, PhD; Samuel Franklin³, PhD; Kristen Platt³, PhD

¹College of Medicine, University of Kentucky, Lexington, KY, United States

²Department of Biomedical Sciences, College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO, United States

³Department of Neuroscience, College of Medicine, University of Kentucky, Lexington, KY, United States

* these authors contributed equally

Corresponding Author:

Kristen Platt, PhD
Department of Neuroscience
College of Medicine
University of Kentucky
780 Rose Street
MS 207 Medical Sciences Building
Lexington, KY, 40536
United States
Phone: 1 8592186648
Fax: 1 8593235946
Email: platt.kristen@uky.edu

Abstract

Background: As medical and allied health curricula adapt to increasing time constraints, ethical considerations, and resource limitations, digital innovations are becoming vital supplements to donor-based anatomy instruction. While prior studies have examined the effectiveness of prosection versus dissection and the role of digital tools in anatomy learning, few resources align interactive digital modules directly with hands-on prosection experiences.

Objective: This project addresses that gap by introducing an integrated, curriculum-aligned platform for self-guided cadaveric learning.

Methods: We created Anatomy Interactives, a web-based laboratory manual structured to complement prosection laboratories for MD, DPT, and PA students. Modules were developed using iSpring Suite (iSpring Solutions Incorporated) and included interactive labeled images, donor photographs, and quiz-style self-assessments. Learners engaged with modules before, during, or after laboratory sessions. PA/DPT and MD students completed postcourse surveys evaluating module use and perceived impact. MD student examination scores from a 2023 cohort (no module access) were compared to a 2024 cohort (with access) to evaluate effectiveness.

Results: A total of 147 students completed the survey (31 PA/DPT and 116 MD). The majority reported using modules for 1-2 hours per week and found them helpful for both written and laboratory examinations. MD students in the 2024 cohort performed better on all 3 examinations compared to the 2023 cohort, with 2 examination median differences reaching statistical significance (Mann-Whitney U, $P<.001$). Qualitative feedback highlighted accessibility, content reinforcement, and user engagement as key benefits.

Conclusions: Interactive modules integrated with prosection laboratories enhanced learner engagement and performance. This hybrid digital-donor model shows promise for scalable, learner-centered gross anatomy education.

(*JMIR Med Educ* 2026;12:e85028) doi:[10.2196/85028](https://doi.org/10.2196/85028)

KEYWORDS

digital learning; education; medical; gross anatomy; prosection; hybrid anatomy education

Introduction

Gross anatomy remains a foundational component of health professions education, providing essential knowledge for clinical reasoning, procedural competency, and disease management [1]. However, students in prosection-based courses often struggle to transition from routine passive study to active, tactile engagement during laboratory sessions. Most existing digital resources are intended for use before or after the laboratory rather than functioning as interactive tools to engage the learner at the dissection table. As a result, students may feel disconnected from the donor during prosection, limiting the depth and integration of their learning experience.

Historically, cadaveric dissection has been regarded as the pedagogical gold standard for anatomy education [2,3]. However, contemporary anatomical curricula face mounting logistical, ethical, and financial constraints. These constraints include high costs for donor procurement and preservation, limited laboratory space and faculty, exposure to formaldehyde and other chemicals, and ethical concerns. These forces are driving the adoption of alternative teaching approaches such as prosection, peer-assisted learning, and digital platforms [4].

A growing body of literature suggests that prosection-based learning can be as effective as, and in some cases more efficient than, dissection in facilitating anatomical knowledge acquisition and clinical relevance [5-7]. A comparative assessment demonstrated that nonmedical undergraduate students reported similar academic outcomes and levels of engagement regardless of whether they were taught through dissection or prosection [3]. Likewise, in a study of veterinary students, prosections prepared by qualified staff outperformed traditional dissection in promoting both efficiency and student satisfaction [8]. Coker et al [9] report that students who learned by prosection performed better on questions specifically involving complex or deep structures, but simultaneously demonstrated higher satisfaction with and preference for a dissection-based model.

Concurrently, the integration of digital technologies into anatomy curricula is transforming how learners engage with spatial content. Radiographic imaging is a time-tested methodology that has been incorporated into undergraduate medical education with increasing frequency [10-12]. Ultrasound-based anatomy instruction has grown in popularity in recent years [13,14]. In addition, meta-analytic findings underscore the effectiveness of virtual- and augmented-reality platforms in reinforcing anatomical understanding and enhancing learner autonomy [15,16]. This shift was accelerated by the COVID-19 pandemic, during which hybrid and online formats became critical for continuity in medical education [17].

Other commercially available digital tools serve to augment the didactic anatomy learning experience. For example, the Anatomage and Sectra digital dissection technologies rely on the Visible Human Project work to generate a donor experience in physical format reminiscent of a donor table [18,19]. Use of virtual dissection as a supplement to donor dissection has been viewed favorably by students [20]. The Anatomage also offers unique advantages, such as blood flow modeling that cannot be replicated in a donor [21]. However, these devices are often

proposed as an alternative to the study of the donor body and do not address the gap wherein digital media can complement the study of the prosected donor in real time. Models such as hybrid prosection-based anatomy laboratories demonstrated high student satisfaction, with learners appreciating the flexibility and clarity offered by digital supplements [7].

Several institutions have developed innovative digital platforms to complement or enhance traditional gross anatomy instruction, exemplifying how technology can be leveraged to support donor-based learning. For instance, the University of Michigan's BlueLink program offers students an extensive library of presentation slide-like materials rich with cadaveric images, facilitating self-directed study and review [22]. These resources, which also include laboratory manuals and interactive files, are designed to provide a flexible and widely accessible digital ecosystem for anatomy education. Similarly, Texas Tech University Health Sciences Center has developed a gross anatomy program featuring a varied suite of online tools [23]. Their offerings include prelaboratory cadaveric dissection videos, practice quizzes for reinforcing knowledge of cadaveric material, and various other interactive learning modules, providing multifaceted support for students beyond traditional laboratory settings.

From a theoretical perspective, multimodal learning and constructivist principles provide a useful lens for understanding why integrated digital prosection tools may enhance anatomy learning. Multimodal learning theory posits that students build stronger and more durable knowledge when information is presented through complementary channels, such as visual, spatial, textual, and tactile, which is an especially relevant consideration in gross anatomy [24]. Constructivist frameworks further suggest that learners deepen understanding by actively engaging with material, generating connections, and iteratively refining mental models during hands-on exploration, and this approach has been shown to facilitate more meaningful learning than traditional methods in the context of human anatomy [25]. Past constructivist approaches in anatomical sciences education include techniques such as casting and creating models [26,27]. Digital resources that integrate with prosection-based laboratory work have the potential to support meaning-making in real time, helping students actively construct anatomical understanding while drawing on multiple modes of representation.

Despite the growing prevalence of prosection-based anatomy coursework and the surge of fully virtual learning models, few resources are designed specifically to bridge digital learning with hands-on prosected donor experiences in a structured, curriculum-aligned format. The Anatomy Interactives platform was developed to fill this gap, providing guided modules that integrate interactive digital content with cadaveric prosections [28]. Unlike static resources or purely virtual simulations, this tool is designed to reinforce laboratory-based learning through active engagement, self-assessment, and iterative feedback in a digital format while students are present in the gross anatomy laboratory.

This study evaluates the effectiveness of the Anatomy Interactives platform in 2 different student populations and implementation formats. This includes analyses of student

engagement, self-reported preparedness, and academic performance. This work will interest medical and health professions students and educators whose classroom is the cadaveric laboratory. These learners often face challenges integrating didactic anatomical knowledge with hands-on experiences, particularly when transitioning between digital prelaboratory preparation and in-laboratory prosection-based learning. By comparing student experiences and performance across 2 cohorts, one with access to the digital modules and one without, we aim to provide evidence for the impact of integrated, digital-prosection hybrid models on anatomy education for future health professionals.

Methods

Ethical Considerations

All work with living subjects (ie, surveys, website traffic, and grades) was reviewed by the University of Kentucky Institutional Review Board (IRB) and determined to be exempt (IRB Protocol #94449) from human subjects research oversight. All procedures comply with the World Medical Association Declaration of Helsinki. The donor registration form of the University of Kentucky Willed Body Program includes language of consent for “education, general research, and/or to further innovative technologies.” This project used anonymous, optional surveys. No identifiable or sensitive personal data were collected, and all responses were stored without linkage to participants. Informed consent was collected via the cover page of the survey with approved study details, contact information for the University of Kentucky Office of Research Integrity, and a final question asking, “Do you consent to your anonymized survey data being used in future research publications or other research purposes?” Participants received no compensation for taking part in the survey.

Curricular Context

We implemented this educational resource as a required component of the first-year gross anatomy courses at the University of Kentucky, serving learners in the MD, DPT, and PA programs. During the course laboratory days, the medical

students rotate between physical examination with point-of-care ultrasound, prosection in the gross anatomy laboratory, and radiology-focused integrated clinical case topic sessions. This manuscript only considers the gross anatomy component of the laboratory experiences. All medical students participated in donor-based laboratory sessions using prosected donors, while PA and DPT students used an alternating dissection model (half dissected one day; the other half dissected the opposite day, with peer teaching at the end of each session—a long-used dissection model) [29]. We specifically designed the Anatomy Interactives modules to supplement these in-person experiences. Students had no prerequisite training in anatomy beyond the concurrent lectures offered within the course. Facilitators (anatomy instructors) had no specific training on the Anatomy Interactives.

Development of Anatomy Interactives

Overview

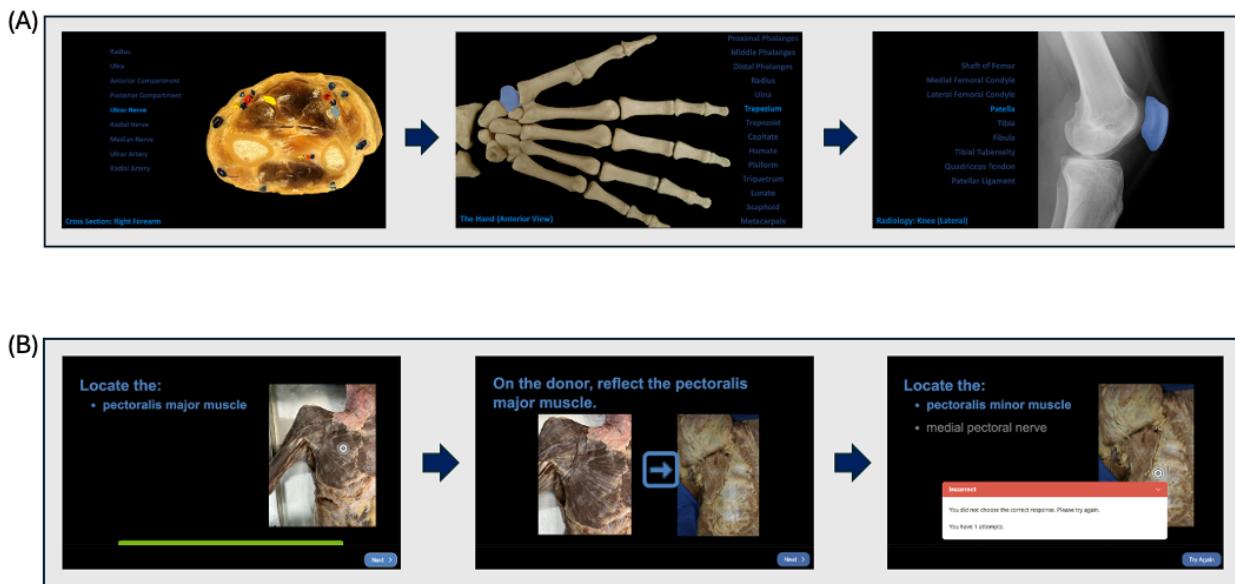
We created the Anatomy Interactives website to function as a digital laboratory manual, aiming to enhance anatomical structure identification and spatial reasoning through the use of prosected donor images.

Our primary goal was to provide a guided, active learning experience that reinforced anatomical concepts outside of the traditional laboratory setting. We executed the development process in 2 distinct phases, representing the type of materials that we created.

Phase 1: Foundational Content Module Development

In this phase, we generated labeled anatomical images from diverse modalities, including arteriograms, cross-sections, computed tomography scans, radiographs, and osteological photographs (Figure 1A). We embedded these images with hover-over identification features, where selecting an anatomical term would highlight the corresponding structure on the image, and vice versa. We initially built this content in Microsoft PowerPoint and subsequently exported it to HTML format using iSpring Suite 11 to ensure web accessibility.

Figure 1. Sample view of Anatomy Interactives Website. (A) Examples of fixed labeled images from the hover-over identification resource. Screenshots from cross-sectional, osteology, and radiology image sets showing labeled anatomical structures. Structures and corresponding names are highlighted for visual reinforcement of anatomical terminology. (B) Example Anatomy Interactives module interface: pectoral muscle identification task. Screenshot from the pectoral region module showing a pin-on-image question asking students to identify a muscular structure. Immediate feedback is provided upon incorrect identification, with instructional guidance on adjusting the donor to reveal the underlying structure.



Phase 2: Prosection Laboratory Module Development

In the second phase, we focused on developing interactive, donor-based modules (Figure 1B). We dedicated each module to a specific anatomical subregion (eg, foot and thigh) and incorporated a series of image-based questions, such as pin-on-image tasks, multiple-choice questions, and check-all-that-apply questions. We augmented these with pop-up content designed to provide relevant clinical and conceptual context. Each module culminated in a score based on accuracy, offering students formative feedback on their understanding. We then grouped multiple modules to comprise a regional “laboratory” based on curricular structure. We developed six laboratories: (1) back and upper limb, (2) lower limb, (3) thorax, (4) abdomen, (5) pelvis, and (6) head and neck.

Implementation

Overview

At the beginning of the semester, we oriented students to the Anatomy Interactives platform via an in-class demonstration. We hosted the modules on the institution’s learning management system, and they were available on demand throughout the course. We suggested students use the modules before, during, and after their scheduled laboratory sessions to reinforce learning. During the laboratory time for the MD students, the main exercise was for them to complete the module with their group (approximately 8 students per group, with 2 donors). However, we did not mandate completion of the modules for course credit for any student (MD, PA, and DPT). Students used these modules using university-sponsored iPads and any personal devices that could access the internet. The modules were all made available to students from the start of their gross anatomy course, allowing them to integrate both types of

resources into prelaboratory preparation and ongoing review throughout the semester.

MD Students

Students in the MD program had 80-minute gross anatomy prosection laboratory sessions 6 weeks out of the 9-week course. MD students collaborated with their table group to use the digital Prosection Laboratory Modules as they worked with their donor. Students answered multiple-choice questions or select-all-that-apply questions as a team, and compared the location of the item on the hotspot questions to the anatomy of their group’s specific donor. We made the Foundational Content Modules available for students to self-study on their own time to prepare for assessments. The fall 2024 cohort of MD students had access to the Anatomy Interactives modules as part of their standard curriculum. The fall 2023 cohort of MD students, serving as a comparison group, completed the same gross anatomy course without access to these modules. Both cohorts completed the course with the same professors with nearly identical assessments. Both groups had full-time access to the anatomy laboratory and participated in the same number of small-group laboratory sessions, the only difference being the use of the modules.

PA/DPT students

In spring 2024, we provided PA/DPT students access during their shared dissection-based gross anatomy laboratory course. The format of this course consisted of 2 days of 2-hour lectures followed by 2-hour laboratories per week for 16 weeks. We encouraged students to use the Foundational Content Modules and Prosection Laboratory Modules as self-study to prepare for laboratories and course assessments.

Assessment and Evaluation Strategy

Overview

We used a mixed-methods approach to evaluate the usage and effectiveness of the Anatomy Interactives platform. This strategy included student surveys, website usage data, and measures of academic performance.

Surveys

Surveys were distributed at the conclusion of the respective course to all PA/DPT students and MD students who had access to the modules. The surveys aimed to assess perceived utility, engagement, and the learning impact of the platform. Surveys included questions including which modules were used, time spent using the modules, time spent in the laboratory, opinions regarding score correlation with module usage, opinions regarding preparation level and module usage, study purposes when using the modules, and use of modules with the donor or in a stand-alone study. We collected open-ended responses pertaining to likes, dislikes, and recommendations for the modules and analyzed them thematically.

Website Analytics

We collected web traffic data from Google Analytics. These included metrics such as daily active users and usage patterns, particularly around examination periods and laboratory sessions, to understand engagement trends.

Academic Performance

We obtained written and practical anatomy examination scores from the Office of Medical Education (determined as not human research per the University of Kentucky IRB) for 2 MD student cohorts: the 2023 cohort (N=206) without access to the modules, and the 2024 cohort (N=205) with access. To ensure baseline academic comparability between student cohorts, we compared Medical College Admission Test (MCAT) scores, overall undergraduate grade point averages (uGPA), and science undergraduate grade point averages (sGPA) for both MD cohorts

to test for differences. We then conducted a comparison of examination performance between these 2 MD cohorts to assess the educational impact of the Anatomy Interactives modules. We first tested for normality of the data and determined that our data was not normally distributed. Thus, we analyzed significance using an unpaired Mann-Whitney *U* test with statistical significance set at $P<.05$ and assessed effect size by calculating the rank-biserial correlation coefficient (r_{rb}). Some examination scores were not available in the dataset, and therefore, the number of participants changes per examination; this was accounted for in the data analysis. Statistical analyses were conducted in GraphPad Prism (version 10.6.1; GraphPad Software, LLC). r_{rb} was manually calculated using the simplified formula as previously described [30].

Results

Analysis of the student surveys (N=31 for PA/DPT and N=116 for MD) and website trafficking data provided significant insight into the usage of the modules by students.

For the PA/DPT cohort, 23 out of 31 (74%) students reported using the modules for 1-2 hours per week in addition to the 5-6 hours spent with the donor in allotted laboratory time. Of the medical students surveyed, 85 out of 116 (73%) students report using the laboratory modules 1-2 hours per week. Regarding the nature of their usage, 22 out of 31 (70%) PA/DPT students and 79 out of 116 (68%) medical students used the modules both with the donor and as a stand-alone study tool (Figure 2). The website data indicate the modules are heavily trafficked leading up to the examination, the number of users increasing drastically in the 1-3 days leading up to an examination for PA/DPT and medical students (Figure 2). Module page traffic also increases on laboratory days for medical students, as well as the day before and the day after a laboratory session (Figure 3). The laboratory modules had much higher traffic than the hover-over labeled images.

Figure 2. Student self-report of module usage type. Percentage of surveyed MD students (n=116) reporting use of the Anatomy Interactives modules in each context: with the donor during laboratory sessions, as a stand-alone study tool, or both. Percentage of surveyed PA/DPT students (n=31) reporting use of the modules in each context.

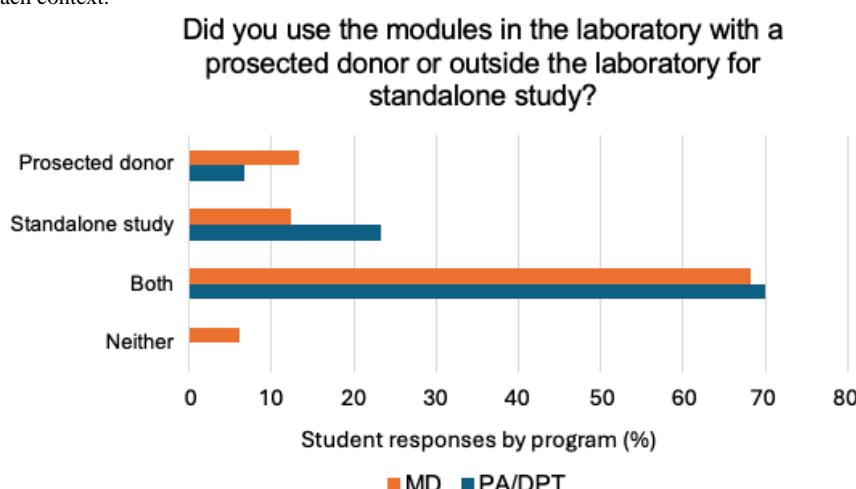
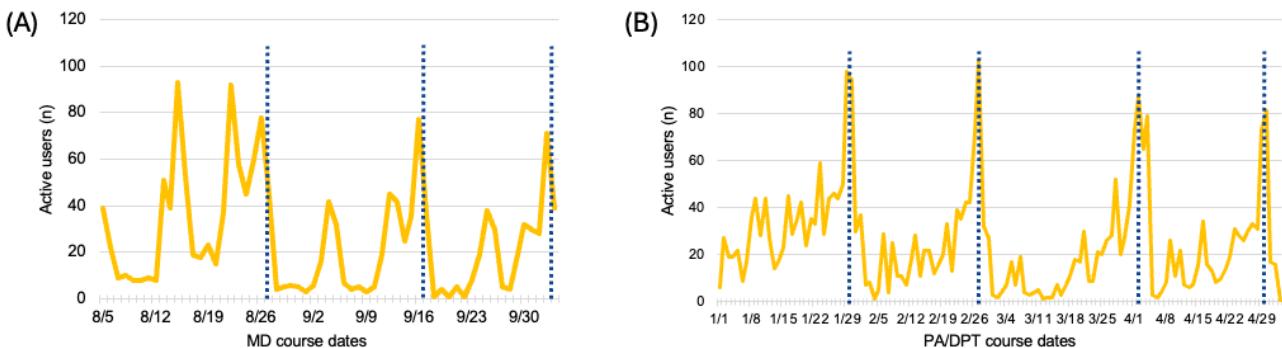


Figure 3. Daily users of the Anatomy Interactives platform for MD and PA/DPT courses. (A) Number of MD student users per day during the MD gross anatomy course (August 5, 2024, to October 24, 2024). (B) Number of PA/DPT student users per day during the PA/DPT gross anatomy course (January 2, 2024, to April 30, 2024). Spikes correspond to laboratory sessions and examination periods. Vertical dashed lines indicate examination dates.



When asked about the relationship between their module usage and examination score, 23 out of 31 (74%) PA/DPT students reported that the modules helped them on their written examination, and 26 out of 31 (84%) believed the modules helped them on their laboratory examination. The PA/DPT cohort felt more prepared for the examination when they had a laboratory module to support the curriculum. The medical student survey revealed 80 out of 116 (69%) of that cohort believed their assessment scores reflected the time and effort they spent on modules.

To examine the modules' impact, examination scores from 2 different cohorts were compared (Table 1). First-year medical students in 2023 did not use the modules as part of their laboratory donor experience, while the medical students in 2024

did have access to the new resource. For all 3 measures we used to test comparability (MCAT scores, uGPA, and sGPA), the data were not normally distributed (Shapiro-Wilk test, $P<.001$) and not significantly different (Mann-Whitney U , $P<.001$) between the 2 MD cohorts. All MD cohort examination score data failed normality testing (Shapiro-Wilk test, $P<.001$), so groups were compared using nonparametric testing. Briefly, Mann-Whitney U tests were used to test for cohort examination score differences, and r_{rb} values were calculated and used to assess effect size. Table 1 summarizes medians, P values, Mann-Whitney U , and the r_{rb} values from these tests. Students in the 2024 cohort scored higher on both the written and practical portions across examination 1 and examination 3, with moderate effect sizes for the written portions and total examination scores.

Table 1. Comparison of anatomy examination scores between MD cohorts with and without module access. Median written, laboratory, and total scores for 3 examinations in the 2023 cohort (no module use) and the 2024 cohort (module use), with P values, Mann-Whitney U , and rank-biserial correlations.

	2023 MD cohort, median (IQR), %	2024 MD cohort, median (IQR), %	P value	Mann-Whitney U	Rank-biserial correlation
Examination 1 medians					
Written portion	82 (74-88)	86 (79-94)	<.001	15,130	.28
Laboratory portion	90 (85-95.5)	92.5 (85-97.5)	.01	18,153	.14
Total	86.1 (78.9-90)	88.9 (82.2-94.4)	<.001	15,778	.25
Examination 2 medians					
Written portion	83 (76-90)	84 (78-90)	.13	19,201	.09
Laboratory portion	92.5 (87.5-95)	92.5 (87.5-95)	.39	19,984	.05
Total	87.8 (81.1-91.1)	87.8 (82.2-92.2)	.16	19,327	.08
Examination 3 medians					
Written portion	85 (79.5-92)	92 (84-96)	<.001	12,420	.41
Laboratory portion	92.5 (87.5-95)	92.5 (87.5-97.5)	.49	20,092	.04
Total	87.8 (83.3-92.2)	92.2 (87.5-97.5)	<.001	14,645	.30

When asked about the purpose of their module usage, 27 PA/DPT students used the modules for general studying, 8 before laboratory for preparation, 24 on laboratory day for preparation, 23 for understanding current content, and 25 for targeted examination review. A review of medical students' qualitative feedback on the modules revealed students appreciated the accessibility of the modules, reinforcement of

course content, engagement provided by the active-learning style, and the complementary nature of the modules and the prospected donor experience. Accessibility and flexibility were mentioned as advantages by 43 students, with 27 stating an appreciation for the convenience and ease of using the modules on their own time, and 16 mentioning they were helpful during independent examination preparation. The complementary

nature of the modules to the prosected donor experience was highlighted by 19 students who detailed the usefulness of having the resource available during the laboratory session. Furthermore, 41 students mentioned liking how the modules support course content and reinforce lecture material. Technological characteristics were mentioned by many students, with 29 students giving positive feedback on the technology and user interface of the modules, while 48 students thought it was an area for improvement.

Discussion

Principal Findings

The Anatomy Interactives digital laboratory manual achieved its intended goal of supplementing the prosection-based gross anatomy curriculum with an interactive, learner-centered resource that promoted structure recognition, spatial understanding, and self-directed learning. Moving toward a student-centered learning approach in anatomy may have positive effects on student performance [31]. While learners report that student-centered approaches promote their independent learning and peer communication, they perceive that faculty-led learning increases their anatomical knowledge [32]. Before the implementation of the modules at our institution, the prosection laboratory had become a completely passive faculty-led environment with minimal student participation. By offering a digitally guided experience aligned with the prosected donor material, this innovation supported student engagement and potentially improved academic performance across multiple learner groups.

Both MD and PA/DPT learners found the modules helpful in preparing for examinations and reinforcing course content. Website analytics corroborated student usage of the modules as a study material outside of class, as website usage spiked before laboratory sessions and assessments. This suggests that the modules were integrated into learners' study routines in a timely and meaningful way. It is still uncertain whether engagement with various digital media in anatomical education correlates to improved student outcomes and satisfaction with the learning experience [33]. However, the goal of implementing the digital modules during laboratory time was to promote engagement as a means to foster a constructivist learning environment [34]. That is to say, during the laboratory, the modules were a student-centered method to have learners interact with the donors and the course content in a meaningful way, such that they could develop or "construct" their own knowledge base as it related to the gross anatomical structures. The outcomes observed suggest that we successfully addressed the initial challenge of providing anatomy learners with a resource that bridges traditional prosection-based instruction and flexible, learner-driven study to enhance both student satisfaction and academic performance.

Most notably, medical students who used the modules in 2024 outperformed the previous cohort (2023) on both written and practical assessments, even after controlling for prior academic performance (ie, MCAT scores, uGPA, and sGPA), indicating a potential causal relationship between module usage and examination performance. While the use of MCAT scores is an

imperfect means of normalizing between cohorts, it is a method that has been used by similar studies [9]. More work is warranted to assess whether the use of the digital laboratory modules directly improves course performance compared to baseline (faculty-led prosection session with no digital modules).

Recent work has drawn light to the importance of educational design that creates flourishing medical and health sciences learners and practitioners [35]. Independent of developing student knowledge recall, the anatomy laboratory offers a unique environment in which to promote the holistic development of future clinicians [36,37]. In fact, with careful course design, the anatomy laboratory can be harnessed as a means to promote the development of flourishing physicians and health care providers [38]. The Anatomy Interactives laboratory modules could be enhanced in future iterations to encourage learners to embrace a sense of meaning and belonging in line with these goals.

Limitations

This educational resource was implemented and evaluated within a single institution, limiting the generalizability of the results. The anatomy curriculum and access to prosected donors may differ at other institutions, affecting feasibility and relevance. While the resource is available for use in any program, the module progression is modeled after the University of Kentucky's gross anatomy curriculum and may not directly follow along the curriculum of other institutions. To access the resource, students simply need a device that connects to the internet. To implement the use of the modules in a digital-donor laboratory experience, institutions must have access to prosected donors and a space to conduct a laboratory session, likely with course director oversight.

Additionally, the evaluation relied on student self-reports and examination scores as primary metrics, which, while informative, do not fully capture long-term knowledge retention, clinical application, or behavioral change. The module usage and examination data were examined on a cohort level, rather than correlated on an individual level, which limits the interpretation of their association. Furthermore, although MCAT scores, uGPA, and sGPA were used to control for incoming academic aptitude, unmeasured variables such as instructor differences, schedule differences, or peer learning dynamics could have influenced performance outcomes.

Future Directions

During implementation, we found that introducing the resource early in the course and orienting students to its features increased usage. Embedding the modules via a link from the learning management system allowed easy access and made integration into student study workflows more seamless. However, we also learned that variability in student technology preferences and learning styles may affect engagement; while some praised the interactivity of the modules, others noted opportunities for improvement in interface design and navigation.

Future iterations of the Anatomy Interactives platform will aim to improve user interface design, enhance accessibility across devices, and incorporate more clinically relevant content, such as radiographic correlations and pathology examples. Broader implementation at multiple institutions would help test

generalizability and effectiveness across diverse learner populations and curricula. Additionally, incorporating longitudinal tracking and correlating module usage with clinical anatomy performance in later coursework or clerkships could strengthen claims about its educational impact.

On a broader scale, this work supports ongoing shifts in anatomy education toward multimodal, learner-driven resources. The findings may inform curricular policy regarding the balance between traditional laboratory experiences and flexible digital tools, particularly as programs navigate logistical constraints, student wellness, and educational equity.

Acknowledgments

The authors would like to acknowledge the donors who gifted their bodies to the Willed Body Program at the University of Kentucky. Generative AI was not used in any part of producing this manuscript.

Funding

The article processing fee (APP) was paid by the University of Kentucky Department of Neuroscience Research Enhancement Award (KP). The funder had no involvement in this study's design, data collection, analysis, interpretation, or the writing of this paper.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to concerns of student educational data protection, but are available in an anonymous format from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: CS (equal), SF (equal), SLC (equal), KP (equal)
Data curation: CS (equal), SLC (equal), KP (equal)
Formal analysis: CS (lead), SLC (lead), KP (supporting)
Methodology: CS (equal), SF (equal), SLC (equal), KP (equal)
Writing – original draft: CS (equal), SLC (equal), KP (equal)
Writing – review & editing: CS (equal), SLC (equal), KP (equal)

Conflicts of Interest

None declared.

References

1. Wilson AB, Miller CH, Klein BA, Taylor MA, Goodwin M, Boyle EK, et al. A meta-analysis of anatomy laboratory pedagogies. *Clin Anat* 2018;31(1):122-133. [doi: [10.1002/ca.22934](https://doi.org/10.1002/ca.22934)] [Medline: [28612403](https://pubmed.ncbi.nlm.nih.gov/28612403/)]
2. Yeager VL. Learning gross anatomy: dissection and prosection. *Clin Anat* 1996;9(1):57-59. [doi: [10.1002/\(SICI\)1098-2353\(1996\)9:1<57::AID-CA12>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1098-2353(1996)9:1<57::AID-CA12>3.0.CO;2-9)] [Medline: [8838283](https://pubmed.ncbi.nlm.nih.gov/8838283/)]
3. McWatt SC, Newton GS, Umphrey GJ, Jadeski LC. Dissection versus prosection: a comparative assessment of the course experiences, approaches to learning, and academic performance of non-medical undergraduate students in human anatomy. *Anat Sci Educ* 2021;14(2):184-200. [doi: [10.1002/ase.1993](https://doi.org/10.1002/ase.1993)] [Medline: [32539226](https://pubmed.ncbi.nlm.nih.gov/32539226/)]
4. Quek FF. Revolutionising anatomy education: the current role of digital technologies in enhancing anatomy learning for undergraduate medical students. *Cureus* 2024;16(12):e75919. [doi: [10.7759/cureus.75919](https://doi.org/10.7759/cureus.75919)] [Medline: [39711926](https://pubmed.ncbi.nlm.nih.gov/39711926/)]
5. McMenamin PG, McLachlan J, Wilson A, McBride JM, Pickering J, Evans DJR, et al. Do we really need cadavers anymore to learn anatomy in undergraduate medicine? *Med Teach* 2018;40(10):1020-1029. [doi: [10.1080/0142159X.2018.1485884](https://doi.org/10.1080/0142159X.2018.1485884)] [Medline: [30265177](https://pubmed.ncbi.nlm.nih.gov/30265177/)]
6. Lackey-Cornelison WL, Bauler LD, Smith J. A comparison of the effectiveness of dissection and prosection on short-term anatomic knowledge retention in a reciprocal peer-teaching program. *Adv Physiol Educ* 2020;44(2):239-246. [doi: [10.1152/advan.00031.2020](https://doi.org/10.1152/advan.00031.2020)] [Medline: [32412383](https://pubmed.ncbi.nlm.nih.gov/32412383/)]
7. Schurr AF, Burg BJ, Dickinson E, Granatosky MC. No cuts, no buts: satisfaction of first-year medical students with a hybrid prosection-based model for learning gross anatomy during the COVID-19 pandemic. *Anat Sci Educ* 2022;15(5):827-838 [FREE Full text] [doi: [10.1002/ase.2205](https://doi.org/10.1002/ase.2205)] [Medline: [35726438](https://pubmed.ncbi.nlm.nih.gov/35726438/)]
8. Clement SM, Ubben TA, Yates DT. Cadaveric prosections prepared by qualified instructional staff were more efficient and effective teaching modalities for veterinary gross anatomy than in-class dissections by students. *J Vet Med Educ* 2024;51(5):593-609. [doi: [10.3138/jvme-2024-0031](https://doi.org/10.3138/jvme-2024-0031)] [Medline: [39499889](https://pubmed.ncbi.nlm.nih.gov/39499889/)]

9. Coker C, Rogers RS, Freed B, Steele R, Kinde MN, Danilova G, et al. Battle of the sections: student outcomes and course feedback support combined prosection and dissection laboratory formats to maximize student success. *Anat Sci Educ* 2024;17(5):1012-1025. [doi: [10.1002/ase.2420](https://doi.org/10.1002/ase.2420)] [Medline: [38570916](https://pubmed.ncbi.nlm.nih.gov/38570916/)]
10. Hussey D, Shaw AV, Brian PL, Lazarus MD. Learning head and neck anatomy through a radiological imaging platform. *MedEdPORTAL* 2022;18:11230 [FREE Full text] [doi: [10.15766/mep_2374-8265.11230](https://doi.org/10.15766/mep_2374-8265.11230)] [Medline: [35342790](https://pubmed.ncbi.nlm.nih.gov/35342790/)]
11. Marsland MJ, Tomic D, Brian PL, Lazarus MD. Pelvis anatomy tutorial using radiological images. *MedEdPORTAL* 2018;14:10778 [FREE Full text] [doi: [10.15766/mep_2374-8265.10778](https://doi.org/10.15766/mep_2374-8265.10778)] [Medline: [30800978](https://pubmed.ncbi.nlm.nih.gov/30800978/)]
12. Lack CM, Pena J, Grubb EXL, Shen E, Hiatt KD, Benayoun MD, et al. Cranial nerve anatomy using a modular and multimodal radiologic approach. *MedEdPORTAL* 2022;18:11261 [FREE Full text] [doi: [10.15766/mep_2374-8265.11261](https://doi.org/10.15766/mep_2374-8265.11261)] [Medline: [35720637](https://pubmed.ncbi.nlm.nih.gov/35720637/)]
13. So S, Patel RM, Orebaugh SL. Ultrasound imaging in medical student education: impact on learning anatomy and physical diagnosis. *Anat Sci Educ* 2017;10(2):176-189. [doi: [10.1002/ase.1630](https://doi.org/10.1002/ase.1630)] [Medline: [27286419](https://pubmed.ncbi.nlm.nih.gov/27286419/)]
14. Bell FE, Neuffer FH, Haddad TR, Epps JC, Kozik ME, Warren BC. Active learning of the floor of mouth anatomy with ultrasound. *Anat Sci Educ* 2019;12(3):310-316. [doi: [10.1002/ase.1839](https://doi.org/10.1002/ase.1839)] [Medline: [30414266](https://pubmed.ncbi.nlm.nih.gov/30414266/)]
15. McBain KA, Habib R, Laggis G, Quaiattini A, M Ventura N, Noel GPJC. Scoping review: the use of augmented reality in clinical anatomical education and its assessment tools. *Anat Sci Educ* 2022;15(4):765-796. [doi: [10.1002/ase.2155](https://doi.org/10.1002/ase.2155)] [Medline: [34800073](https://pubmed.ncbi.nlm.nih.gov/34800073/)]
16. Odogwu TS, Mohamed EAH, Mishu L, Umahi I. Effect of virtual reality simulation on anatomy learning outcomes: a systematic review. *Cureus* 2025;17(4):e81893. [doi: [10.7759/cureus.81893](https://doi.org/10.7759/cureus.81893)] [Medline: [40342485](https://pubmed.ncbi.nlm.nih.gov/40342485/)]
17. Attardi SM, Harmon DJ, Barremkala M, Bentley D, Brown KM, Dennis JF, et al. An analysis of anatomy education before and during COVID-19: August-December 2020. *Anat Sci Educ* 2022;15(1):5-26 [FREE Full text] [doi: [10.1002/ase.2152](https://doi.org/10.1002/ase.2152)] [Medline: [34758206](https://pubmed.ncbi.nlm.nih.gov/34758206/)]
18. Al-Redouan A, Dudin A, Urbanek AJ, Olsson E, Kachlik D. Visible human project based applications can prompt integrating cross-sectional anatomy into the medical school curriculum when combined with radiological modalities: a three-year cross-sectional observational study. *Ann Anat* 2025;257:152357. [doi: [10.1016/j.aanat.2024.152357](https://doi.org/10.1016/j.aanat.2024.152357)] [Medline: [39577816](https://pubmed.ncbi.nlm.nih.gov/39577816/)]
19. Ackerman MJ. The visible human project. *Stud Health Technol Inform* 2022;288:134-140. [doi: [10.3233/SHTI210988](https://doi.org/10.3233/SHTI210988)] [Medline: [35102835](https://pubmed.ncbi.nlm.nih.gov/35102835/)]
20. Darras KE, Spouge R, Hatala R, Nicolaou S, Hu J, Worthington A, et al. Integrated virtual and cadaveric dissection laboratories enhance first year medical students' anatomy experience: a pilot study. *BMC Med Educ* 2019;19(1):366 [FREE Full text] [doi: [10.1186/s12909-019-1806-5](https://doi.org/10.1186/s12909-019-1806-5)] [Medline: [31590672](https://pubmed.ncbi.nlm.nih.gov/31590672/)]
21. Kavvadia E, Katsoula I, Angelis S, Filippou D. The anatomage table: a promising alternative in anatomy education. *Cureus* 2023;15(8):e43047 [FREE Full text] [doi: [10.7759/cureus.43047](https://doi.org/10.7759/cureus.43047)] [Medline: [37692592](https://pubmed.ncbi.nlm.nih.gov/37692592/)]
22. Alsup K, Fox G. University of Michigan BlueLink Photographic Atlas. 2014. URL: <https://sites.google.com/a/umich.edu/bluelink/resources/bluelink> [accessed 2025-12-13]
23. Gest TR. Texas Tech University Health Sciences Center Gross Anatomy Home. 2000. URL: <https://anatomy.elpaso.ttuhs.edu/index.html> [accessed 2012-12-13]
24. Mayer RE. Using multimedia for e - learning. *J Comput Assisted Learn* 2017;33(5):403-423 [FREE Full text] [doi: [10.1111/jcal.12197](https://doi.org/10.1111/jcal.12197)]
25. Mota MF, Mata FRD, Aversi-Ferreira TA. Constructivist pedagogic method used in the teaching of human anatomy. *Int J Morphol* 2010;28(2):369-374. [doi: [10.4067/s0717-95022010000200005](https://doi.org/10.4067/s0717-95022010000200005)]
26. Hermiz DJ, O'Sullivan DJ, Lujan HL, DiCarlo SE. Constructivist learning of anatomy: gaining knowledge by creating anatomical casts. *Anat Sci Educ* 2011;4(2):98-104. [doi: [10.1002/ase.206](https://doi.org/10.1002/ase.206)] [Medline: [21337719](https://pubmed.ncbi.nlm.nih.gov/21337719/)]
27. Rodenbaugh DW, Lujan HL, DiCarlo SE. Learning by doing: construction and manipulation of a skeletal muscle model during lecture. *Adv Physiol Educ* 2012;36(4):302-306 [FREE Full text] [doi: [10.1152/advan.00093.2012](https://doi.org/10.1152/advan.00093.2012)] [Medline: [23209011](https://pubmed.ncbi.nlm.nih.gov/23209011/)]
28. Platt K, Franklin S, Case S. University of Kentucky College of Medicine Anatomy Interactives. 2021. URL: <https://medapp.mc.uky.edu/anatomyinteractives> [accessed 2025-12-11]
29. Brueckner JK, MacPherson BR. Benefits from peer teaching in the dental gross anatomy laboratory. *Eur J Dent Educ* 2004;8(2):72-77. [doi: [10.1111/j.1600-0579.2003.00333.x](https://doi.org/10.1111/j.1600-0579.2003.00333.x)] [Medline: [15059083](https://pubmed.ncbi.nlm.nih.gov/15059083/)]
30. Wendt H. Dealing with a common problem in social science: a simplified rank - biserial coefficient of correlation based on the U statistic. *Euro J Social Psych* 1972;2(4):463-465 [FREE Full text] [doi: [10.1002/ejsp.2420020412](https://doi.org/10.1002/ejsp.2420020412)]
31. Wilson AB, Brown KM, Misch J, Miller CH, Klein BA, Taylor MA, et al. Breaking with tradition: a scoping meta-analysis analyzing the effects of student-centered learning and computer-aided instruction on student performance in anatomy. *Anat Sci Educ* 2019;12(1):61-73. [doi: [10.1002/ase.1789](https://doi.org/10.1002/ase.1789)] [Medline: [29659155](https://pubmed.ncbi.nlm.nih.gov/29659155/)]
32. Abdullah E, Lone M, Balta JY. Student-centered learning in the anatomy laboratory: medical students' perspective. *Med Sci Educ* 2020;30(4):1459-1464 [FREE Full text] [doi: [10.1007/s40670-020-01094-w](https://doi.org/10.1007/s40670-020-01094-w)] [Medline: [34457813](https://pubmed.ncbi.nlm.nih.gov/34457813/)]
33. Pickering JD, Swinnerton BJ. Exploring the dimensions of medical student engagement with technology-enhanced learning resources and assessing the impact on assessment outcomes. *Anat Sci Educ* 2019;12(2):117-128 [FREE Full text] [doi: [10.1002/ase.1810](https://doi.org/10.1002/ase.1810)] [Medline: [30168911](https://pubmed.ncbi.nlm.nih.gov/30168911/)]

34. Mukhalalati BA, Taylor A. Adult learning theories in context: a quick guide for healthcare professional educators. *J Med Educ Curric Dev* 2019;6:2382120519840332 [FREE Full text] [doi: [10.1177/2382120519840332](https://doi.org/10.1177/2382120519840332)] [Medline: [31008257](https://pubmed.ncbi.nlm.nih.gov/31008257/)]
35. Maurana CA, Fritz JD, Witten AA, Williams SE, Ellefson KA. Advancing flourishing as the north star of medical education: a call for personal and professional development as key to becoming physicians. *Med Teach* 2024;46(12):1539-1543. [doi: [10.1080/0142159X.2024.2412791](https://doi.org/10.1080/0142159X.2024.2412791)] [Medline: [39481004](https://pubmed.ncbi.nlm.nih.gov/39481004/)]
36. Ghosh SK. Cadaveric dissection as an educational tool for anatomical sciences in the 21st century. *Anat Sci Educ* 2017;10(3):286-299. [doi: [10.1002/ase.1649](https://doi.org/10.1002/ase.1649)] [Medline: [27574911](https://pubmed.ncbi.nlm.nih.gov/27574911/)]
37. Čavka M, Petanjek Z, Zanze Beader M, Mathew T, Brassett C, Nagy AD, et al. Balancing tradition and innovation: rethinking the dichotomy in anatomy teaching. *Croat Med J* 2025;66(1):62-70 [FREE Full text] [doi: [10.3325/cmj.2025.66.62](https://doi.org/10.3325/cmj.2025.66.62)] [Medline: [40047163](https://pubmed.ncbi.nlm.nih.gov/40047163/)]
38. Kranzlein J, Sharpe B, Robinson M, Burbank E, Pearson WG. Repurposing the anatomy lab: a developmental model for physician formation and human flourishing. *Anat Sci Educ* 2025;18(9):897-904. [doi: [10.1002/ase.70103](https://doi.org/10.1002/ase.70103)] [Medline: [40717029](https://pubmed.ncbi.nlm.nih.gov/40717029/)]

Abbreviations

IRB: Institutional Review Board

MCAT: Medical College Admission Test

sGPA: science undergraduate grade point average

uGPA: undergraduate grade point average

Edited by J Gentges; submitted 29.Sep.2025; peer-reviewed by M Pascoe, A Al-Redouan; comments to author 03.Nov.2025; accepted 08.Dec.2025; published 13.Jan.2026.

Please cite as:

Sumner C, Case SL, Franklin S, Platt K

Interactive, Image-Based Modules as a Complement to Prosection-Based Anatomy Laboratories: Multicohort Evaluation

JMIR Med Educ 2026;12:e85028

URL: <https://mededu.jmir.org/2026/1/e85028>

doi: [10.2196/85028](https://doi.org/10.2196/85028)

PMID:

©Caroline Sumner, Sami L Case, Samuel Franklin, Kristen Platt. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 13.Jan.2026. 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.

Original Paper

Disaster Medicine Training for Medical Students in Lebanon: Quasi-Experimental Comparison of e-Learning and Face-to-Face Modalities

Ali Msheik^{1,2}, MD, MSc; Ruben Peralta^{3,4}, MD, MAMSE; Zeinab Al Mokdad⁵, MSc; Christine Bartulec⁶, MD; Linda Abou-Abbas^{6,7}, PhD; Dima Anani¹, MD; Hussein Al Mawla¹, MD; Hassan Salame⁸, MD; Mohamad Moussa¹, MD; Awsan Bahattab⁹, MD, PhD

¹Faculty of Medical Sciences, Lebanese University, Beirut, Beyrouth, Lebanon

²Neurological Surgery, Neuroscience Institute, Hamad Medical Corporation, Doha, Qatar

³Trauma Surgery, Hamad Medical Corporation, Doha, Qatar

⁴Department of Surgery, Universidad Nacional Pedro Henríquez Ureña, Santo Domingo, Dominican Republic

⁵Medical Ethics, Faculty of Medical Sciences, Lebanese University, Beirut, Lebanon

⁶International Committee of the Red Cross, Beirut, Lebanon

⁷Institut National de Santé Publique, Épidémiologie Clinique et Toxicologie-Liban, Beirut, Lebanon

⁸Head of the Program Evaluation Committee, Faculty of Medical Sciences, Lebanese University, Beirut, Lebanon

⁹Center for Research and Training in Disaster Medicine, Università degli Studi del Piemonte Orientale “Amedeo Avogadro”, Piemonte, Italy

Corresponding Author:

Ali Msheik, MD, MSc

Neurological Surgery

Neuroscience Institute

Hamad Medical Corporation

Doha

3050

Qatar

Phone: 974 72157731

Email: dr.alimsheik@gmail.com

Abstract

Background: Despite global advocacy for its integration into medical curricula, disaster medicine (DM) education remains underdeveloped, especially in fragile settings where such training is urgently needed. In Lebanon, a country facing political and economic crises, students face significant barriers to in-person education.

Objective: This study aims to evaluate the effectiveness of e-learning versus face-to-face (F2F) approaches in improving knowledge retention and provides insight into the practical considerations of implementing DM courses in such settings.

Methods: This quasi-experimental study used a Solomon 4-group design to evaluate e-learning and F2F DM courses for second-to fifth-year medical students at the Lebanese University. A total of 205 participants, stratified by academic year, were divided between the 2 modalities. Knowledge was assessed before the course, after the course, and at 1-month follow-up. Confidence, competency, and satisfaction were evaluated after the course using validated tools.

Results: Of 205 participants, 56.6% (n=116) favored e-learning. Both modalities improved knowledge and knowledge retention, with no significant difference between the 2 groups. Fifth-year students achieved the highest gains in knowledge, particularly in the e-learning group. Similarly, no significant difference in satisfaction was observed across modalities, although F2F was preferred overall, except among fifth-year students, who preferred e-learning. Confidence levels were also similar across both modalities, but F2F scored higher for skills like triaging. Feedback emphasized the relevance of the course and advocated for integration of DM into the medical curriculum, and adding practical sessions.

Conclusions: Integrating DM education into the fifth-year curriculum, prior to externship, can enhance preparedness and promote knowledge retention and application in real-world disaster settings. The study shows that e-learning is an effective modality for improving knowledge acquisition and retention in DM. Although feasibility and efficiency were not measured directly, the successful delivery of the course to geographically dispersed students suggests potential practical advantages. Combining F2F

practical sessions for specific targeted topics in a blended curriculum is recommended to further enhance medical students' confidence for future disaster response. These findings support broader policy efforts to institutionalize DM in medical curricula, particularly in fragile and resource-limited settings.

(*JMIR Med Educ* 2026;12:e80409) doi:10.2196/80409

KEYWORDS

disaster medicine; education, medical, undergraduate; educational technology; teaching methods; program evaluation; knowledge; surveys and questionnaires; self-efficacy; Lebanon

Introduction

The need for disaster medicine (DM) training in medical education has been recognized since the 1970s, but its integration into the curricula has remained limited until the early 2000s, when DM was incorporated into the curricula of US medical schools. Since then, international frameworks and associations such as the Sendai Framework for Disaster Risk Reduction [1], and the Association of American Medical Colleges and the World Association for Disaster and Emergency Medicine have called for the incorporation of DM and mass casualty incident education in medical schools using scientific evidence-based and flexible approaches [2,3]. Despite these recommendations, limited efforts have been made by universities to implement such training [2].

The lack of DM education has left medical students unprepared to participate in disaster response [4,5]. In contrast, those who have participated in elective DM courses have reported positive results, with high satisfaction, and improved understanding, knowledge, attitudes, and skills toward disaster response [2,6-9]. Still, evaluating these courses remains complicated, due to a lack of standardization across curricula and assessment methods [10,11]. Furthermore, the availability of such training is concentrated in high-income countries and Global North institutes, with limited accessibility for global participation, particularly in fragile and conflict-affected countries, which disproportionately experience higher frequency and impact of disasters [12,13].

Lebanon, home to 5.4 million people and hosting over a million refugees, presents an urgent need for DM education. The country is facing repeated manmade and public health disasters compounded by political instability and an economic crisis, which highlighted the country's fragile health care system [12,14-18]. Yet, undergraduate medical curricula still lack DM education, which remains limited to postgraduate in-service training for practitioners and emergency physicians only. Incorporating DM as a component of the medical curricula constitutes a critical strategy in medical education in Lebanon.

The Lebanese University Faculty of Medical Sciences (LUFMS) is the only public medical school and serves many low-income students all over the country, with around 130 graduates annually. Due to ongoing hostilities and the economic crisis, many low-income students from remote areas are unable to afford the money required to travel to and from campus. The recent COVID-19 pandemic has led to the development of e-learning alternatives that have made it even more accessible to students from these areas. e-learning can provide a

cost-effective, equitable, and even practical and interactive solution for DM education [18-24]. However, implementation of such training in crisis-affected settings poses operational challenges [25,26].

This study aimed to evaluate the effectiveness of e-learning versus face-to-face (F2F) approaches in DM education for improving knowledge acquisition and retention among medical students. DM education was delivered through both e-learning and F2F modalities because DM is not included in the undergraduate medical curriculum at LUFMS, and no standardized format for its delivery exists. Evaluating both modalities was therefore essential to determine the most feasible, equitable, and effective approach for nationwide implementation in a context where financial constraints, transportation barriers, and recurrent disruptions frequently hinder in-person attendance. This study also assessed students' self-perceived confidence, competence, satisfaction, and engagement to inform the acceptability, effectiveness, and feasibility of implementing such courses in fragile or disaster-prone settings.

Methods

DM Course Development

The online course used the International Committee of the Red Cross (ICRC) DM e-learning tool. This e-learning tool was an existing, predeveloped ICRC course and was not created by the authors for this study. The development of the ICRC tool was based on the standard principles that are used by multiple courses and learning resources, including the major incident medical management and support course, World Health Organization, and Médecins Sans Frontières, with each module referencing these standards as relevant [27-36]. Development of the F2F course was based on the course contents of the e-learning tool. Both e-learning and F2F DM courses used in this study were based on 10 core topics (Table 1). The training is designed to promote effective learning through interactive, experience-based methods that encourage active engagement and reflection. This approach is informed by constructivist theory, which emphasizes that learners build knowledge through meaningful experiences. To support motivation and a sense of autonomy, the training incorporates principles from self-determination theory, offering flexible learning options that allow individuals to take ownership of their learning process. Additionally, the structure of the content is guided by cognitive load theory, ensuring that information is presented in a clear and manageable way to enhance understanding and retention. The design of the course schedule incorporated a degree of flexibility to accommodate institutional curriculum requirements and the blended e-learning modality used. Of note, principles

of responder and scene safety, including self-safety, hazard identification, and ensuring a safe environment before initiating triage or patient care, were incorporated across multiple modules, particularly the introduction to DM, prehospital

disaster management, and triage modules, consistent with ICRC, World Health Organization, Médecins Sans Frontières, and major incident medical management and support course standards.

Table 1. Topics and subtopics included in the courses in this study.

Topic	Subtopic
Introduction to disaster medicine and taxonomy	<ul style="list-style-type: none"> Disaster medicine definitions “disaster” and “MCI^a” Difference between DM^b and EM^c Public health principles in disasters Disaster management phases: 3 phases The concept of “hazard vulnerability analysis”
Triage in disasters	<ul style="list-style-type: none"> MCI triage definitions Different triaging systems EM versus DM triaging
Hospital disaster preparation and response	<ul style="list-style-type: none"> Laws and DM Internal vs external hospital incidents What is an MCI plan? Integration of medical staff
Health consequences of different disasters	<ul style="list-style-type: none"> Types of disasters and the impact of each Natural vs man-made disasters: the health impact
Prehospital disaster management	<ul style="list-style-type: none"> MCI: disposition, treatment, and transport Disaster plans and the control chain Functional response rate
Psychosocial care	<ul style="list-style-type: none"> Psychic reactions and disaster exposure Treatment of incident stress reactions
Presentation of past disasters and public health emergencies	<ul style="list-style-type: none"> Beirut post explosion COVID-19 pandemic
Pediatric injuries in MCI	<ul style="list-style-type: none"> Trauma assessment and triage Pediatric-specific injuries Psychological and emotional support Pediatric resuscitation
MCI and burn injuries	<ul style="list-style-type: none"> Burn severity index Burn resuscitation Wound care and dressing Pain management Psychological support
Management of the dead bodies after disasters	<ul style="list-style-type: none"> Disaster victim identification Forensic pathology and autopsy Temporary mortuary facilities Body handling Documentation Cultural and religious considerations Public health and hygiene

^aMCI: mass casualty incident.

^bDM: disaster medicine.

^cEM: emergency medicine.

Study Design

This is a quasi-experimental controlled study using a Solomon 4-group design [37]. Quasi-experimental designs are appropriate when random assignments are impractical, allowing for the evaluation of interventions in real-world settings while

maintaining a degree of control over confounding variables [38]. In the Lebanese context, constraints such as transportation difficulties, financial limitations, and geographical barriers rendered random allocation to e-learning and F2F groups unfeasible. Therefore, participants were assigned based on their ability to attend on-campus sessions, considering factors like

proximity and access to reliable transportation rather than their personal preference or academic performance, and selection bias related to motivation was minimal. The design balances trade-offs between internal and external validity, feasibility, inclusiveness, and applicability in educational real-life settings in resource-constrained environments [39]. Furthermore, the Solomon 4-group design enhances the robustness of the study by controlling confounding variables and potential pretest sensitization effects [37].

Target Population

The target population consisted of medical students between the second and fifth academic years. All students from the second to the fifth academic years received identical course content, as DM is not part of the formal curriculum at any level. Standardizing the content ensured comparability across groups and isolated the effect of the delivery modality rather than differences in academic year. No interns were included in this study.

Sampling

The total number of the targeted population was 512. The sample size was calculated using G*Power software to detect the effect size (Cohen $d=0.4$) for the primary outcome (posttraining effectiveness difference between e-learning and F2F, measured by the significant difference in knowledge scores between the groups), with a significance level of .05, a desired power of 0.80, and a 2-tailed test. The estimated minimum sample size was 174 participants (87 in each group). The population count per group was determined with the assistance of delegates from each year for the medical students via convenience sampling, such as open registration on a first-come, first-served basis, regardless of their academic year. Group assignments were determined based on students' logistical capabilities. Those with the ability to attend on-campus sessions, considering factors such as proximity to campus, access to reliable transportation, and affordability, were assigned to the F2F cohort. Conversely,

students facing logistical barriers that precluded on-campus attendance were placed in the e-learning cohort.

DM Course Implementation

The study involved delivering a DM course using identical educational materials through 2 modalities: e-learning and F2F sessions. The course was delivered over 3 days per week for 2 weeks. The F2F sessions were implemented by the principal investigator at LUFMS. All synchronous e-learning sessions were also delivered by the principal investigator to ensure standardization across modalities. The digital version for the e-learning groups was a set of modules built onto a learning management system, which was delivered to the respective subgroups in a synchronized fashion by the principal investigator, with the lectures on the same dates to the assigned subgroups. Both groups received a digital version of the respective teaching material at the end of each session, while the F2F group received the slides for the respective session.

Evaluation Framework and Outcomes

The first 2 levels of Kirkpatrick's model were used to assess knowledge, knowledge retention, learning, and satisfaction. Additionally, self-efficacy, as a predictor of skill application, was measured through perceived behavioral outcomes, specifically confidence (self-perceived competence), and was guided by social cognitive theory [40].

1. Assessment of the students' knowledge: Students' knowledge assessment consisted of 20 multiple-choice questions (Table 2). The 20-item multiple-choice questions assessment included 2 questions per core topic, as summarized in Table 2; the full questionnaire is provided in Multimedia Appendix 1. Each question had 5 possible answers, and only 1 answer was correct.

Based on Solomon 4-group design model, students within each year and learning modality were divided randomly into 2 groups [37]. Each academic year was subdivided into 4 subgroups (A, B, C, and D; Table 3).

Table 2. Multiple-choice question topics per core subject.

Topic	Multiple choice questions
Introduction to disaster medicine and taxonomy	<ul style="list-style-type: none"> Question 1: Which statement is correct about DM^a?
Triage in disasters	<ul style="list-style-type: none"> Question 2: What is the purpose of triaging in a medical emergency? Question 3: In the triage system, which color is typically associated with the highest priority level?
Hospital disaster preparedness and response	<ul style="list-style-type: none"> Question 4: What is the role of MOH^b in DM? Question 5: What is an MCI^c plan? Question 6: Role of hospitals and prehospital units in preparedness Question 7: Which means of communication is most important in DM?
Health consequences of different disasters	<ul style="list-style-type: none"> Question 8: Which type of disaster affects the respiratory system? Question 9: Which statement is correct?
Prehospital and hospital disaster management	<ul style="list-style-type: none"> Question 10: Who controls during a disaster? Question 11: What is the role of an incident commander? Question 12: In an MCI, who is typically responsible for declaring a patient deceased? Question 13: What is the role of hospitals in disaster preparedness?
Psychosocial care	<ul style="list-style-type: none"> Question 14: What does PTSD^d stand for? Question 15: What are treatment modalities for stress during disasters?
Presentation of past disasters and public health emergencies	<ul style="list-style-type: none"> Question 16: Recall the hospitals damaged during the incident Question 17: What is COVID-19 pandemic most significant impact worldwide?
Pediatric injuries in MCI	<ul style="list-style-type: none"> Question 18: In an MCI involving pediatric injuries, which of the following is essential to identify first?
MCI and burn injuries	<ul style="list-style-type: none"> Question 19: In the context of mass casualty incidents involving burn injuries, which aspect is essential for initial assessment?
Management of the dead bodies after disasters	<ul style="list-style-type: none"> Question 20: When managing dead bodies after disasters, which of the following should be prioritized first?

^aDM: disaster management.^bMOH: Ministry of Health.^cMCI: mass casualty incident.^dPTSD: posttraumatic stress disorder.**Table 3.** Subgroup classification of the patients within each academic year and the timing of the assessments based on Solomon 4-group design model.

Time points	Medical students: second-fifth year (subgroups A-D)
First week	Phase I: precourse assessment: Subgroups A and B had face-to-face sessions, and subgroups C and D had e-learning in the first and second weeks
Second week “end of course”	Phase II: postcourse assessment online survey for all subgroups
1 month post course completion	Phase III: assessment for all subgroups

The assessment was done over 3 phases:

- Precourse assessment: conducted 15 minutes before the course for subgroups A and C only.
- Postcourse assessment: conducted after the end of the final session for all subgroups. The pre- and post-assessments used the same questionnaire details.
- Follow-up assessments: conducted 1 month after the finalization of the DM educational course to evaluate knowledge retention.

- Assessment of the satisfaction and recommendations of students: the assessment of satisfaction was done using a

tool developed by Han et al [38]. The tool consisted of questions with a rating scale (Multimedia Appendix 2) [35,37]:

- “Reaction” of the participants: how they felt and thought about the course
- “Learning” of the participants: how they described the increase in knowledge after the course.

In addition, 3 open questions addressed the experience of the participants. The first question addressed the most unique aspect of the course. The second question addressed the suggestions

of the participants for improvement of the course. The last question requested additional comments from the participants.

1. Evaluation of self-efficacy: a postcourse evaluation of confidence was conducted for all subgroups in each academic year. The evaluation of the confidence level was measured using a validated Disaster Preparedness Evaluation Tool [38,39]. The original version was adapted to the educational material implemented in the course (Multimedia Appendix 3) [35,36]. Students will be asked their level of confidence on a Likert scale as “strongly disagree,” “disagree,” “neutral,” “agree,” and “strongly agree.”

Data Collection Tools and Methods

The student’s ID, their academic year, and evaluation data were confidentially collected using the Google Forms survey tool, which was converted to a Microsoft Excel sheet and saved anonymously (Multimedia Appendix 4). The handling and access of the data were limited to the principal investigator to ensure data security. All data were stored in accordance with institutional data protection policies.

Reliability Assessment

A post hoc reliability analysis check of the scale used to assess the knowledge, satisfaction, and confidence level showed excellent internal consistency (Cronbach α was 0.82, 0.92, and 0.963, respectively).

Data Analysis

Quantitative Analysis

Descriptive statistics, including percentages, means, and SD values, were used to describe the sample and summarize survey scores. Knowledge scores were reported as the number of correct responses. Repeated measures ANOVA, also referred to as a within- and between-subjects ANOVA for correlated samples, was used to detect any overall differences between related means at different points in time. Chi-square and Fisher exact tests were used to evaluate differences in the items of the satisfaction scale and confidence scale across the different options. An independent-sample *t* test was used to compare satisfaction and confidence scores in the overall sample and stratified by year of education between the 2 learning modalities. Data were analyzed using the Statistical Package for the Social Sciences

(version 27.0; IBM Corp). A *P* value less than .05 was considered significant.

Qualitative Analysis

Open question responses were inductively analyzed and coded, and recurrent themes were presented in percentages.

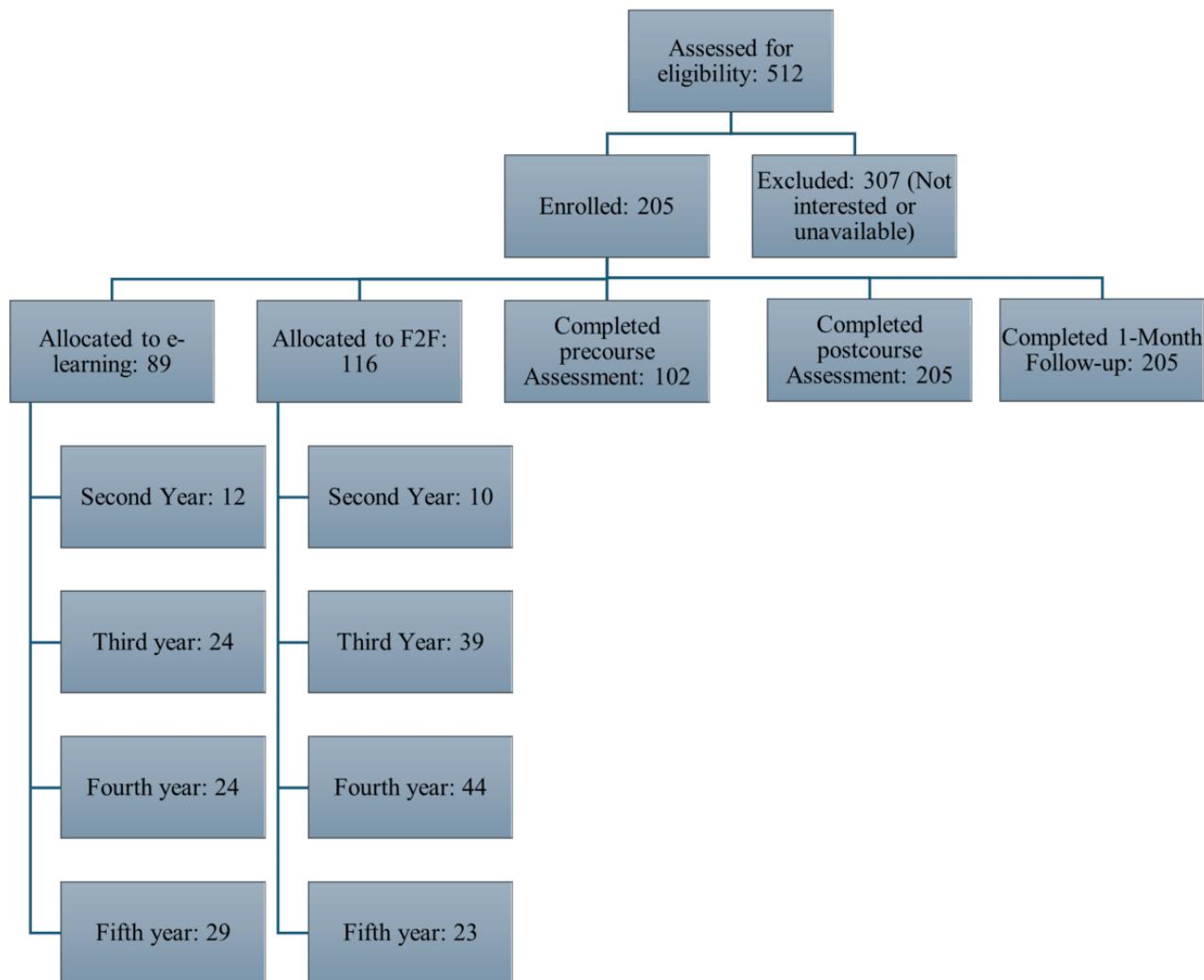
Ethical Considerations

This project was conducted according to the applied ethical guidelines and received the necessary approvals from the relevant ethical review boards. The study was approved by the Institutional Review Board of Al Zahraa Hospital University Center under reference number 12/2023. Additionally, the study received an exemption from the Ethics Review Board of the ICRC in Geneva. The ICRC Ethics Review Board confirmed that the research proposal presented minimal risks to participants, which were adequately addressed and minimized through the proposed procedures. The exemption reference number for this study is 0623. Informed consent was obtained from all participants, and strict confidentiality was maintained in the handling of all data to ensure participants’ privacy and well-being throughout the study.

Results

Descriptive Statistics

Participants were recruited, and the course was conducted between March 2024 and June 2024. The participant flowchart is presented in Figure 1. Out of 512 students at LUFMS from the second to the fifth academic years, 205 (40%) voluntarily participated in the study, of which 116 (56.6%) joined the e-learning cohort, and 89 (43.4%) joined the F2F cohort between March and June 2024. The third and fourth academic year students represented 30.7% (63/205) and 33.2% (68/205) of the participants, respectively. Fifth academic year students were 25.4% (52/205), and the second academic year students were the least (22/205, 10.7%). The distribution of students according to the crosstabulation of the academic year and learning modality is depicted in Figure 1. Participation rates from the second and fifth academic years were comparable across the F2F and e-learning groups, with 12 (13.9%) and 10 (8.6%) students from the second year, and 29 (32.5%) and 23 (19.8%) students from the fifth year, respectively. Nearly 61.9% (39/63) of the third academic year students and 64.7% (44/66) of the fourth academic year students joined the F2F cohort.

Figure 1. Participant flow diagram showing distribution by academic year and learning modality (F2F vs e-learning). F2F: face-to-face.

Assessment of Precourse Sensitization

Based on Solomon 4-group design model [37], a summarized comparison for the overall sample is presented in **Table 4**. The analysis shows that pretested and non-pretested groups demonstrated comparable scores at baseline, at the immediate postcourse assessment, and at 1-month follow-up ($P=.60, .40,$

and $.64$, respectively). These results confirm that the precourse assessment did not introduce sensitization effects or bias the learning outcomes, thereby supporting the internal validity of the study design. Hence, precourse assessment is not a confounding factor and did not affect the outcome across the subgroups.

Table 4. Summary of Solomon 4-group analysis for the overall sample.

Overall	Pretest	Posttest	Post 1 month
F2F^a			
Experimental, mean (SD)	6.3 (2.3)	12.3 (1.8)	14.3 (2.4)
Control, mean (SD)	— ^b	12.2 (1.7)	15.7 (3.9)
e-Learning			
Experimental, mean (SD)	6.6 (2.3)	12 (2.1)	14.6 (3.5)
Control, mean (SD)	—	13.5 (3.7)	16.4 (4.4)
<i>P</i> value between groups	.60	.40	.64

^aF2F: face-to-face.

^bNot applicable.

Assessment of the Knowledge Level in the Overall Sample and Across the Academic Years

For the overall sample, both F2F and e-learning groups demonstrated significant improvements from precourse to immediate postcourse and post 1 month, with no significant statistical difference between both modalities overall ($P=.40$ and $.64$) and across each year (Multimedia Appendix 5). The F2F cohort showed a mean increase of 5.9 (95% CI 5.35-6.45; $P<.001$) and 7.9 (95% CI 7.34-8.46; $P<.001$) from precourse to postcourse and post 1 month, respectively (Table 5). Similarly, the e-learning group exhibited a mean increase of 5.4

(95% CI 4.69-6.11; $P<.001$) and 8 (95% CI 7.11-8.89; $P<.001$) over the same periods. In the second year, the F2F group showed a significant increase of 5.6 (95% CI 3-8.2; $P=.04$) and 7.6 (95% CI 5.68-9.52; $P=.002$) from precourse to postcourse and post 1 month, respectively. However, the e-learning group did not show significant improvement, with a mean increase of 4 (95% CI 0.83-7.17; $P=.74$) and 7 (95% CI 4.57-9.43; $P=.14$), indicating less effectiveness in the second year. Of note, all mean change values were calculated at the individual level and then averaged within each subgroup; no pooled scoring was used.

Table 5. Mean change in knowledge scores from precourse to postcourse and 1-month follow-up.

Cohort	Sample size	Postcourse, mean (SD)		1-month follow-up, mean (SD)		<i>P</i> value
		Mean (SD)	95% CI	Mean (SD)	95% CI	
F2F total ^a	116	5.9 (3)	5.35-6.45	7.9 (3.1)	7.34-8.46	<.001
e-learning total	89	5.4 (3.4)	4.69-6.11	8 (4.3)	7.11-8.89	<.001
Second-year F2F	10	5.6 (4.2)	3-8.20	7.6 (3.1)	5.68-9.52	.002
Second-year e-learning	12	4 (5.6)	0.83-7.17	7 (4.3)	4.57-9.43	.14
Third-year F2F	39	5.3 (2.1)	4.64-5.96	7.3 (2.3)	6.58-8.02	<.001
Third-year e-learning	24	5.8 (3.2)	4.52-7.08	8.3 (4.3)	6.58-10.02	<.001
Fourth-year F2F	44	7.1 (3.6)	6.04-8.16	9 (3.9)	7.85-10.15	.003
Fourth-year e-learning	24	4.9 (3.3)	3.58-6.22	7.2 (4.8)	5.28-9.12	<.001
Fifth-year F2F	23	6.1 (2.7)	5-7.20	8.2 (3.4)	6.81-9.59	<.001
Fifth-year e-learning	29	6.1 (3)	5.01-7.19	10.1 (2.6)	9.15-11.05	<.001

^aF2F: face-to-face.

The fifth-year students exhibited the highest mean score increases within their respective learning modalities. The F2F group showed a mean increase of 6.1 (95% CI 5-7.2; $P=.005$) and 8.2 (95% CI 6.81-9.59; $P<.001$) from precourse to postcourse and post 1 month, respectively. The e-learning group showed a comparable increase of 6.1 (95% CI 5.01-7.19; $P=.005$) and 10.1 (95% CI 9.15-11.05; $P<.001$) over the same periods.

Assessment of the Satisfaction Level in the Overall Sample and Across the Academic Years

The results show that the total satisfaction score (SD) was slightly higher for the F2F cohort (mean 87.6, SD 11.3; 95% CI 85.54-89.66) compared to e-learning (mean 86, SD 14.2; 95% CI 83.05-88.95). Only the e-learning cohort of the fifth academic year students showed a higher score (mean 90.2, SD 11.7; 95% CI 85.94-94.46) compared to F2F (mean 87.5, SD 9.6; 95% CI 83.58-91.42; Table 6). No statistically significant differences in satisfaction scores were observed between F2F and e-learning modalities in the overall sample ($P=.38$) or within any academic year subgroup.

Table 6. Mean change in the satisfaction score compared to the precourse assessment: at the postcourse assessment, directly postcourse.

Cohort	Postcourse assessment, mean (SD)	Sample size, n	95% CI	P value ^a
F2F ^b total	87.6 (11.3)	116	85.54-89.66	.38
e-learning total	86 (14.2)	89	83.05-88.95	— ^c
Second-year F2F	87.6 (15)	10	78.3-96.9	.81
Second-year e-learning	86.2 (11.32)	12	79.8-92.6	—
Third-year F2F	87.5 (11.5)	39	83.89-91.11	.51
Third-year e-learning	84.9 (16.65)	24	78.24-91.56	—
Fourth-year F2F	87.6 (11.5)	44	84.2-91	.35
Fourth-year e-learning	84.5 (13.6)	24	79.06-89.94	—
Fifth-year F2F	87.5 (9.6)	23	83.58-91.42	.37
Fifth-year e-learning	90.2 (11.7)	29	85.94-94.46	—

^aP values calculated using independent-sample Welch *t* tests.

^bF2F: face-to-face.

^cNot available.

Although the comparison of the total satisfaction between F2F and e-learning showed statistically nonsignificant *P* values, comparison of the individual satisfaction scores of each question between the 2 modalities was statistically significant for only 4 questions in favor of F2F (Figure 2). Students were mostly

satisfied with the vividness of the course and its relevance to their medical career, enhancement of their knowledge of disaster management, and their engagement during the course (Figure 3).

Figure 2. Horizontal bar graph showing the percentage of satisfaction of the participants with each learning modality in the statistically significant questions. F2F: face-to-face.

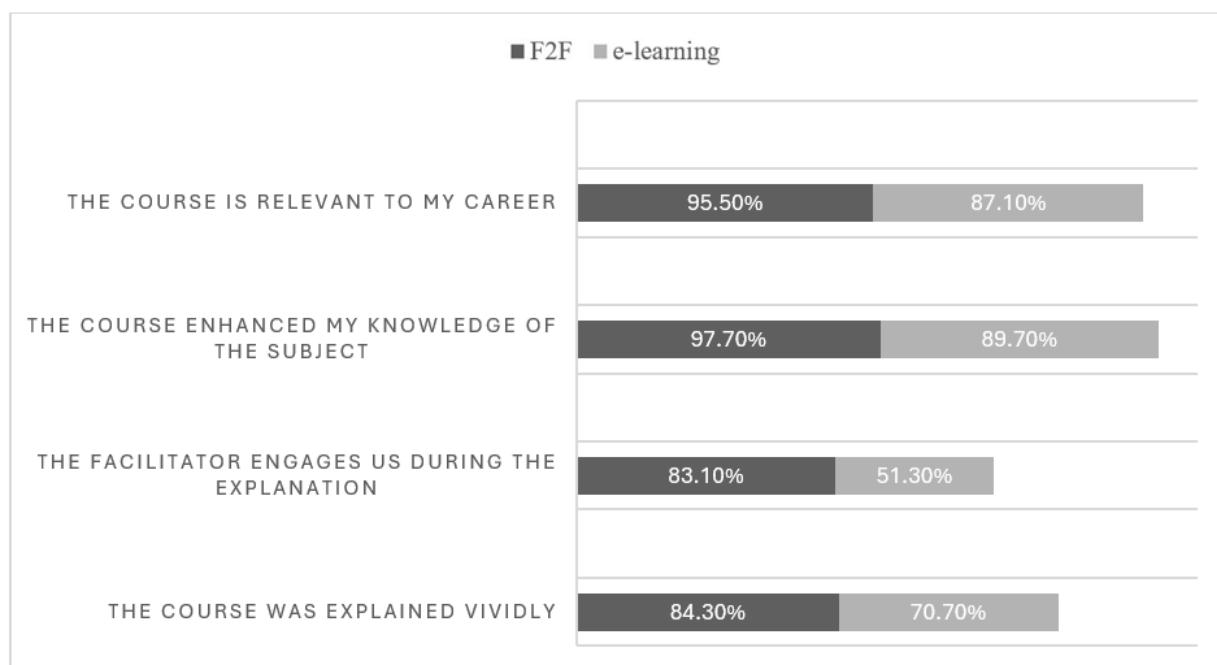
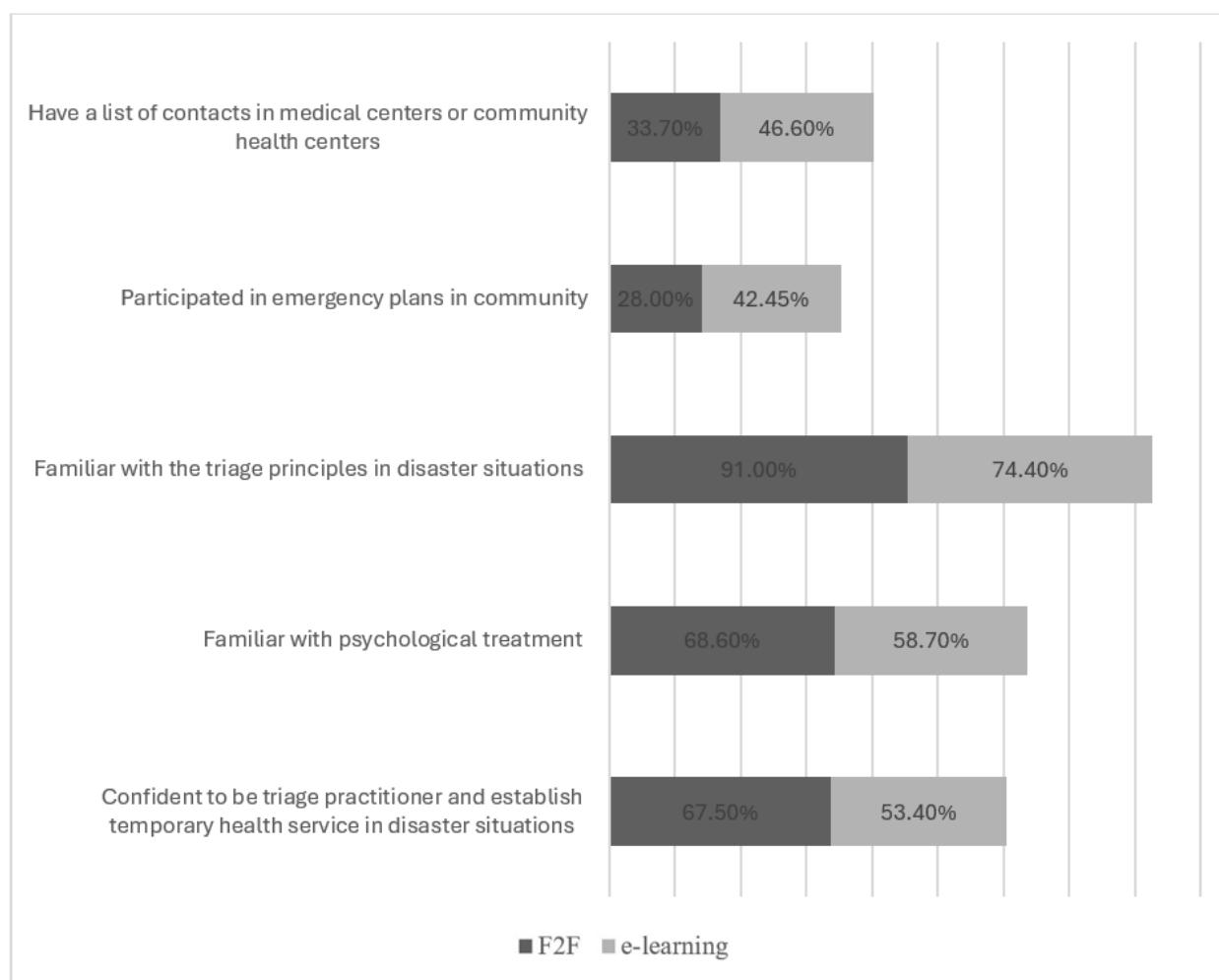


Figure 3. Horizontal bar graph showing the percentage of confidence of the participants with each learning modality in the statistically significant questions. F2F: face-to-face.



Assessment of the Confidence Level in the Overall Sample and Across the Academic Years

The results show that the total confidence score (SD) was similar between F2F (mean 125.5, SD 20.6; 95% CI 121.75-129.25) and e-learning (mean 126.5, SD 23.8; 95% CI 121.56-131.44). Only the F2F cohort of the second academic year students showed scores that were like the overall score, yet better than e-learning, with a mean F2F score of 125.4 (SD 22.5; 95% CI 111.45-139.35) compared to a mean of 116 (SD 13.9; 95% CI

108.14-123.86) for e-learning (Table 7). The e-learning cohort of the fifth academic year students showed higher scores compared to F2F, with a F2F score of mean 124.9 (SD 20.1; 95% CI 116.69-133.11) compared to mean 135.3 (SD 21.71; 95% CI 127.4-143.2) for e-learning. No statistically significant differences in confidence scores were observed between learning modalities in the overall sample ($P=.75$). A marginal trend favoring the e-learning group was noted in the fifth-year subgroup ($P=.08$), although it did not reach statistical significance.

Table 7. Mean change in the confidence score compared to the precourse assessment: at the postcourse assessment, directly postcourse.

Cohort	Postcourse assessment, mean (SD)	Sample size, n	95% CI	P value ^b
F2F ^a total	125.5 (20.6)	116	121.75-129.25	.75
e-learning total	126.5 (23.8)	89	121.56-131.44	— ^c
Second-year F2F	125.4 (22.5)	10	111.45-139.35	.27
Second-year e-learning	116 (13.9)	12	108.14-123.86	—
Third-year F2F	124.7 (23.1)	39	117.45-131.95	.99
Third-year e-learning	124.6 (28.4)	24	113.24-135.96	—
Fourth-year F2F	127 (18.7)	44	121.47-132.53	.70
Fourth-year e-learning	125 (21.1)	24	116.56-133.44	—
Fifth-year F2F	124.9 (20.1)	23	116.69-133.11	.08
Fifth-year e-learning	135.3 (21.71)	29	127.4-143.2	—

^aF2F: face-to-face.

^bP values calculated using independent-sample Welch *t* tests.

^cNot available.

Although the comparison of the total confidence between F2F and e-learning showed no statistically significant *P* values (Table 7, the comparison of the individual confidence score of each question between the 2 modalities was statistically significant for only 5 questions. Students were mostly confident about their knowledge of triaging principles in a disaster through F2F learning (105/116, 91%) versus e-learning (66/89, 74.4%). Similarly, students expressed a higher familiarity with psychological treatment and the ability to triage patients in disaster situations through F2F learning (80/116, 68.6% and 78/116, 67.5%, respectively) relative to e-learning (52/89, 58.7% and 48/89, 53.4%, respectively). The confidence level was below 50% for either modality when students were asked about participation in community emergency plans and having contacts in community health centers.

Unique Aspect of the Course

The participants highlighted the distinctiveness of the course. Around 85% (174/205) of the participants appreciated the relevance of the course to real-life situations and disaster management. Nearly 75% (153/205) of the participants found engagement with a variety of topics, including mass casualty incidents, triaging, and the management of specific disasters like white phosphorus. The innovative teaching methods and the adaptation of the course to the Lebanese situation were praised by 70% (143/205) of the participants, while 65% (133/205) valued the case studies that helped to effectively manage the disasters.

Areas of Improvement

Around 60% (123/205) of the participants suggested the inclusion of interactive and practical sessions to reinforce through hands-on practice. Nearly 55% (112/205) of the participants suggested improvement of the scheduling of the sessions, while some proposed that it should become a mandatory part of the curriculum. Half of the participants noted that the artificial intelligence voice used in the e-learning course was mildly distracting. Around 45% (92/205) suggested that

live sessions would enhance the course's effectiveness, especially for complex medical terms and scenarios.

Additional Comments

About 80% (164/205) of the participants expressed appreciation for the course and gratitude for the opportunity to learn about disaster management. Nearly 70% (143/205) recognized the importance of the course, considering the recent events in Lebanon. A desire for the incorporation of similar courses into the curriculum was noted by 65% (133/205) of the participants. There were also suggestions from 55% (112/205) of participants to expand the course to include more practical exercises, community-based projects, and timely distribution of course materials to ensure better knowledge retention.

Discussion

Principal Findings

This study evaluated whether e-learning is comparable to F2F instruction in DM training for LUFMS medical students and contributes to a growing body of evidence supporting the effectiveness of e-learning [40-45]. The results demonstrated significant improvements in knowledge and retention for both modalities, with no statistical difference between them (*P*=.40). However, F2F training resulted in higher confidence in practical skills like triage (91% vs 74.4%), while e-learning resulted in superior retention among fifth-year students. Satisfaction was high in both groups, although qualitative feedback emphasized the need for more hands-on sessions (60%). The results have direct implications for the delivery of DM education and the expectation of outcomes, especially in countries of protracted conflict like Lebanon [46,47]. A blended DM training approach with the skill-building benefits of targeted F2F sessions offers the most pragmatic solution. Still, the equivalence in knowledge gains supports e-learning as a viable alternative in resource-limited settings.

While both delivery methods were effective across academic years, the higher mean gains in knowledge and retention

observed among fifth-year e-learning students can be attributed to their advanced medical knowledge compared to students in the preceding academic years. The LUFMS curriculum typically incorporates diagnostic courses within the fifth year, and students commonly undertake a 6-week externship after the fifth academic year. This confluence of advanced knowledge and practical experience may have motivated the fifth-year students to excel in the assessments, recognizing the opportunity to apply their learning in real-life and clinical settings [6]. Additionally, the fifth academic year students were the only group who reported satisfaction and confidence levels with e-learning higher compared to F2F due to their 2-year antecedent exposure to online learning. Given their advanced medical knowledge and diagnostic expertise, the students' ability to integrate prior knowledge with new concepts acquired through e-learning and independent research likely played a significant role in their favorable outcomes.

Both modalities showed low confidence in community emergency plan participation and health center contacts. This can be attributed to the fact that the participants have not engaged in hospital situations and do not interact with hospital staff except their academic mentors. Therefore, they have limited knowledge about the emergency plans in hospitals. This suggests DM implementation should be timed with clinical rotations to maximize its relevance and effectiveness.

Limitations

This study has several limitations, which stem from the difficulty of conducting similar studies in a volatile context such as Lebanon, a country with a history of disasters and emergencies, which also impacts the educational settings. Nevertheless, the findings are particularly relevant and have direct application and impact in practice. Voluntary recruitment and the lack of random assignments of the participants may have compromised internal validity, although the real-world applicability, along

with sufficient sample size, enhances the study's external validity, especially for comparable settings. Furthermore, the application of the Solomon 4-group design model classification of the participants allowed for control confounding and pretest sensitization [37], enhancing the internal validity of the study. The short follow-up period limits the conclusion about long-term retention. Still, the significant increase in knowledge over time suggests the consolidation of learning over time due to its real application in their setting. Finally, the reliance on self-reported learning and confidence can introduce bias, but the results were also triangulated by qualitative data and objective assessment with a validated tool [48-51], which also enhances the applicability and internal validity of the study.

Although "lack of interest" was the most frequently selected reason for nonparticipation, this explanation likely underrepresents the true contextual barriers faced by students. Informal feedback indicated that transportation costs, fuel shortages, intermittent electricity and internet access, work obligations, and exam schedules limited their ability to participate. Therefore, the high nonparticipation rate should be interpreted within Lebanon's unstable socioeconomic context, where such logistical constraints, not genuine disinterest in DM, are more plausible explanations for nonenrollment.

Conclusions

This study demonstrates that e-learning and F2F DM education are comparably effective in improving knowledge retention, confidence, and perceived competence among Lebanese medical students. The findings support integrating a DM course into the fifth-year curriculum before externships to enhance preparedness and application of skills. A blended approach combining theoretical e-learning with practical F2F sessions may further strengthen practical competencies. These results provide evidence to guide curriculum development and DM training implementation at the LUFMS.

Acknowledgments

The authors would like to thank all students, faculty, and department who contributed to the fulfillment of this project. All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Funding

This study was funded by the Agence Française de Développement as part of its partnership project with the International Committee of the Red Cross and Rafik Hariri University Hospital. AFD had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the article.

Data Availability

Data are available with the corresponding author upon request.

Authors' Contributions

Conceptualization: AM, CB, HS, and RP

Methodology: AM, CB, DA, and AB

Ethical approval: ZM and HAM

Project administration: AM, HAM, MM, and AB

Supervision: MM, AB, and RP
Teaching: AM
Resources: AM and AB
Investigation: AM, HAM, and RP
Data curation: AM, LA, and AB
Formal analysis: AM, LA, and MM
Validation: LA and MM
Original draft: AM
Review and editing: AM, CB, ZM, HS, MM, and AB
Visualization: AM

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questions for knowledge for phases I, II, and III.

[[PDF File \(Adobe PDF File\), 336 KB - mededu_v12i1e80409_app1.pdf](#)]

Multimedia Appendix 2

Questions for the evaluation of the satisfaction of participants in the DM course utilizing the Likert scale.

[[DOCX File, 18 KB - mededu_v12i1e80409_app2.docx](#)]

Multimedia Appendix 3

Questions for evaluations of the confidence of participants in the disaster management educational course utilizing the Likert scale.

[[DOCX File, 19 KB - mededu_v12i1e80409_app3.docx](#)]

Multimedia Appendix 4

All data collection schemes.

[[DOCX File, 15 KB - mededu_v12i1e80409_app4.docx](#)]

Multimedia Appendix 5

Mean knowledge scores for the four groups over the three assessment phases in the overall sample and stratified by year of education and according to the Solomon 4-group design.

[[DOCX File, 19 KB - mededu_v12i1e80409_app5.docx](#)]

References

1. Kim PY, Sawe HR, Wallis LA. The Sendai Framework and emergency care. *Afr J Emerg Med* 2016 Mar;6(1):3-4 [[FREE Full text](#)] [doi: [10.1016/j.afjem.2015.09.001](https://doi.org/10.1016/j.afjem.2015.09.001)] [Medline: [30456056](#)]
2. Ingrassia PL, Ragazzoni L, Tengattini M, Carenzo L, Della Corte F. Nationwide program of education for undergraduates in the field of disaster medicine: development of a core curriculum centered on blended learning and simulation tools. *Prehosp Disaster Med* 2014 Oct;29(5):508-515 [[FREE Full text](#)] [doi: [10.1017/S1049023X14000831](https://doi.org/10.1017/S1049023X14000831)] [Medline: [25155942](#)]
3. Chong K, Abdul NF, Musa S, Shahron S, Anuar N. Challenges and opportunities of disaster education program among UUM students. 2018. URL: <https://ui.adsabs.harvard.edu/abs/2018AIPC.2016b0038L/abstract> [accessed 2026-01-03]
4. Ragazzoni L, Ingrassia PL, Gugliotta G, Tengattini M, Franc JM, Corte FD. Italian medical students and disaster medicine: awareness and formative needs. *Am J Disaster Med* 2013;8(2):127-136. [doi: [10.5055/ajdm.2013.0119](https://doi.org/10.5055/ajdm.2013.0119)] [Medline: [24352928](#)]
5. International FOMSA. IFMSA Policy Statement on Disasters and Emergencies 2012. Accessed February 6. 2014. URL: <http://imccam2012.files.wordpress.com/2012/07/psam12-disasters-and-emergencies.pdf> [accessed 2026-01-03]
6. Kaji AH, Coates W, Fung C. A disaster medicine curriculum for medical students. *Teach Learn Med* 2010;22(2):116-122. [doi: [10.1080/10401331003656561](https://doi.org/10.1080/10401331003656561)] [Medline: [20614377](#)]
7. Mortelmans LJ, De Cauwer HG, Van Dyck E, Monballyu P, Van Giel R, Van Turnhout E. Are Belgian senior medical students ready to deliver basic medical care in case of a H5N1 pandemic? *Prehosp Disaster Med* 2009;24(5):438-442. [doi: [10.1017/s1049023x00007287](https://doi.org/10.1017/s1049023x00007287)] [Medline: [20066648](#)]
8. Hermann S, Gerstner J, Weiss F, Aichele S, Stricker E, Gorgati E, et al. Presentation and evaluation of a modern course in disaster medicine and humanitarian assistance for medical students. *BMC Med Educ* 2021;21(1):610 [[FREE Full text](#)] [doi: [10.1186/s12909-021-03043-6](https://doi.org/10.1186/s12909-021-03043-6)] [Medline: [34893083](#)]

9. Bhattacharya S, Singh A, Semwal J, Marzo RR, Sharma N, Goyal M, et al. Impact of a training program on disaster preparedness among paramedic students of a tertiary care hospital of North India: A single-group, before-after intervention study. *J Educ Health Promot* 2020;9:5 [FREE Full text] [doi: [10.4103/jehp.jehp_423_19](https://doi.org/10.4103/jehp.jehp_423_19)] [Medline: [32154300](https://pubmed.ncbi.nlm.nih.gov/32154300/)]
10. Ashcroft J, Byrne M, Brennan PA, Davies RJ. Preparing medical students for a pandemic: a systematic review of student disaster training programmes. *Postgrad Med J* 2021;97(1148):368-379 [FREE Full text] [doi: [10.1136/postgradmedj-2020-137906](https://doi.org/10.1136/postgradmedj-2020-137906)] [Medline: [32518075](https://pubmed.ncbi.nlm.nih.gov/32518075/)]
11. Voicescu GT, Lamine H, Lo on i AE, Lupan-Murean EM, Luka S, Ulerio JG, et al. Monitoring and evaluation in disaster management courses: a scoping review. *BMC Med Educ* 2025;25(1):188 [FREE Full text] [doi: [10.1186/s12909-025-06659-0](https://doi.org/10.1186/s12909-025-06659-0)] [Medline: [39910476](https://pubmed.ncbi.nlm.nih.gov/39910476/)]
12. Bahattab AA, Linty M, Trentin M, Truppa C, Hubloue I, Della Corte F, et al. Availability and characteristics of humanitarian health education and training programs: a web-based review. *Prehosp Disaster Med* 2022;37(1):132-138. [doi: [10.1017/S1049023X21001333](https://doi.org/10.1017/S1049023X21001333)] [Medline: [34928199](https://pubmed.ncbi.nlm.nih.gov/34928199/)]
13. Saeed Bahattab AA, Amat Camacho N, von Reding H, Rogers M, Lajolo CP, Salio F, et al. Climate, conflict, and epidemic: global and regional disaster profiles and implications for emergency medical teams. *Prehosp Disaster Med* 2025 May 21;40(S1):s3-s3. [doi: [10.1017/s1049023x25000305](https://doi.org/10.1017/s1049023x25000305)]
14. Lebanon. World Bank Group. URL: <https://data.worldbank.org/country/lebanon> [accessed 2026-01-03]
15. Lebanon profile-timeline. BBC. 2021 Jan 08. URL: <https://www.bbc.com/news/world-middle-east-14649284> [accessed 2024-01-03]
16. Beirut explosion: what we know so far.: BBC; 2020 Aug 5. URL: <https://www.bbc.com/news/world-middle-east-53668493> [accessed 2024-01-03]
17. Crisis briefing. International Committee of the Red Cross. URL: <https://www.rescue.org/country/lebanon> [accessed 2024-01-03]
18. Lebanon. World Health Organization. 2023 Apr 03. URL: <https://www.who.int/countries/lbn/> [accessed 2026-01-03]
19. Bahattab A, Caviglia M, Martini D, Hubloue I, Della Corte F, Ragazzoni L. Scenario-based e-simulation design for global health education: theoretical foundation and practical recommendations. *J Med Internet Res* 2023;25:e46639 [FREE Full text] [doi: [10.2196/46639](https://doi.org/10.2196/46639)] [Medline: [37902810](https://pubmed.ncbi.nlm.nih.gov/37902810/)]
20. Zarifsanaiey N, Farrokhi M, Karimian Z, Hoseini S, Chahartangi F, Shahraki HR. Lesson learned from assessing teachers' and students' perspectives regarding the quality of e-learning in medical education during the pandemic: a mixed-methods study. *BMC Med Educ* 2024;24(1):171 [FREE Full text] [doi: [10.1186/s12909-024-05160-4](https://doi.org/10.1186/s12909-024-05160-4)] [Medline: [38388898](https://pubmed.ncbi.nlm.nih.gov/38388898/)]
21. Imran R, Fatima A, Elbayoumi Salem I, Allil K. Teaching and learning delivery modes in higher education: looking back to move forward post-COVID-19 era. *The International Journal of Management Education* 2023;21(2):100805. [doi: [10.1016/j.ijme.2023.100805](https://doi.org/10.1016/j.ijme.2023.100805)]
22. Mahdavi Ardestani SF, Adibi S, Golshan A, Sadeghian P. Factors influencing the effectiveness of e-learning in healthcare: a fuzzy ANP study. *Healthcare (Basel)* 2023;11(14) [FREE Full text] [doi: [10.3390/healthcare11142035](https://doi.org/10.3390/healthcare11142035)] [Medline: [37510477](https://pubmed.ncbi.nlm.nih.gov/37510477/)]
23. De Felice S, Hamilton A, Ponari M, Vigliocco G. Learning others is good, others is better: the role of social interaction in human acquisition of new knowledge. *Philos Trans R Soc Lond B Biol Sci* 2023;378(1870):20210357 [FREE Full text] [doi: [10.1098/rstb.2021.0357](https://doi.org/10.1098/rstb.2021.0357)] [Medline: [36571126](https://pubmed.ncbi.nlm.nih.gov/36571126/)]
24. Luke K. Twelve tips for using synchronous virtual classroom technologies in medical education. *MedEdPublish* (2016) 2021;10:66 [FREE Full text] [doi: [10.15694/mep.2021.000066.1](https://doi.org/10.15694/mep.2021.000066.1)] [Medline: [38486516](https://pubmed.ncbi.nlm.nih.gov/38486516/)]
25. Alarifi BN, Song S. Online vs in-person learning in higher education: effects on student achievement and recommendations for leadership. *Humanit Soc Sci Commun* 2024;11(1). [doi: [10.1057/s41599-023-02590-1](https://doi.org/10.1057/s41599-023-02590-1)]
26. Bahattab A, Zain O, Linty M, Amat Camacho N, Von Schreeb J, Hubloue I, et al. Development and evaluation of scenario-based e-simulation for humanitarian health training: a mixed-methods action research study. *BMJ Open* 2024;14(8):e079681 [FREE Full text] [doi: [10.1136/bmjopen-2023-079681](https://doi.org/10.1136/bmjopen-2023-079681)] [Medline: [39107011](https://pubmed.ncbi.nlm.nih.gov/39107011/)]
27. Major Incident Medical Management and Support (MIMMS) course. URL: https://www.coemed.org/courses/course_desc/mimms [accessed 2024-01-03]
28. Health Emergency and Disaster Risk Management Framework.: WHO; 2019. URL: <https://www.who.int/publications/item/health-emergency-and-disaster-risk-management-framework> [accessed 2024-01-03]
29. Naidoo M, Lee J, Trelles M, Wallis L, Chu KM. Preventing avoidable hospital admissions after emergency care in humanitarian settings: a cross-sectional review of Médecins Sans Frontières emergency departments. *BMJ Open* 2021;11(7):e049785 [FREE Full text] [doi: [10.1136/bmjopen-2021-049785](https://doi.org/10.1136/bmjopen-2021-049785)] [Medline: [34257097](https://pubmed.ncbi.nlm.nih.gov/34257097/)]
30. Desmond M, Schwengel D, Chilson K, Rusy D, Ingram K, Ambardekar A, et al. Paediatric patients in mass casualty incidents: a comprehensive review and call to action. *Br J Anaesth* 2022;128(2):e109-e119 [FREE Full text] [doi: [10.1016/j.bja.2021.10.026](https://doi.org/10.1016/j.bja.2021.10.026)] [Medline: [34862001](https://pubmed.ncbi.nlm.nih.gov/34862001/)]
31. Schaefer T, Nunez L. Burn resuscitation and management. Orlando, FL: StatPearls Publishing; Oct 03, 2024.
32. Hughes A, Almeland SK, Leclerc T, Ogura T, Hayashi M, Mills J, et al. Recommendations for burns care in mass casualty incidents: WHO Emergency Medical Teams Technical Working Group on Burns (WHO TWGB) 2017-2020. *Burns* 2021;47(2):349-370 [FREE Full text] [doi: [10.1016/j.burns.2020.07.001](https://doi.org/10.1016/j.burns.2020.07.001)] [Medline: [33041154](https://pubmed.ncbi.nlm.nih.gov/33041154/)]

33. Suwalowska H, Amara F, Roberts N, Kingori P. Ethical and sociocultural challenges in managing dead bodies during epidemics and natural disasters. *BMJ Glob Health* 2021 Nov;6(11) [FREE Full text] [doi: [10.1136/bmigh-2021-006345](https://doi.org/10.1136/bmigh-2021-006345)] [Medline: [34740913](#)]

34. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;337:a1655 [FREE Full text] [doi: [10.1136/bmj.a1655](https://doi.org/10.1136/bmj.a1655)] [Medline: [18824488](#)]

35. Glasgow RE, Emmons KM. How can we increase translation of research into practice? Types of evidence needed. *Annu Rev Public Health* 2007;28:413-433. [doi: [10.1146/annurev.publhealth.28.021406.144145](https://doi.org/10.1146/annurev.publhealth.28.021406.144145)] [Medline: [17150029](#)]

36. Corpuz JCG. A dignified death: management of dead bodies during COVID-19. *J Public Health (Oxf)* 2022 Jun 27;44(2):e281-e282 [FREE Full text] [doi: [10.1093/pubmed/fdab182](https://doi.org/10.1093/pubmed/fdab182)] [Medline: [34096608](#)]

37. French D, Miles L, Elbourne D, Farmer A, Gulliford M, Locock L, et al. Reducing bias in trials from reactions to measurement: the MERIT study including developmental work and expert workshop. *Health Technol Assess* 2021;25(55):1-72 [FREE Full text] [doi: [10.3310/hta25550](https://doi.org/10.3310/hta25550)] [Medline: [34553685](#)]

38. Han SJ, Chun J. Validation of the disaster preparedness evaluation tool for nurses—the korean version. *Int J Environ Res Public Health* 2021;18(3) [FREE Full text] [doi: [10.3390/ijerph18031348](https://doi.org/10.3390/ijerph18031348)] [Medline: [33540899](#)]

39. Wang J, Lu S, Sun X, Wang F, Wan M, Chen H, et al. Psychometric evaluation of the disaster preparedness evaluation tool (dpet) on emergency nurses in mainland China: two cross-sectional studies. *Disaster Med Public Health Prep* 2022;16(3):1083-1090. [doi: [10.1017/dmp.2021.39](https://doi.org/10.1017/dmp.2021.39)] [Medline: [33949305](#)]

40. Sobaih AEE, Palla IA, Baquee A. Social media use in e-learning amid COVID 19 pandemic: Indian students' perspective. *Int J Environ Res Public Health* 2022;19(9) [FREE Full text] [doi: [10.3390/ijerph19095380](https://doi.org/10.3390/ijerph19095380)] [Medline: [35564771](#)]

41. Bahattab A, Trentin M, Hubloue I, Della Corte F, Ragazzoni L. Humanitarian health education and training state-of-the-art: a scoping review. *Front Public Health* 2024;12:1343867 [FREE Full text] [doi: [10.3389/fpubh.2024.1343867](https://doi.org/10.3389/fpubh.2024.1343867)] [Medline: [39135925](#)]

42. Roman M, Plooreanu A. The effectiveness of the emergency eLearning during COVID-19 pandemic. The case of higher education in economics in Romania. *International Review of Economics Education* 2021;37:100218. [doi: [10.1016/j.iree.2021.100218](https://doi.org/10.1016/j.iree.2021.100218)]

43. Maatuk AM, Elberkawi EK, Aljawarneh S, Rashaideh H, Alharbi H. The COVID-19 pandemic and E-learning: challenges and opportunities from the perspective of students and instructors. *J Comput High Educ* 2022;34(1):21-38 [FREE Full text] [doi: [10.1007/s12528-021-09274-2](https://doi.org/10.1007/s12528-021-09274-2)] [Medline: [33967563](#)]

44. Fauzi MA. E-learning in higher education institutions during COVID-19 pandemic: current and future trends through bibliometric analysis. *Heliyon* 2022;8(5):e09433 [FREE Full text] [doi: [10.1016/j.heliyon.2022.e09433](https://doi.org/10.1016/j.heliyon.2022.e09433)] [Medline: [35600443](#)]

45. Schilly K, Huhn M, Visker JD, Cox C. Evaluation of a disaster preparedness curriculum and medical students' views on preparedness education requirements for health professionals. *Disaster Med Public Health Prep* 2024;18:e8. [doi: [10.1017/dmp.2023.230](https://doi.org/10.1017/dmp.2023.230)] [Medline: [38282524](#)]

46. Xia Y, Hu Y, Wu C, Yang L, Lei M. Challenges of online learning amid the COVID-19: college students' perspective. *Front Psychol* 2022;13:1037311 [FREE Full text] [doi: [10.3389/fpsyg.2022.1037311](https://doi.org/10.3389/fpsyg.2022.1037311)] [Medline: [36619114](#)]

47. Doumi R, Alassaf W, Aljahany M, Aldobeaban S, Mirghani HA, BinSaleh Y, et al. The effectiveness and benefits of disaster simulation training for undergraduate medical students in Saudi Arabia. *Adv Med Educ Pract* 2024;15:707-714 [FREE Full text] [doi: [10.2147/AMEP.S453879](https://doi.org/10.2147/AMEP.S453879)] [Medline: [39050119](#)]

48. Mubarak AF, Amiruddin R, Gaus S. The effectiveness of disaster prevention and mitigation training for the students in disaster prone areas. *IOP Conf. Ser.: Earth Environ. Sci* 2019;235:012055. [doi: [10.1088/1755-1315/235/1/012055](https://doi.org/10.1088/1755-1315/235/1/012055)]

49. Barrimah I, Adam I, Al-Mohaimeed A. Disaster medicine education for medical students: is it a real need? *Med Teach* 2016;38 Suppl 1:S60-S65. [doi: [10.3109/0142159X.2016.1142515](https://doi.org/10.3109/0142159X.2016.1142515)] [Medline: [26984036](#)]

50. Tipton E, Hallberg K, Hedges LV, Chan W. Implications of small samples for generalization: adjustments and rules of thumb. *Eval Rev* 2017 Oct;41(5):472-505. [doi: [10.1177/0193841X16655665](https://doi.org/10.1177/0193841X16655665)] [Medline: [27402612](#)]

51. Stuart EA, Bradshaw CP, Leaf PJ. Assessing the generalizability of randomized trial results to target populations. *Prev Sci* 2015;16(3):475-485 [FREE Full text] [doi: [10.1007/s11121-014-0513-z](https://doi.org/10.1007/s11121-014-0513-z)] [Medline: [25307417](#)]

Abbreviations

DM: disaster medicine

LUFMS: Lebanese University Faculty of Medical Sciences

F2F: face-to-face

ICRC: International Committee of the Red Cross

Edited by T Gladman; submitted 11.Jul.2025; peer-reviewed by H Hanada, J Wilkinson; comments to author 03.Nov.2025; revised version received 05.Dec.2025; accepted 14.Dec.2025; published 28.Jan.2026.

Please cite as:

Msheik A, Peralta R, Al Mokdad Z, Bartulec C, Abou-Abbas L, Anani D, Al Mawla H, Salame H, Moussa M, Bahattab A. Disaster Medicine Training for Medical Students in Lebanon: Quasi-Experimental Comparison of e-Learning and Face-to-Face Modalities. *JMIR Med Educ* 2026;12:e80409

URL: <https://mededu.jmir.org/2026/1/e80409>

doi:10.2196/80409

PMID:

©Ali Msheik, Ruben Peralta, Zeinab Al Mokdad, Christine Bartulec, Linda Abou-Abbas, Dima Anani, Hussein Al Mawla, Hassan Salame, Mohamad Moussa, Awsan Bahattab. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 28.Jan.2026. 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.

Enhancing Engagement With Endocrine Guidelines and Fostering Medical Student Interest Through Concise Medical Information Cines: Qualitative Co-Design Study

Rahul Sagu¹, MBChB, FHEA; Sangamithra Ravi¹, BSc FHEA; Aspasia Manta², MSc, MD; Punith Kempegowda^{2,3}, MSc, MBBS, MD, PhD, SFHEA, FRCP

¹Birmingham Medical School, College of Medicine and Health, University of Birmingham, Birmingham, United Kingdom

²Department of Applied Health Sciences, School of Health Sciences, College of Medicine and Health, University of Birmingham, Birmingham, United Kingdom

³Department of Endocrinology and Diabetes, Queen Elizabeth Hospital Birmingham, University Hospitals Birmingham NHS Foundation Trust, Mindelsohn Way, Birmingham, United Kingdom

Corresponding Author:

Punith Kempegowda, MSc, MBBS, MD, PhD, SFHEA, FRCP

Department of Applied Health Sciences, School of Health Sciences, College of Medicine and Health, University of Birmingham, Birmingham, United Kingdom

Abstract

Background: There is a need to modernize the dissemination of clinical guidelines, making them more accessible and engaging for health care professionals. Concise Medical Information Cines (CoMICs) are peer-reviewed videos created by medical students that distill complex guidelines into learner-friendly visuals.

Objective: This study aimed to describe the process of co-designing an audiovisual version of a clinical guideline and explore the experiences of co-designing audiovisual guideline summaries using the CoMICs model.

Methods: A 4-part CoMICs series on glucocorticoid-induced adrenal insufficiency was codeveloped by clinicians and medical students through 10 iterative steps. A patient version of these CoMICs was then created in multiple languages. Semistructured interviews with authors, reviewers, and student collaborators assessed the clarity, usability, trustworthiness, and educational value of these CoMICs. Reflexive thematic analysis then identified key themes.

Results: CoMICs improved guideline accessibility, comprehension, and global adaptability, while the collaborative process promoted interdisciplinary learning and underscored the efficacy of audiovisual tools for complex content. Student collaborators reported greater confidence in interpreting and communicating clinical guidance, renewed interest in endocrinology, and a deeper appreciation of its academic dimensions.

Conclusions: Cocreating audiovisual resources, such as CoMICs, enhances guideline dissemination. Student involvement can foster curiosity, encourage academic career pathways, and reshape engagement with evidence-based medicine.

(*JMIR Med Educ* 2026;12:e83711) doi:10.2196/83711

KEYWORDS

guidelines; education; endocrinology; dissemination; implementation

Introduction

Clinical practice guidelines are a cornerstone of evidence-based medicine, offering systematically developed recommendations to optimize patient care [1]. Despite their importance, ensuring that guidelines are effectively disseminated and implemented in clinical settings remains a persistent challenge. Health care professionals often face barriers such as time constraints, cognitive overload, lack of awareness of updates, and perceived limited contextual relevance—factors that reduce the usability of traditional long-form, text-based documents in fast-paced environments [2].

Dissemination refers to the targeted distribution of evidence-based knowledge and materials to specific audiences to increase awareness and understanding [3]. Conventional dissemination approaches, such as printed summaries, educational events, audit and feedback loops, and reminder systems, have shown variable success [4]. While multifaceted interventions tend to outperform single strategies, their implementation is often constrained by organizational resources, and they fail to achieve consistent improvements in clinical behaviors or patient outcomes. Implementation, in this context, involves systematically integrating that knowledge into real-world practice to improve health care outcomes. More importantly, these strategies frequently overlook the importance

of engagement, accessibility, and real-time utility—features that are increasingly essential in today's diverse and dynamic health care systems [5]. These limitations underscore the need for supplementary dissemination strategies that can deliver key recommendations in more accessible, engaging, and time-efficient formats [1].

Emerging audiovisual tools offer an opportunity to bridge this gap. By reducing cognitive load, enhancing recall, and improving user engagement, visual formats can complement traditional learning formats [6]. One such approach is Concise Medical Information Cines (CoMICs)—short, peer-reviewed, visually engaging videos designed to translate complex clinical guidance into digestible content [7]. Using infographics, narration, and visual storytelling, CoMICs aim to make guidelines more approachable and usable for both early-career and experienced professionals.

In parallel with growing calls to diversify how students experience specialty fields and engage with research, coproducing educational resources such as CoMICs allows students to gain exposure to academic medicine in a hands-on, creative manner. The Endocrine Society's Medical School Engagement Program similarly seeks to broaden medical students' exposure to endocrinology [8]. By aligning with the goals of the Medical School Engagement Program, CoMICs offer a complementary, hands-on platform that both disseminates core guideline content and actively engages students in the specialty. While the earlier CoMICs study investigated the dissemination and engagement patterns of specialty-focused educational videos on social media platforms, the translation of a clinical guideline into a structured audiovisual format has not been examined in previous CoMICs research. Therefore, in this study, we describe the process of codeveloping a 4-part CoMICs series based on the glucocorticoid-induced adrenal insufficiency (GIAI) guideline, showcasing an innovative approach to translating clinical guidance into accessible resources. We also explored medical students' experiences of co-designing the CoMICs and how this influenced their learning.

Methods

Overview

This study was conducted between October 2024 and May 2025 at the University of Birmingham. The process of CoMICs creation, aligned with the knowledge-to-action framework, has been described in detail in our previous publication [7]. To inform the adoption of CoMICs for guideline dissemination, we conducted a structured exploratory literature review using PubMed. The search strategy focused on 3 intersecting themes: dissemination methods, barriers to the uptake of clinical guidelines, and health care professional behaviors. The review was limited to English-language articles involving health care professionals but included studies from all publication years to ensure comprehensive coverage. Key findings highlighted the limitations of traditional dissemination techniques, such as printed materials, educational sessions, and audit-feedback cycles, and underscored the need for more engaging and accessible formats, particularly in time-pressed clinical environments [4]. These insights shaped the rationale for the

CoMICs approach. A 4-part audiovisual series was created to convey the core content of the Joint European Society of Endocrinology and Endocrine Society guideline on GIAI. The primary focus of this work is qualitative, centered on the co-design and evaluation of CoMICs. The literature review served as a supporting framework for this study. Each video distilled key diagnostic, therapeutic, and follow-up recommendations into concise, visually engaging episodes. Crucially, medical students played an active role throughout the ten-step CoMICs development process, which included the following:

1. The CoMICs team contacted the corresponding guideline author to establish collaboration, define the scope of the project, and confirm participation in an iterative review.
2. Guideline authors shared supplementary educational resources (eg, slide decks, figures, explanatory notes, or key references) to support accurate interpretation of the guideline.
3. The CoMICs team developed an initial script and visual storyboard using PowerPoint (Microsoft Corp), translating key guideline recommendations into simplified, learner-friendly visual narratives.
4. The storyboard underwent clinical review by guideline authors, who provided feedback on accuracy, emphasis, and clarity, after which suggested revisions were incorporated.
5. A preliminary video draft without narration or background audio was produced to enable focused review of structure, visuals, and content flow.
6. Feedback from guideline authors, clinical experts, and peer reviewers was collated by the CoMICs team and used to refine content, visuals, and sequencing.
7. A prefinal version incorporating narration and refined visuals was produced and circulated to the authors for final clinical validation.
8. Final comments and minor revisions were addressed to ensure consistency with the guideline and clarity for the intended audience.
9. The completed CoMIC was approved by the corresponding author and prepared for professional use.
10. Approved videos were disseminated through clinical networks and social media platforms to maximize accessibility, reach, and engagement [9].

This process not only supported the technical production of CoMICs but also served as a structured educational opportunity, exposing students to academic publishing, clinical content curation, and interdisciplinary teamwork in endocrinology. The CoMICs videos were disseminated through multiple online platforms to maximize reach and engagement. All videos were uploaded to YouTube (Google LLC) as the primary hosting platform and shared via social media channels, including X (X Corp, formerly Twitter), Instagram (Meta Platforms, Inc), and WhatsApp (Meta Platforms, Inc) groups within the CoMICs network. The posts were accompanied by short educational captions and reposted by CoMICs members on the aforementioned platforms to enhance discoverability.

Multilingual Patient-Focused CoMICs Development

In parallel, a patient-facing version of the CoMICs series was developed to improve health literacy and access among lay audiences. These 4 short videos were co-designed with individuals living with adrenal insufficiency and deliberately avoided technical language. Patient advocates provided feedback throughout script development, ensuring clarity, relevance, and the inclusion of relatable, everyday examples. Practical guidance on symptom recognition, treatment adherence, and self-management was integrated.

To further enhance accessibility, multilingual versions were created. Medical students fluent in English and another language of interest to this project produced initial transcripts, which were then reviewed by native speakers with clinical or health communication experience. This ensured that idiomatic clarity, cultural appropriateness, and medical accuracy were preserved across language versions. Languages for the multilingual video adaptations were selected pragmatically based on the connections within the CoMICs network and the availability of fluent volunteer translators. This approach ensured linguistic accuracy while allowing culturally relevant adaptations for each audience. This process enhanced the reach and relevance of the CoMICs series to non-English-speaking audiences, promoting global health equity [10].

Qualitative Interviews to Explore Experience and Impact

Purposive sampling was used to recruit participants with direct involvement in the CoMICs project or relevant expertise. These included the following:

- Medical students and health care professionals involved in the CoMICs cocreation
- Clinical experts who authored or peer-reviewed the GIAI guideline

Among the CoMICs team members interviewed, there were 10 medical students and 2 physicians. The medical students were in their penultimate or final year of study, while the 2 physicians included 1 foundation physician and 1 honorary research clinical fellow. All participants demonstrated an interest in endocrinology, having been active members of the CoMICs initiative for at least 1 year and engaged in ongoing development of the CoMICs guideline. Their participation reflected both their exploratory career interests and commitment to evidence-based knowledge dissemination.

Semistructured interviews explored participants' experiences with clinical guidelines, perceptions of the CoMICs format, and the personal or professional impact of their involvement. Students were also asked to reflect on how the experience shaped their views on endocrinology, academic collaboration, and the role of creative dissemination in medicine. The interview guide was aligned with the 2 objectives of the study and focused on clarity, usability, coproduction dynamics, and suggestions for future improvement. All interviews were audio-recorded and transcribed verbatim. Saturation was considered reached when participants could no longer provide additional insights or perspectives relevant to the research questions.

Ethical Considerations

The study was reviewed and received ethics approval from the Science, Technology, Engineering, and Mathematics Committee at the University of Birmingham (ERN_2965-Aug2024). All data were anonymized to safeguard the participants' information. Written informed consent was obtained from all participants before the interview, in accordance with institutional guidance and the approved ethics protocol. Participation was voluntary, and participants were not compensated for their time.

Data Analysis

Interview data were analyzed using reflexive thematic analysis following the six-phase framework by Braun and Clarke [10]:

1. Familiarization: transcripts were reviewed in depth, with initial impressions documented.
2. Generating initial codes: RS undertook repeated familiarization with the transcripts before manually coding meaningful segments of text. Initial codes were developed inductively and generated directly from the data at a semantic level. Coding was data-driven and iterative, with reflexive notes maintained throughout to document analytic decisions and support ongoing refinement. Preliminary codes were shared with the research team, and team discussions informed the refinement and consolidation of the initial coding set.
3. Constructing initial themes: related codes were clustered to form preliminary candidate themes.
4. Reviewing themes: themes were refined for coherence and distinctiveness, with weak or overlapping themes revised or removed.
5. Defining and naming themes: final themes were clearly defined, named, and supported with representative quotations through group discussion. Generative artificial intelligence assistance was used only to refine subthemes, and all analytic decisions were confirmed through team consensus.
6. Producing the report: themes were organized in relation to the study objectives and integrated with relevant literature. This approach enabled a nuanced understanding of the practical and pedagogical value of audiovisual dissemination and provided insight into how cocreation experiences can foster medical students' interest in specialty areas and academic pathways. The themes were contextualized using findings from the literature review, highlighting the potential for CoMICs to address longstanding dissemination challenges while also serving as a platform for early academic engagement.

Results

Overview

The 4-part guideline CoMICs series for health care professionals has accumulated 1218 views on YouTube and 1331 views on X (as of November 22, 2025) since its creation. The patient and public versions were created in English, Bengali, Serbian, Tamil, Greek, Georgian, and Brazilian Portuguese and have been viewed 1884 times on YouTube and 2420 times on X (as of November 22, 2025) since their creation. The final CoMICs

videos are publicly accessible via YouTube ([Multimedia Appendix 1](#)).

Thematic Analysis of Interviews

Overview

A total of 15 participants—12 (80%) medical students and 3 (20%) senior health care professionals—involved in the

Table . Thematic analysis results.

Theme and subtheme	Code
Accessibility and usability	<ul style="list-style-type: none"> • Use NICE^a for guidelines • Written guidelines are too long • Prefer short videos • Guidelines difficult to navigate in clinical settings • Multilingual versions broaden access • Supports diverse learner backgrounds
Visual and cognitive engagement	
Quick refreshers	<ul style="list-style-type: none"> • Use videos as refreshers before clinics • Videos used during breaks or on commutes
Clinician use cases	<ul style="list-style-type: none"> • Visuals improve learning and retention • Animations clarify physiology and treatments
Credibility and trust	<ul style="list-style-type: none"> • Expert review adds legitimacy • Professional tone builds trust • Referencing and consistency enhance credibility
Empowerment through cocreation	
Workflow friction	<ul style="list-style-type: none"> • Students gained confidence and communication skills • Challenges in feedback and team coordination • Improved time and project management skills
Inclusivity and cultural reach	<ul style="list-style-type: none"> • Multilingual versions increase reach • Non-English speakers benefit from translations • Can be shared with patients from diverse backgrounds

^aNICE: National Institute for Health and Care Excellence.

Accessibility and Usability

Participants reported that traditional guidelines were often inaccessible in time-constrained settings, particularly during clinical work or exam preparation. Students and clinicians alike praised CoMICs as practical alternatives that distilled core messages into short, visually engaging formats. Furthermore, participants highlighted that multilingual CoMICs videos broadened accessibility and usability for parents and learners from diverse linguistic backgrounds, enhancing inclusivity and reach, especially where written English guidelines pose barriers:

Even just a quick 2-minute summary video is more useful than scrolling through 30 pages. [Medical student]

development or review of the GIAI CoMICs series were interviewed. Thematic analysis revealed 5 overarching themes with associated subthemes and codes ([Table 1](#)), with supporting illustrative quotes.

Visual and Cognitive Engagement

CoMICs' design—combining narration, visual metaphors, and sequential storytelling—was credited with enhancing comprehension and retention. Students appreciated the ability to review content quickly and repeatedly, while health care professionals used the videos as just-in-time learning tools before clinical duties:

The animations helped me understand the physiology without flipping through the textbook. [CoMICs student team member]

Credibility and Trust

The rigorous review process, which involved guideline authors and clinical experts, was critical to establishing credibility.

Students valued knowing their contributions would be clinically validated, while reviewers endorsed the professional standard of the final product:

If you know the guideline authors reviewed it, it feels legitimate—not just a student project. [Medical student]

Empowerment Through Cocreation

Student participants expressed a strong sense of empowerment, seeing their work translated into tangible educational resources. They highlighted growth in communication, critical thinking, and teamwork skills. The experience also offered early exposure to academic endocrinology and knowledge translation, reinforcing interest in future academic careers:

I never thought a video I helped make would be used by actual doctors. [Medical student]

Despite some communication delays and revision bottlenecks, students viewed the challenges as valuable professional learning opportunities.

Inclusivity and Cultural Reach

Multilingual adaptations were particularly valued for expanding access. Students noted that having videos in multiple languages allowed them to share resources with patients and peers from diverse backgrounds, enhancing equity in education and patient engagement:

The Bangla version is great—I could share it with patients too. [Medical student]

This inclusive element positioned CoMICs as a globally relevant tool that could strengthen both practitioner education and patient-facing communication.

Discussion

Principal Findings

This study represents a novel contribution to the field of clinical guideline dissemination and medical education. To the best of our knowledge, it is among the first to explore the cocreation of audiovisual, narrative-based clinical resources—CoMICs—that are not only grounded in evidence-based guidelines but also designed for multilingual and multicultural relevance. Unlike previous CoMICs initiatives that focused on broad health topics and social media engagement, this study uniquely evaluates the transformation of a single clinical guideline into structured audiovisual outputs and examines the educational impact of the cocreation process on medical students. The visual structure of CoMICs supports the cognitive load theory by minimizing extraneous load and enabling learners to process guideline content more efficiently [11]. Previous reviews have emphasized the limited effectiveness of traditional dissemination methods (eg, printed guidelines and lectures) unless paired with contextual relevance, interactivity, and credible authorship [5,7,12,13]. CoMICs address these limitations through visual storytelling, coproduction with guideline authors, and multilingual delivery, making them a more adaptable and engaging format for real-world use. This aligns with emerging innovative

dissemination practices, which emphasize concise, visually driven formats to improve the reach and uptake of health evidence [14].

Furthermore, this study supports growing calls in the health care education literature for multifaceted, user-driven dissemination models [15]. Notably, CoMICs serve not only as tools to improve practitioner guideline uptake but also as platforms for early-career professional development. Medical students involved in the production process gained firsthand experience in academic endocrinology, communication, interprofessional collaboration, and health education. Such active involvement can foster future interest in both clinical specialties and research careers—an impact rarely explored in the dissemination literature.

Our findings can be contextualized within dissemination science frameworks. The Framework for Knowledge Transfer by Lavis et al [16] proposes that effective dissemination depends on 5 core questions: what should be transferred, to whom, by whom, how, and with what effect. CoMICs directly addresses each of these domains: it condenses evidence-based guideline messages (what), targets medical students and patients (to whom), involves credible experts and peers (by whom), delivers content through short multilingual videos and social media (how), and measures engagement and understanding (with what effect). By mapping CoMICs to these principles, our approach provides a structured, theoretically grounded model for disseminating evidence through digital media.

The strength of this study lies in its dual-impact design, which simultaneously addresses the challenges of guideline dissemination and supports the development of future academic clinicians. The cocreative model provided students with meaningful, mentored engagement in real-world implementation efforts, resulting in a finished product with practical utility for health care professionals and patients alike. The inclusion of multilingual and culturally relevant content further extends CoMICs' global reach, especially in resource-limited or linguistically diverse health care settings. Additionally, by capturing reflections from both clinical experts and student collaborators, the study offers a rich, balanced qualitative perspective on how innovation, collaboration, and mentorship can enhance both educational quality and clinical relevance.

Limitations

We did not measure actual changes in clinical behavior or patient outcomes. While the findings indicate strong potential, objective evidence of impact remains limited. Moreover, because the interviewees were directly involved in developing the resources, positive bias may have influenced responses, potentially limiting the generalizability of the findings. It is also important to note that quantitative metrics, such as video views and engagement figures, were included solely to contextualize the qualitative findings, as the primary aim of this study was to explore the co-design experience rather than to evaluate dissemination outcomes. Furthermore, evaluating the broader impact of CoMICs on students and patients who viewed the content was beyond the scope of this work. Future research [1-7] should focus on guideline adherence, decision-making confidence, and quality of patient care. Several user-generated

insights, such as using the videos as “quick refreshers before clinics” or “during breaks or commutes,” suggest practical pathways for future implementation. These patterns indicate that CoMICs could be integrated into just-in-time learning strategies, on-shift clinical preparation, and student revision routines. Future work should examine how such informal use cases can be systematically incorporated into guideline uptake workflows and clinical education pathways.

Conclusions

CoMICs offer a practical and engaging solution to barriers in guideline dissemination, improving accessibility and relevance through visual storytelling and multilingual adaptation. Importantly, involving medical students in their creation fosters early exposure to academic endocrinology and professional skill development, supporting interest in research and specialty careers. Future work should focus on evaluating clinical outcomes, cost-effectiveness, and integration into wider implementation systems, ensuring scalability across specialties and settings.

Acknowledgments

The authors gratefully acknowledge the DEVI (dissemination, education & value-added implementation) collaboration—Josh Banerjee, Pranav Viswanath Iyer, Vardhan Venkatesh, and Mukunth Kowsik—for their contributions to the production of the guideline Concise Medical Information Cines (CoMICs). They also thank Cristina Ronchi and Leonie van Hulsteijn for their valuable review of the guideline CoMICs during development. They are also grateful to the Department of Applied Health Sciences, School of Health Sciences, College of Medicine and Health, University of Birmingham, for supporting this study. We used the generative artificial intelligence tool (GPT-5.1) [17] to refine the language and grammar during manuscript preparation, which were further reviewed and revised by the study group.

Funding

This research received no particular grant from any funding agency in the public, commercial, or not-for-profit sectors. PK received support from the National Institute for Health and Care Research (NIHR) through his Advanced Clinician Scientist Fellowship (grant NIHR303671). Additionally, he was supported by the Midlands Patient Safety Research Collaboration and the NIHR Race, Equity, and Diversity in Careers Incubator. The views expressed in this study are those of the author and do not necessarily reflect the official positions of the NIHR or the Department of Health and Social Care.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: PK (lead), RS (equal), AM (supporting).

Data curation: RS (lead), PK (supporting).

Formal analysis: RS (lead), PK (supporting).

Funding acquisition: PK.

Investigation: RS (lead), PK (equal), SR (supporting).

Methodology: RS (lead), PK (equal).

Project administration: RS (lead), PK (equal).

Resources: PK (lead), RS (equal), SR (supporting).

Supervision: PK (lead), AM (supporting).

Validation: RS (lead), PK (equal), SR (supporting), AM (supporting).

Visualization: RS (lead), PK (supporting), SR (supporting).

Writing – original draft: RS (lead).

Writing – review & editing: RS (lead), PK (equal), SR (supporting), AM (supporting).

Conflicts of Interest

None declared.

Multimedia Appendix 1

YouTube hyperlinks to the 4-part glucocorticoid-induced adrenal insufficiency guideline Concise Medical Information Cines series and the 7 multilingual patient versions.

[[DOCX File, 15 KB - mededu_v12i1e83711_app1.docx](#)]

References

1. Panteli D, Legido-Quigley H, Reichebner C, Ollenschläger G, Schäfer C, Busse R. Clinical practice guidelines as a quality strategy. In: Busse R, Klazinga N, Panteli D, editors. Improving Healthcare Quality in Europe: Characteristics, Effectiveness and Implementation of Different Strategies: European Observatory on Health Systems and Policies; 2019.
2. McArthur C, Bai Y, Hewston P, Giangregorio L, Straus S, Papaioannou A. Barriers and facilitators to implementing evidence-based guidelines in long-term care: a qualitative evidence synthesis. *Implement Sci* 2021 Jul 9;16(1):70. [doi: [10.1186/s13012-021-01140-0](https://doi.org/10.1186/s13012-021-01140-0)] [Medline: [34243789](#)]
3. Owoeye OB, Rauvola RS, Brownson RC. Dissemination and implementation research in sports and exercise medicine and sports physical therapy: translating evidence to practice and policy. *BMJ Open Sport Exerc Med* 2020 Nov 13;6(1):e000974. [doi: [10.1136/bmjssem-2020-000974](https://doi.org/10.1136/bmjssem-2020-000974)] [Medline: [33304607](#)]
4. 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](https://doi.org/10.1002/14651858.CD006732.pub4)] [Medline: [30025154](#)]
5. Squires JE, Sullivan K, Eccles MP, Worswick J, Grimshaw JM. Are multifaceted interventions more effective than single-component interventions in changing health-care professionals' behaviours? An overview of systematic reviews. *Implement Sci* 2014 Oct 6;9:152. [doi: [10.1186/s13012-014-0152-6](https://doi.org/10.1186/s13012-014-0152-6)] [Medline: [25287951](#)]
6. Allen Moore W, Russell Smith A. Effects of video podcasting on psychomotor and cognitive performance, attitudes and study behaviour of student physical therapists. *Innov Educ Teach Int* 2012;49(4):401-414. [doi: [10.1080/14703297.2012.728876](https://doi.org/10.1080/14703297.2012.728876)]
7. Elhairy M, Malhotra K, Goyal K, Bardus M, Simba And CoMICs, Kempegowda P. A SIMBA CoMICs initiative to cocreating and disseminating evidence-based, peer-reviewed short videos on social media: mixed methods prospective study. *JMIR Med Educ* 2024 Oct 30;10:e52924. [doi: [10.2196/52924](https://doi.org/10.2196/52924)] [Medline: [39475500](#)]
8. Medical school engagement program. Endocrine Society. 2025. URL: <https://www.endocrine.org/our-community/expanding-field/medical-school-engagement-program> [accessed 2025-08-18]
9. Venkatesh V, Mathiopoulou M, Kumar K, et al. Multilingual menopause comics: combating misinformation and enhancing global health literacy through peer-reviewed educational videos. *Endocr Abstr* 2025;110:1076. [doi: [10.1530/endoabs.110.P1076](https://doi.org/10.1530/endoabs.110.P1076)]
10. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006 Jan;3(2):77-101. [doi: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)]
11. Young JQ, Van Merriënboer J, Durning S, Ten Cate O. Cognitive Load Theory: implications for medical education: AMEE Guide No. 86. *Med Teach* 2014 May;36(5):371-384. [doi: [10.3109/0142159X.2014.889290](https://doi.org/10.3109/0142159X.2014.889290)] [Medline: [24593808](#)]
12. van der Wees PJ, Jamtvedt G, Rebbeck T, de Bie RA, Dekker J, Hendriks EJ. Multifaceted strategies may increase implementation of physiotherapy clinical guidelines: a systematic review. *Aust J Physiother* 2008;54(4):233-241. [doi: [10.1016/s0004-9514\(08\)70002-3](https://doi.org/10.1016/s0004-9514(08)70002-3)] [Medline: [19025503](#)]
13. Prior M, Guerin M, Grimmer-Somers K. The effectiveness of clinical guideline implementation strategies--a synthesis of systematic review findings. *J Eval Clin Pract* 2008 Oct;14(5):888-897. [doi: [10.1111/j.1365-2753.2008.01014.x](https://doi.org/10.1111/j.1365-2753.2008.01014.x)] [Medline: [19018923](#)]
14. Ross-Hellauer T, Tennant JP, Banelyté V, et al. Ten simple rules for innovative dissemination of research. *PLoS Comput Biol* 2020 Apr;16(4):e1007704. [doi: [10.1371/journal.pcbi.1007704](https://doi.org/10.1371/journal.pcbi.1007704)] [Medline: [32298255](#)]
15. Chan WV, Pearson TA, Bennett GC, et al. ACC/AHA special report: clinical practice guideline implementation strategies: a summary of systematic reviews by the NHLBI implementation science work group: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol* 2017 Feb 28;69(8):1076-1092. [doi: [10.1016/j.jacc.2016.11.004](https://doi.org/10.1016/j.jacc.2016.11.004)] [Medline: [28132746](#)]
16. Lavis JN, Robertson D, Woodside JM, McLeod CB, Abelson J, Knowledge Transfer Study Group. How can research organizations more effectively transfer research knowledge to decision makers? *Milbank Q* 2003;81(2):221-248. [doi: [10.1111/1468-0009.t01-1-00052](https://doi.org/10.1111/1468-0009.t01-1-00052)] [Medline: [12841049](#)]
17. ChatGPT 5.1. OpenAI. URL: <https://openai.com/index/gpt-5-1/> [accessed 2025-05-25]

Abbreviations

CoMICs: Concise Medical Information Cines

GIAI: glucocorticoid-induced adrenal insufficiency

Edited by A Stone; submitted 10.Sep.2025; peer-reviewed by C Provvidenza, J Kaur, MS Shafi; revised version received 22.Dec.2025; accepted 24.Dec.2025; published 27.Jan.2026.

Please cite as:

Sagu R, Ravi S, Manta A, Kempegowda P

Enhancing Engagement With Endocrine Guidelines and Fostering Medical Student Interest Through Concise Medical Information Cines: Qualitative Co-Design Study

JMIR Med Educ 2026;12:e83711

URL: <https://mededu.jmir.org/2026/1/e83711>

doi: [10.2196/83711](https://doi.org/10.2196/83711)

© Rahul Sagu, Sangamithra Ravi, Aspasia Manta, Punith Kempegowda. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 27.Jan.2026. 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.

Original Paper

Web-Based Virtual Environment Versus Face-To-Face Delivery for Team-Based Learning of Anesthesia Techniques Among Undergraduate Medical Students: Randomized Controlled Trial

Darunee Sripadungkul¹, MD; Suhattaya Boonmak¹, MD; Monsicha Somjit¹, MD; Narin Plailaharn¹, MD; Wimonrat Sriraj¹, MD; Polpun Boonmak¹, MD

Department of Anesthesiology, Faculty of Medicine, Khon Kaen University, Muang Khon Kaen, Khon Kaen, Thailand

Corresponding Author:

Suhattaya Boonmak, MD

Department of Anesthesiology, Faculty of Medicine

Khon Kaen University

123 Ni Muang

Muang Khon Kaen, Khon Kaen,
Thailand

Phone: 66 813698630

Email: suhabo@kku.ac.th

Abstract

Background: Foundational knowledge of anesthesia techniques is essential for medical students. Team-based learning (TBL) improves engagement. Web-based virtual environments (WBVEs) allow many learners to join the same session in real time while being guided by an instructor.

Objective: This study aimed to compare a WBVE with face-to-face (F2F) delivery of the same TBL curriculum in terms of postclass knowledge and learner satisfaction.

Methods: We conducted a randomized, controlled, assessor-blinded trial at a Thai medical school from August 2024 to January 2025. Eligible participants were fifth-year medical students from the Faculty of Medicine, Khon Kaen University, who attended the anesthesiology course at the department of anesthesiology. Students who had previously completed the anesthesiology course or were unable to comply with the study protocol were excluded. They were allocated to one of the groups using a computer-generated sequence, with concealment of allocation to WBVE (on the Spatial platform) or F2F sessions. Both groups received identical 10-section content in a standardized TBL sequence lasting 130 minutes. Only the delivery mode differed (Spatial WBVE vs classroom F2F). The primary outcome was the postclass multiple-choice questionnaire score. The secondary outcome was learner satisfaction. Individual knowledge was assessed before and after the session using a 15-item questionnaire containing multiple-choice questions via Google Forms. Satisfaction was measured immediately after class on a 5-point Likert scale. Outcome scoring and data analysis were blinded to group assignment. Participants and instructors were not blinded.

Results: In total, 79 students were randomized in this study (F2F: n=38, 48%; WBVE: n=41, 52%). We excluded 2% (1/41) of the students in the WBVE group due to incomplete data. There were complete data for the analysis for 78 participants (F2F: n=38, 49%; WBVE: n=40, 51%). Preclass scores were similar between groups (F2F: mean 6.03, SD 2.05; WBVE: mean 6.20, SD 2.04). Postclass knowledge did not differ significantly (F2F: mean 11.24, SD 1.93; WBVE: mean 10.40, SD 2.62; mean difference 0.88, 95% CI -0.18 to 1.94; $P=.12$). Learner satisfaction favored F2F learning across multiple domains, including overall course satisfaction. Overall satisfaction favored F2F learning (mean difference 0.42, 95% CI 0.07-0.77; $P=.01$). Both groups ran as planned. No adverse events were reported. No technical failures occurred in the WBVE group.

Conclusions: In this trial, WBVE-delivered TBL produced similar short-term knowledge gains to F2F delivery, but learner satisfaction was lower in the WBVE group. Unlike many previous studies, this trial compared WBVE and F2F delivery while keeping the TBL curriculum and prespecified outcomes identical across groups. These findings support WBVEs as a scalable option when physical space, learner volume, or constraints are present. However, lower satisfaction in the WBVE highlights the real-world need for improved facilitation, user experience design, and technical readiness before broader implementation.

Trial Registration: Thai Clinical Trials Registry TCTR20240708012; <https://www.thaicalinicaltrials.org/show/TCTR20240708012>

(*JMIR Med Educ* 2026;12:e80097) doi:[10.2196/80097](https://doi.org/10.2196/80097)

KEYWORDS

anesthesia; computer-assisted instruction; distance; education; internet; learning; medical; problem-based learning; students; teaching; virtual reality

Introduction

Foundational knowledge of anesthesia techniques is important for medical students. Medical students are required to have knowledge of anesthesia administration, pharmacology, procedures, complication management, and interpretation of anesthesia records. Traditional didactic lectures offer a clear, efficient way to organize and deliver this content. It also allows broad coverage of key concepts in a limited time. However, limited interactivity can reduce engagement and encourage passive learning [1]. Interactive lectures address these limitations with small-group discussions, problem-solving exercises, and simulations. These activities create a more active classroom and support clinical reasoning. They also improve knowledge retention and provide immediate feedback to guide learning [2,3].

Team-based learning (TBL) is an effective approach to promote collaborative and active learning. TBL links to constructivist learning theory, the interactive-constructive-active-passive framework, and cognitive load theory. A simple sequence includes individual and team readiness assurance, application exercises, and immediate feedback. Students can move beyond just completing tasks to engaging with the content and peers. Additionally, it encourages a deeper understanding of concepts through discussion. Students can focus on what matters, think more deeply, and retain knowledge longer [4-6]. However, TBL also has its limitations. Challenges include managing group dynamics, resolving conflicts arising from differing opinions, managing time effectively, and distributing the workload. It also requires appropriate activity design, resources, and assessments [3,7,8].

Interactive lectures and TBL are used to facilitate the topic of anesthesia techniques for fifth-year medical students in our department. This learning technique demonstrates motivation, problem-based solving, and communication. Advances in medical education technology, such as web-based virtual environments (WBVEs), can support synchronous interaction, shared workspaces, and simulation-like experiences. WBVEs are virtual worlds that enable real-time interactions between users and digital objects through technologies such as virtual and augmented reality. They provide immersive experiences, facilitate collaboration, and support interactive and realistic simulations [9-13]. Compared with conventional TBL, WBVEs may make classes easier to join, easier to expand to more students, and more supportive of group work [7,12,14].

However, comparative evidence remains limited. Existing studies are constrained by heterogeneous reporting, small sample sizes, and nonrandomized designs, particularly in anesthesia education [15-17]. Moreover, learners' diverse preferences and learning styles should be considered when implementing WBVEs [18]. These limitations justify a randomized trial that isolates delivery mode by comparing WBVE with face-to-face (F2F) instruction under an equivalent TBL design. We

conducted a randomized controlled trial comparing WBVE (on the Spatial platform [19]) versus F2F delivery of the same TBL curriculum in anesthesia techniques among fifth-year medical students. Both groups used identical objectives, materials, facilitation, and assessments. Only the delivery mode differed. The primary outcome was postclass knowledge, and the secondary outcome was learner satisfaction. We hypothesized that there would be between-group differences in postclass knowledge and learner satisfaction.

Methods

Study Design

This study was a randomized, controlled, single-blinded trial. Double blinding was not feasible as participants were aware of their assigned groups. This study adhered to the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines [20] to ensure comprehensive and transparent reporting of randomized controlled trial data.

Participants

We recruited fifth-year medical students from the Faculty of Medicine, Khon Kaen University, who attended the anesthesiology course at the department of anesthesiology. Students who had previously completed the anesthesiology course or were unable to comply with the study protocol were excluded. All participants provided written informed consent. The study was conducted from August 26, 2024, to January 7, 2025. Withdrawal criteria included technical issues, such as internet disruptions or software glitches on the Spatial platform [19].

Randomization and Recruitment

We used cluster randomization at the class level to minimize contamination between teaching groups. Students were grouped into clusters based on their scheduled learning sessions. Each cluster comprised approximately 20 students. Clusters were then randomly allocated in a 1:1 ratio to either the WBVE or F2F group using computer-generated random numbers. We used a block size of 2 to ensure that each group had a similar number of clusters. The randomization sequence was generated and implemented by investigators who were not involved in teaching or assessment to reduce the risk of allocation bias. For clusters randomized to the WBVE group, students who declined research participation received F2F teaching delivered by another instructor, scheduled at the same time and following an identical teaching plan. For clusters randomized to the F2F group, students who declined research participation attended the same F2F session with the same instructor but were not asked to complete the posttest or satisfaction questionnaire.

Interventions

Preintervention Process

After recruitment into the study, all participants in both groups were given access to the learning material via online learning in Google Classroom, including a slide presentation and supplementary video material, 2 days before the teaching session. On the teaching day, students first received a 5-minute briefing about study procedures and learning objectives. They then completed a questionnaire containing 15 multiple-choice questions (MCQs) before the test in 15 minutes to assess baseline knowledge. Participants also completed a brief questionnaire capturing student characteristics. The test specification table is provided in [Multimedia Appendix 1](#).

Session Allocation and Team Setup

After baseline assessments, students attended their assigned session (either a WBVE or an F2F session). In each group, students were organized into 4 TBL teams of 5 to 6 students, supervised by 1 instructor. Each session lasted 130 minutes.

Content and Structure Are Common to Both Groups

Both groups participated in a single, standardized 130-minute TBL session. The content was identical, organized into 10 sections on anesthesia techniques (listed in [Multimedia Appendix 1](#)). Each section began with core knowledge, followed by a problem-solving scenario aligned with the same learning objectives. Students progressed sequentially through the modules to complete the required tasks.

F2F Group

The 10 modules described earlier were delivered F2F using a standard TBL sequence. At the start of the session, the instructor reviewed the learning objectives and ground rules and confirmed

team assignments. For each module ([Multimedia Appendix 1](#)), students first received a brief minilecture to consolidate core concepts. Subsequently, teams completed a scenario-based application task at their respective tables. They recorded a single team answer. They received immediate feedback, followed by a short whole-class debrief that compared rationales across teams. The room layout comprised 4 separated team tables with a central screen, and teams were asked to discuss quietly to minimize cross talk. Materials were identical to those in the other group (handouts mirroring the slides and identical question stems). Any clarifying questions were addressed at the table before the debrief.

WBVE Group

Sessions ran on the Spatial platform (Spatial Systems, Inc; Thai localization) [19] using desktop computers on the university's secure local network. On first log-in, students created an account and entered a virtual space with 3 areas (a classroom, an operating room, and a common room). Each was equipped with detailed 3D anesthesia equipment to support the team's application tasks. The same 10 interactive modules described earlier were delivered in sequence. Each module began with brief core content, followed by a scenario-based application task aligned with the learning objectives. Students moved through the environment with keyboard-controlled avatars and collaborated in real time via built-in voice and text chat. Embedded multimedia (short videos, slide decks, interactive cases, and simple educational games) supported engagement. The instructor monitored team rooms, offered real-time guidance, provided immediate feedback, and answered questions as they arose. Any clarifying questions were addressed in the common room before the debrief. A detailed description of the WBVE is provided in [Figure 1](#).

Figure 1. The web-based virtual environment ran on the Spatial platform: (A) A classroom. Students learned modules 1 to 7 in this classroom, organized into 4 teams of 5 to 6 students each, supervised by 1 instructor. (B) A classroom. Students navigated the environment using keyboard-controlled avatars and collaborated in real time via built-in voice and text chat. Students moved to the operating room by clicking the button (*). (C) An operating room. Students learned modules 8 and 9 in this room, including how to read anesthetic record data. Each group also discussed 4 case studies for team-based learning. (D) An operating room. After finishing the questions, students moved to the common room by clicking the button (*). (E) A common room. Students in each group asked clarifying questions in this room, after which the instructor conducted a debriefing for module 10. If students wanted to review the material again or go to an operating room or a classroom, they could return by clicking the button (*). (F) A common room. A simulated operating room had been created to allow students to learn in a realistic virtual environment before entering the actual operating room. If students had extra time while waiting for their classmates, they could explore this room. Various games were also available to help train their minds.



Postintervention Process

Immediately after the teaching session, 15 minutes were allocated for outcome measurement. All students completed a questionnaire containing 15 MCQs after the test via Google Forms, using the exact test specification as before the test, under invigilated conditions. They then completed a satisfaction questionnaire in Google Forms (5-point Likert scale; instrument and scoring details are provided in [Multimedia Appendix 1](#)). The forms were identical for both groups and were accessed through locked links. Answer keys were withheld until all submissions had been received. In addition, we enabled automatic scoring and time stamp logging, restricted responses to a single submission per account, and exported raw data directly from Google Forms into the analysis dataset.

Data Collection

Pre- and postintervention knowledge evaluations were assessed using a 15-item MCQ test mapped to a test specification table ([Multimedia Appendix 1](#)). The authors designed and developed a questionnaire on students' satisfaction with the learning process, drawing on previous studies [21,22], and adjusted its content for use in a Thai context ([Multimedia Appendix 1](#)). Student satisfaction was evaluated immediately after the intervention using a 21-item survey, with responses on a 5-point Likert scale (1=strongly disagree, 2=disagree, 3=neither agree nor disagree, 4=agree, and 5=strongly agree). The questionnaire's content validity was confirmed by 3 experts. Internal consistency was excellent, with a Cronbach α of 0.95. We also recorded gender, age, grade point average, and experience with WBVEs (including frequency, proficiency, and comfort).

Statistical Analysis

Data were analyzed using Stata/SE (version 18.0; StataCorp) for Windows. Descriptive statistics summarized participant characteristics. Categorical variables were presented as counts and percentages, and continuous variables were presented as means and SDs. Proportions were calculated using nonmissing denominators. Analyses followed the intention-to-treat principle. Because only 1 (1%) of the 79 participants had missing primary outcome data, analyses were conducted using complete-case analysis. Missing completely at random testing and multiple imputation were not performed due to negligible missingness. Between-group comparisons were conducted using linear mixed modeling. We reported the mean differences and 95% CIs. For the primary outcome (knowledge), we used posttest score as the dependent variable, study group as the factor, and pretest score as a covariate. For multiple tests that were analyzed, the

false discovery rate was controlled at 5% (Benjamini-Hochberg method). Item-level Likert responses were summarized descriptively as means (SDs).

The required sample size was calculated using the postlearning knowledge score from 63 fifth-year medical students who studied the anesthesia technique in the previous academic year (2023), with a mean of 57.94% (SD 14.07). To detect a 10% difference in knowledge score (type I error of 0.05) with 80% power, we determined that a sample size of 32 participants in each group would be required. The dropout rate was accounted for in the sample size, which included at least 40 participants per group, representing approximately 10% of the total dropout rate.

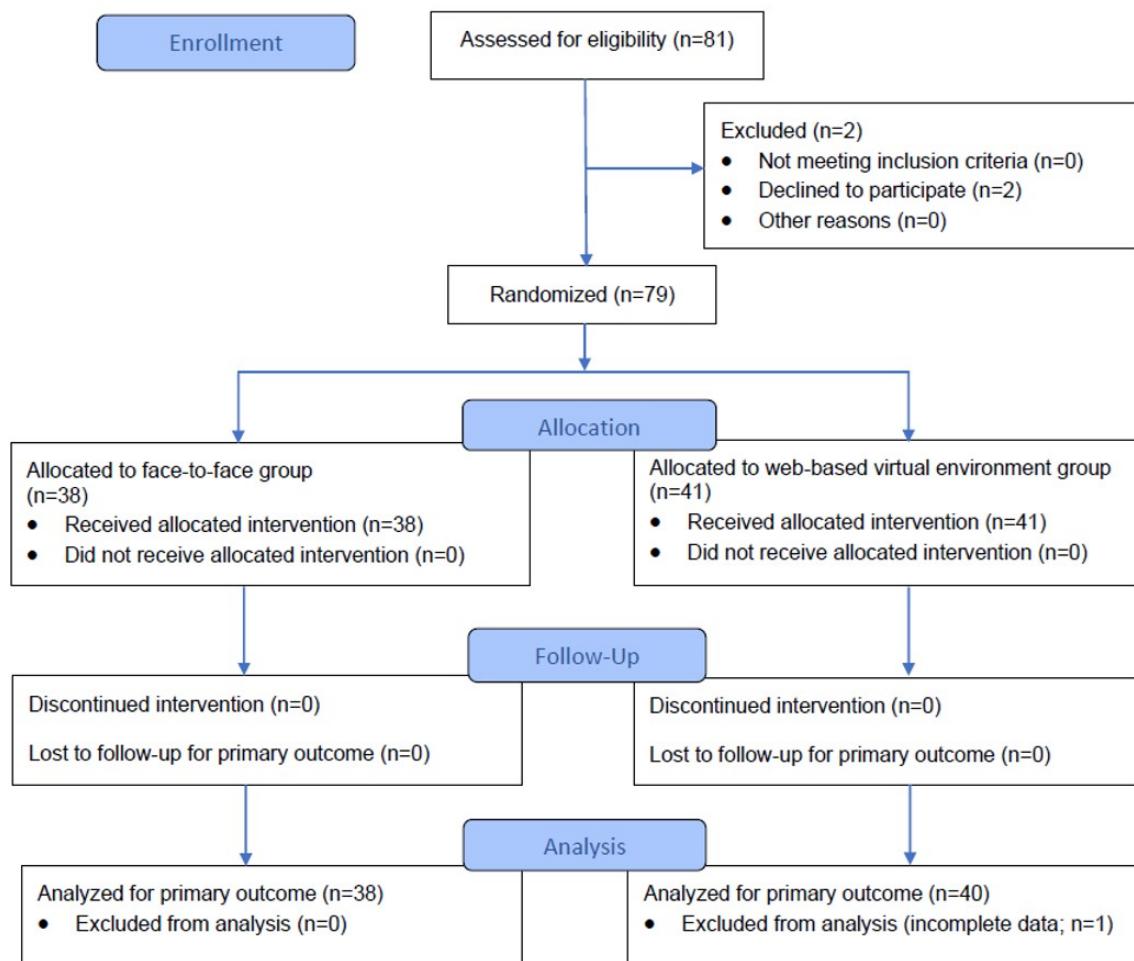
Ethical Considerations

This study received full board approval from the human research ethics committee of Khon Kaen University (HE671294). It was also registered with the Thai Clinical Trials Registry before participant enrollment (TCTR20240708012). This study was conducted in accordance with the World Medical Association Declaration of Helsinki and institutional policies. Participants received clear written information about the study objectives, procedures, potential risks, and data management before participating. Participation was voluntary, and they could withdraw at any time without affecting their academic standing. Those who agreed to take part then provided written informed consent. Study-related data were not collected from them. After participating in this study, students created unique codes instead of using their names, which enabled us to link pre- and posttest data without knowing individual identities to protect privacy and confidentiality. Data were anonymized for analysis and stored on password-protected systems with access restricted to the investigator's team. Students did not receive any financial or in-kind compensation.

Results

Overview

A total of 81 participants were assessed for eligibility; 2 (2%) participants declined to participate. Both were in clusters randomized to the F2F group; therefore, they received F2F teaching as part of the regular course. However, their data were not included in the analysis. Hence, 79 participants were enrolled in the study and allocated into 2 groups: the F2F group (n=38, 48%) and the WBVE group (n=41, 52%); 2% (1/41) of the participants in the WBVE group had incomplete outcome data and were excluded. A total of 78 (99%) participants had complete data in the analysis ([Figure 2](#)).

Figure 2. Flow diagram of the participants in the study.

Participant Characteristics

As shown in [Table 1](#), 78 participants were included, with 38 (49%) assigned to the F2F group and 40 (51%) to the WBVE group. The groups were comparable in gender, age, and grade point average. The proportion of male participants was slightly

higher in the F2F group (25/38, 66%) compared to the WBVE group (21/40, 52%). Most participants in both groups reported limited previous exposure to WBVEs, with more than 60% (51/78) of the participants classified as novices and a few reporting intermediate or higher proficiency. Comfort levels with the technology were predominantly neutral.

Table 1. Baseline characteristics by group in a cluster-randomized trial of a face-to-face (F2F) group versus a web-based virtual environment (WBVE) group for team-based learning among fifth-year medical students (N=78).

	F2F group (n=38)	WBVE group (n=40)
Male, n (%)	25 (66)	21 (52)
Age (y), mean (SD)	22.5 (0.6)	22.5 (0.7)
Grade point average (0-4), mean (SD)	3.50 (0.34)	3.44 (0.33)
Experience with WBVEs, n (%)		
Frequency		
No	19 (50)	26 (65)
Sometimes	18 (47)	14 (35)
Frequently	1 (3)	0 (0)
Proficiency		
Novice	24 (63)	27 (67)
Beginner	10 (26)	7 (17)
Intermediate	4 (11)	5 (13)
Advanced	0 (0)	1 (3)
Expert	0 (0)	0 (0)
Comfort level		
Discomfort	0 (0)	2 (5)
Neutral	27 (71)	27 (67)
Comfort	11 (29)	11 (28)

Knowledge Outcomes

As shown in [Table 2](#), both groups demonstrated significant improvement in knowledge after the intervention. The WBVE group had slightly higher mean pretest scores than the F2F group (mean 6.20, SD 2.04 vs mean 6.03, SD 2.05 for WBVE vs F2F),

but lower posttest scores (mean 10.40, SD 2.62 vs mean 11.24, SD 1.93 for WBVE vs F2F). However, after adjustment using linear mixed modeling and the Benjamini-Hochberg procedure, no statistically significant differences between groups were observed in either pretest or posttest scores ($P=.70$ and $P=.12$, respectively).

Table 2. Knowledge (15-point scale) and learner satisfaction (Likert 1-5) by group in a cluster-randomized trial comparing face-to-face (F2F) versus web-based virtual environment (WBVE) team-based learning among fifth-year medical students (N=78).

	F2F group (n=38), mean (SD)	WBVE group (n=40), mean (SD)	Mean difference (95% CI)	P value
Knowledge (15-point scale)^a				
Pretest knowledge score	6.03 (2.05)	6.20 (2.04)	-0.17 (-1.07 to 0.72)	.70
Posttest knowledge score	11.24 (1.93)	10.40 (2.62)	0.88 (-0.18 to 1.94)	.12
Learner satisfaction (Likert scale 1-5)^b				
Learning topic				
Intellectually challenging and stimulating	4.7 (0.5)	4.4 (0.6)	0.34 (0.05 to 0.62)	.02 ^c
Valuable learning gained	4.9 (0.3)	4.6 (0.6)	0.29 (0.04 to 0.54)	.02 ^c
Increased interest in the subject	4.6 (0.7)	4.5 (0.7)	0.10 (-0.21 to 0.41)	.53
Learning process				
Instructor				
Instructor enthusiasm	4.9 (0.3)	4.7 (0.6)	0.22 (-0.01 to 0.45)	.07
Presentation style holds interest	4.8 (0.6)	4.4 (0.9)	0.36 (-0.01 to 0.74)	.06
Clear explanations	4.8 (0.5)	4.3 (0.8)	0.46 (0.08 to 0.85)	.01 ^c
Facilitates note-taking	4.6 (0.7)	4.4 (1.0)	0.28 (-0.11 to 0.67)	.17
Friendly to students	4.9 (0.2)	4.9 (0.4)	0.10 (-0.06 to 0.25)	.25
Welcomes questions or help	4.9 (0.3)	4.8 (0.5)	0.12 (-0.07 to 0.31)	.25
Genuine interest in students	4.9 (0.3)	4.8 (0.5)	0.07 (-0.10 to 0.24)	.46
Presents alternative viewpoints	4.8 (0.4)	4.6 (0.6)	0.22 (-0.04 to 0.48)	.11
Technique				
Well-prepared materials	4.8 (0.4)	4.3 (1.0)	0.49 (0.08 to 0.90)	.01 ^c
Encourages class discussion	4.9 (0.4)	4.1 (1.0)	0.77 (0.27 to 1.27)	<.001 ^c
Invites the sharing of ideas	4.8 (0.4)	3.9 (1.1)	0.92 (0.34 to 1.50)	<.001 ^c
Encourages and answers questions	4.9 (0.4)	4.6 (0.7)	0.32 (0.04 to 0.60)	.02 ^c
Supports student expression	4.9 (0.3)	4.5 (0.8)	0.42 (0.07 to 0.76)	.009 ^c
Technical satisfaction (audio, video, and media)	4.7 (0.5)	4.4 (0.8)	0.28 (-0.06 to 0.62)	.11
Satisfaction with teaching methods	4.9 (0.3)	4.3 (0.9)	0.54 (0.12 to 0.97)	.004 ^c
Learning outcome				
Content covers objectives	4.9 (0.3)	4.7 (0.5)	0.27 (0.05 to 0.50)	.01 ^c
Consistent with current knowledge	4.9 (0.4)	4.7 (0.5)	0.22 (-0.02 to 0.45)	.07
Overall course satisfaction	4.9 (0.4)	4.5 (0.8)	0.42 (0.07 to 0.77)	.01 ^c

^aKnowledge scores were on a 15-point scale (0-15); higher scores indicated greater knowledge.

^bLikert score (1=very dissatisfied, 2=dissatisfied, 3=neither satisfied nor dissatisfied, 4=satisfied, and 5=very satisfied).

^cStatistically significant difference at $P<.05$.

Learner Satisfaction

As shown in Table 2, learner satisfaction across multiple domains was generally high in both groups. However, several aspects favored the F2F group. For the learning topic, the F2F

group reported higher levels of intellectual challenge and stimulation (mean difference 0.34; 95% CI 0.05-0.62; $P=.02$) as well as greater perceived value in learning gained (mean difference 0.29; 95% CI 0.04-0.54; $P=.02$). For the learning

process, the evaluation was divided into instructor- and technique-related components. The F2F group rated significantly higher on clear explanations (mean difference 0.46; 95% CI 0.08-0.85; $P=.01$) on instructor-related items. It also rated significantly higher on well-prepared materials (mean difference 0.49; 95% CI 0.08-0.90; $P=.01$), encouragement of class discussion (mean difference 0.77; 95% CI 0.27-1.27; $P<.001$), an invitation to share ideas (mean difference 0.92; 95% CI 0.34-1.50; $P<.001$), encouragement and answering questions (mean difference 0.32; 95% CI 0.04-0.60; $P=.02$), support for student expression (mean difference 0.42; 95% CI 0.07-0.76; $P=.009$), and satisfaction with teaching methods (mean difference 0.54; 95% CI 0.12-0.97; $P=.004$). In terms of learning outcomes, the F2F group rated significantly higher on content coverage (mean difference 0.27; 95% CI 0.05-0.50; $P=.01$). While overall course satisfaction was significantly higher in the F2F group mean 4.9, SD 0.4 vs mean 4.5, SD 0.8; mean difference 0.42; 95% CI 0.07-0.77; $P=.01$.

Harms

Neither group experienced adverse events, privacy breaches, or unintended effects. The WBVE session ran as planned, and no technical issues (audio or latency) were observed. There were no withdrawals due to technical problems.

Discussion

Principal Findings

This study compared knowledge and learner satisfaction between F2F and WBVE learning for anesthesia techniques. Both groups were comparable in terms of characteristics, including their experience with WBVEs. All participants showed improvement in knowledge outcomes. However, posttest scores did not show statistically significant differences between groups. Learner satisfaction was generally high in both groups, while it was consistently higher in the F2F group. The F2F group reported significantly higher satisfaction with the learning topic, the learning process (related to learning techniques), and the learning outcome. Overall course satisfaction was higher in the F2F group.

Participant Characteristics

The baseline characteristics of participants in both groups were comparable. There were no significant differences in gender, age, or academic performance. Most participants in both groups reported limited previous experience with WBVEs, with more than 60% (51/78) of the participants classified as novices. Most (54/78, 69%) participants reported neutral comfort ratings. This suggests that students were not very confident with WBVEs but were not uncomfortable with them. This limited exposure may have influenced both their comfort level and engagement with the WBVEs [23,24]. Previous studies have reported that WBVEs are not an obstacle to learning for medical students with limited experience with them [23-25]. The findings indicate that additional training and familiarization sessions with WBVEs may be necessary to enhance learner confidence and optimize engagement [23-25].

Knowledge Outcome

Pre- and posttest results showed improvement in both groups. Posttest scores also did not show statistically significant differences between groups. This suggests that both the WBVE and F2F delivery are effective for teaching. This is consistent with previous studies demonstrating the potential of WBVEs to facilitate cognitive learning and performance in medical students [25-29]. Despite technical and practical challenges, the use of WBVEs can improve knowledge and performance [25-29].

Learner Satisfaction

Our data reported that overall satisfaction was high in both groups. However, across multiple domains of learner satisfaction, the F2F instructional modality was favored. The F2F group reported significantly higher ratings for intellectual stimulation and for the value of what they learned. They also felt that instructors provided clearer explanations and more effectively encouraged idea sharing, questioning, answering, and student expression. Our data showed the benefit of the F2F format in promoting effective engagement, interactive learning dynamics, and real-time interpersonal communication. These findings suggest that conducting TBL in WBVEs is faced with several limitations. Technical constraints and students' limited familiarity with WBVE may affect learner satisfaction [3,23,24]. Additionally, the instructor's facilitation techniques during WBVE should be a concern [23,24].

Cultural context may help explain the differences in satisfaction [29-32]. However, it should be viewed as a possible moderator rather than a fixed or deterministic cause. Thai medical students may defer to instructors and be comfortable with structured guidance and supervision. This may favor F2F settings, where nonverbal cues and immediate instructor feedback are more noticeable. Consistent with previous work in Asia, students have reported greater satisfaction with traditional methods [30]. This appears to be influenced by personality, self-efficacy, and their expectations about instructional design and delivery [30]. In contrast, studies and systematic reviews from Europe and the United States have often found higher satisfaction with WBVEs [29,31]. However, cultural context does not uniformly predict outcomes. Our data and previous studies observed variability across participants that appears to depend on facilitation quality, previous exposure, technology self-efficacy, language demands, and device access [29-32]. During the COVID-19 pandemic, Thai medical students adapted quickly [32]. They reported greater satisfaction with new anesthesiology learning technologies than their teachers. These suggest that growing familiarity can narrow culture-based gaps over time [32]. We view cultural context as one of several interacting factors that affect learner satisfaction.

The variability in learner satisfaction likely reflects implementation issues rather than only the delivery method [29-32]. In WBVEs, instructors cannot rely on being physically in the room, so they must think more carefully about how they facilitate the learning process [10,12,29]. It is essential to use strategies that build social presence and keep students engaged [10,12,29]. Additionally, technical issues, such as latency, audio problems, and complex navigation, can create more burdens

[10,12]. When the sound is delayed, the flow of discussion is broken. It becomes hard to know when to speak or respond [10,12]. Complex navigation may be engaging at first, but students must spend time figuring out how to move between virtual rooms or use the interface. They end up focusing on the platform rather than the TBL tasks and the content itself [2,10,12]. Students' satisfaction is likely to depend on how well learning is facilitated [10,12]. Reliable devices and technical support should be included [10,12]. Therefore, future work should concentrate on strengthening facilitation and improving the user experience.

Implications

Our findings indicate that WBVEs can yield knowledge gains comparable to those of F2F instruction. WBVEs may be especially valuable in settings with limited classroom space, tight schedules, or large cohorts [10,12,16]. However, running TBL in WBVEs requires careful design and facilitation to keep students engaged and improve their satisfaction [2,10,12,29]. Specifically, this includes clearer turn taking, more visible feedback, simpler navigation within the virtual environment, and fewer technical obstacles [2,10,12,29]. This study's main contribution is to separate delivery mode from pedagogy. We directly compared a WBVE with F2F delivery under an identical TBL design using assessor-blinded, prespecified outcomes. These design features extend previous work, which often combined different instructional methods or did not report use and outcome measures as clearly [15-17]. This approach can be implemented with medical students, especially in anesthesiology education.

Limitations

This study has several limitations. A single-center study may limit the generalizability of the findings. The relatively small sample size may limit their applicability to broader and more diverse student populations. Next, a double-masked study was not feasible due to the nature of the intervention. This may have introduced performance or response bias. Additionally, participants had limited previous exposure to WBVEs. This may have dampened their engagement and satisfaction. Cultural factors specific to Thai medical students should be a concern. A preference for structured, instructor-led learning may have contributed to the lower satisfaction reported with WBVE.

Acknowledgments

The authors would like to thank Kaewjai Thepsuthammarat of the clinical epidemiology unit for her expert biostatistical assistance and Suwitcha Sripadungkul for his support and encouragement. Artificial intelligence-based software (Grammarly) was used solely for grammar and language editing. Professional language editors at the Language Institute of Khon Kaen University also reviewed the final manuscript. No artificial intelligence tools were used for data analysis, content generation, or interpretation. The authors remain fully responsible for the accuracy and scientific content.

Funding

This study was funded by the Faculty of Medicine, Khon Kaen University, Thailand (grant IN67086). The funder had no role in the study design, data collection, analysis, interpretation, or manuscript writing.

Future Directions

Future research should include larger sample sizes and multicenter trials. The generalizability of the findings across diverse learner populations should be of concern. In addition, future studies should examine how students' familiarity with WBVEs affects their satisfaction and learning outcomes. They should test whether repeated exposure helps students feel more comfortable and engaged. Moreover, researchers should focus on facilitation techniques to identify the most effective approach for WBVE-based TBL. Simulation-based features within WBVEs can make the environment more interactive and better aligned with students' learning needs. Further work should also study instructional designs tailored to Asian medical education. These designs should incorporate strategies that support collaborative learning and respect cultural norms.

Prior Presentation

Part of this study was previously presented as an oral presentation at the SANCON-ASPA 2025 conference—Scaling New Heights in Pediatric Anesthesia and Beyond (24th Annual Conference of the Society of Anesthesiologists of Nepal [SANCON] and 21st Meeting of the Asian Society of Pediatric Anesthesiologists [ASPA]), held in Kathmandu, Nepal, on April 5, 2025. The abstract was presented under the title "Metaverse versus in-person: impact on anesthesia education and student satisfaction."

Conclusions

This trial is innovative in isolating the delivery mode from pedagogy by directly comparing WBVE to F2F delivery within an identical TBL structure with prespecified end points. This design differs from much of the existing literature, in which shifts in instructional strategy often accompany technological changes. It is challenging to attribute effects solely to the delivery mode. Our findings contribute evidence that a WBVE can deliver comparable short-term knowledge gains in a single standardized session. It has practical implications for scaling instruction when classroom space, faculty time, or scheduling is limited. Nevertheless, lower satisfaction in WBVE underscores the need for real-world implementation. We require stronger facilitation strategies, improved social presence, simplified navigation, and robust technical support to achieve an experience comparable to F2F.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: DS, SB, PB
Data curation: DS, SB, MS, NP, PB
Formal analysis: DS, SB, PB
Investigation: DS, SB, MS, NP, PB
Methodology: DS, SB, MS, NP, WS, PB
Project administration: DS, SB, MS, NP, WS, PB
Supervision: DS, SB, MS, NP, WS, PB
Visualization: DS, SB, MS, NP, WS, PB
Writing—original draft: DS, SB, MS, NP, WS, PB
Writing—review and editing: DS, SB, MS, NP, WS, PB

Conflicts of Interest

None declared.

Multimedia Appendix 1

Instructional package: lesson plan, user guide, and assessment instruments.

[[DOCX File, 7270 KB - mededu_v12i1e80097_app1.docx](#)]

Multimedia Appendix 2

CONSORT-eHEALTH (V 1.6.1).

[[PDF File \(Adobe PDF File\), 409 KB - mededu_v12i1e80097_app2.pdf](#)]

References

1. Solymos O, O'Kelly P, Walshe CM. Pilot study comparing simulation-based and didactic lecture-based critical care teaching for final-year medical students. *BMC Anesthesiol* 2015 Oct 21;15(1):153 [[FREE Full text](#)] [doi: [10.1186/s12871-015-0109-6](https://doi.org/10.1186/s12871-015-0109-6)] [Medline: [26490826](#)]
2. Anas S, Kyrou I, Rand-Weaver M, Karteris E. The effect of online and in-person team-based learning (TBL) on undergraduate endocrinology teaching during COVID-19 pandemic. *BMC Med Educ* 2022 Mar 22;22(1):120 [[FREE Full text](#)] [doi: [10.1186/s12909-022-03173-5](https://doi.org/10.1186/s12909-022-03173-5)] [Medline: [35193577](#)]
3. Liu H, Mi XF, Huang ZZ, Heng BC, Shen WL. Challenges and strategies in developing team-based learning in Chinese medical education. *Med Teach* 2020 Aug 08;42(11):1243-1249. [doi: [10.1080/0142159X.2020.1801995](https://doi.org/10.1080/0142159X.2020.1801995)]
4. Wekerle C, Daumiller M, Janke S, Dickhäuser O, Dresel M, Kollar I. Putting ICAP to the test: how technology-enhanced learning activities are related to cognitive and affective-motivational learning outcomes in higher education. *Sci Rep* 2024 Jul 15;14(1):16295 [[FREE Full text](#)] [doi: [10.1038/s41598-024-66069-y](https://doi.org/10.1038/s41598-024-66069-y)] [Medline: [39009608](#)]
5. Szulewski A, Howes D, van Merriënboer JJ, Sweller J. From theory to practice: the application of cognitive load theory to the practice of medicine. *Acad Med* 2021 Jan 01;96(1):24-30. [doi: [10.1097/ACM.0000000000003524](https://doi.org/10.1097/ACM.0000000000003524)] [Medline: [32496287](#)]
6. Roossien L, van Klaveren LM, de Vos R, Dolmans DH, Boerboom TB. Team-based learning (TBL) in health professions education: a systematic review on its conceptual elements and outcomes. *Med Teach* 2025 May 28;47(12):1933-1947. [doi: [10.1080/0142159X.2025.2506723](https://doi.org/10.1080/0142159X.2025.2506723)]
7. Mulugeta H, Zemedkun A. Implementation of team-based learning for a clinical module of the Ethiopian undergraduate anesthesia curriculum and students' perspectives: a pilot cross-sectional study. *Adv Med Educ Pract* 2023;14:1413-1424 [[FREE Full text](#)] [doi: [10.2147/AMEP.S437710](https://doi.org/10.2147/AMEP.S437710)] [Medline: [38116477](#)]
8. Ho CL. Team-based learning: a medical student's perspective. *Med Teach* 2019 Oct 01;41(9):1087. [doi: [10.1080/0142159X.2018.1555370](https://doi.org/10.1080/0142159X.2018.1555370)] [Medline: [30707851](#)]
9. Wang Y, Li C, Qu L, Cai H, Ge Y. Application and challenges of a metaverse in medicine. *Front Robot AI* 2023 Dec 11;10:1291199 [[FREE Full text](#)] [doi: [10.3389/frobt.2023.1291199](https://doi.org/10.3389/frobt.2023.1291199)] [Medline: [38152305](#)]
10. Lewis KO, Popov V, Fatima SS. From static web to metaverse: reinventing medical education in the post-pandemic era. *Ann Med* 2024 Dec;56(1):2305694 [[FREE Full text](#)] [doi: [10.1080/07853890.2024.2305694](https://doi.org/10.1080/07853890.2024.2305694)] [Medline: [38261592](#)]
11. Seddon I, Rosenberg E, Houston 3rd SK. Future of virtual education and telementoring. *Curr Opin Ophthalmol* 2023 May 01;34(3):255-260. [doi: [10.1097/ICU.0000000000000945](https://doi.org/10.1097/ICU.0000000000000945)] [Medline: [36995108](#)]
12. Ahuja AS, Polascik BW, Doddapaneni D, Byrnes ES, Sridhar J. The digital metaverse: applications in artificial intelligence, medical education, and integrative health. *Integr Med Res* 2023 Mar;12(1):100917 [[FREE Full text](#)] [doi: [10.1016/j.imr.2022.100917](https://doi.org/10.1016/j.imr.2022.100917)] [Medline: [36691642](#)]

13. Bhugaonkar K, Bhugaonkar R, Masne N. The trend of metaverse and augmented and virtual reality extending to the healthcare system. *Cureus* 2022 Oct;14(9):e29071 [FREE Full text] [doi: [10.7759/cureus.29071](https://doi.org/10.7759/cureus.29071)] [Medline: [36258985](https://pubmed.ncbi.nlm.nih.gov/36258985/)]
14. MacNeill H, Masters K, Nemethy K, Correia R. Online learning in health professions education. Part 1: teaching and learning in online environments: AMEE guide no. 161. *Med Teach* 2024 Jan 24;46(1):4-17 [FREE Full text] [doi: [10.1080/0142159X.2023.2197135](https://doi.org/10.1080/0142159X.2023.2197135)] [Medline: [37094079](https://pubmed.ncbi.nlm.nih.gov/37094079/)]
15. Ichikura K, Watanabe K, Moriya R, Chiba H, Inoue A, Arai Y, et al. Online vs. face-to-face interactive communication education using video materials among healthcare college students: a pilot non-randomized controlled study. *BMC Med Educ* 2024 Jul 10;24(1):746. [doi: [10.1186/S12909-024-05742-2](https://doi.org/10.1186/S12909-024-05742-2)]
16. Tokalić R, Poklepović Peričić T, Marušić A. Similar outcomes of web-based and face-to-face training of the GRADE approach for the certainty of evidence: randomized controlled trial. *J Med Internet Res* 2023 Jul 06;25:e43928 [FREE Full text] [doi: [10.2196/43928](https://doi.org/10.2196/43928)] [Medline: [37279050](https://pubmed.ncbi.nlm.nih.gov/37279050/)]
17. Ashokka B, Law LS, Areti A, Burckett-St Laurent D, Zuercher RO, Chin K, et al. Educational outcomes of simulation-based training in regional anaesthesia: a scoping review. *Br J Anaesth* 2025 Mar;134(2):523-534. [doi: [10.1016/j.bja.2024.07.037](https://doi.org/10.1016/j.bja.2024.07.037)] [Medline: [39358185](https://pubmed.ncbi.nlm.nih.gov/39358185/)]
18. Chinnapun D, Narkkul U. Enhancing learning in medical biochemistry by teaching based on VARK learning style for medical students. *Adv Med Educ Pract* 2024 Sep;15:895-902 [FREE Full text] [doi: [10.2147/AMEP.S472532](https://doi.org/10.2147/AMEP.S472532)] [Medline: [39359490](https://pubmed.ncbi.nlm.nih.gov/39359490/)]
19. Spatial Systems. URL: <https://www.spatial.io> [accessed 2024-04-10]
20. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. *J Med Internet Res* 2011 Dec 31;13(4):e126 [FREE Full text] [doi: [10.2196/jmir.1923](https://doi.org/10.2196/jmir.1923)] [Medline: [22209829](https://pubmed.ncbi.nlm.nih.gov/22209829/)]
21. Corbalan M, Plaza I, Hervas E, Aldabas E, Arcega F. Reduction of the students' evaluation of education quality questionnaire. In: Proceedings of the 2013 Federated Conference on Computer Science and Information Systems. 2013 Presented at: FCCSIS '13; September 8-11, 2013; Krakow, Poland p. 659-659 URL: <https://ieeexplore.ieee.org/document/6644083>
22. Fatani TH. Student satisfaction with videoconferencing teaching quality during the COVID-19 pandemic. *BMC Med Educ* 2020 Oct 31;20(1):396 [FREE Full text] [doi: [10.1186/s12909-020-02310-2](https://doi.org/10.1186/s12909-020-02310-2)] [Medline: [33129295](https://pubmed.ncbi.nlm.nih.gov/33129295/)]
23. Elston P, Canale GP, Ail G, Fisher N, Mahendran M. Twelve tips for teaching in virtual reality. *Med Teach* 2024 May 25;46(4):495-499. [doi: [10.1080/0142159X.2023.2285396](https://doi.org/10.1080/0142159X.2023.2285396)] [Medline: [38006603](https://pubmed.ncbi.nlm.nih.gov/38006603/)]
24. Hussain Z, Ng DM, Alnafisee N, Sheikh Z, Ng N, Khan A, et al. Effectiveness of virtual and augmented reality for improving knowledge and skills in medical students: protocol for a systematic review. *BMJ Open* 2021 Aug 16;11(8):e047004 [FREE Full text] [doi: [10.1136/bmjopen-2020-047004](https://doi.org/10.1136/bmjopen-2020-047004)] [Medline: [34400451](https://pubmed.ncbi.nlm.nih.gov/34400451/)]
25. Rodriguez-Florido M, Reyes-Cabrera JJ, Melián A, Hernández-Flores CN, Ruiz-Azola J, Maynar M. Feasibility of teaching and assessing medical students in the metaverse: design and features for its learning efficiency. *J New Approaches Educ Res* 2024 May 06;13(1):9. [doi: [10.1007/S44322-024-00009-6](https://doi.org/10.1007/S44322-024-00009-6)]
26. Plotzky C, Loessl B, Kuhner 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 Jul 27;22(1):222 [FREE Full text] [doi: [10.1186/s12912-023-01384-9](https://doi.org/10.1186/s12912-023-01384-9)] [Medline: [37370124](https://pubmed.ncbi.nlm.nih.gov/37370124/)]
27. Schoeb DS, Schwarz J, Hein S, Schlager D, Pohlmann PF, Frankenschmidt A, et al. Mixed reality for teaching catheter placement to medical students: a randomized single-blinded, prospective trial. *BMC Med Educ* 2020 Dec 16;20(1):510 [FREE Full text] [doi: [10.1186/s12909-020-02450-5](https://doi.org/10.1186/s12909-020-02450-5)] [Medline: [33327963](https://pubmed.ncbi.nlm.nih.gov/33327963/)]
28. Liu S, Xie M, Zhang Z, Wu X, Gao F, Lu L, et al. A 3D hologram with mixed reality techniques to improve understanding of pulmonary lesions caused by COVID-19: randomized controlled trial. *J Med Internet Res* 2021 Oct 10;23(9):e24081 [FREE Full text] [doi: [10.2196/24081](https://doi.org/10.2196/24081)] [Medline: [34061760](https://pubmed.ncbi.nlm.nih.gov/34061760/)]
29. Li Q, Duan H, Zhou X, Sun X, Tao L, Lu X. The use of metaverse in medical education: a systematic review. *Clin Med (Lond)* 2025 May;25(3):100315 [FREE Full text] [doi: [10.1016/j.clinme.2025.100315](https://doi.org/10.1016/j.clinme.2025.100315)] [Medline: [40246026](https://pubmed.ncbi.nlm.nih.gov/40246026/)]
30. Elshami W, Taha MH, Abuzaid M, Saravanan C, Al Kawas S, Abdalla ME. Satisfaction with online learning in the new normal: perspective of students and faculty at medical and health sciences colleges. *Med Educ Online* 2021 Dec 11;26(1):1920090 [FREE Full text] [doi: [10.1080/10872981.2021.1920090](https://doi.org/10.1080/10872981.2021.1920090)] [Medline: [33974523](https://pubmed.ncbi.nlm.nih.gov/33974523/)]
31. Birrenbach T, Zbinden J, Papagiannakis G, Exadaktylos AK, Müller M, Hautz WE, et al. Effectiveness and utility of virtual reality simulation as an educational tool for safe performance of COVID-19 diagnostics: prospective, randomized pilot trial. *JMIR Serious Games* 2021 Oct 08;9(4):e29586 [FREE Full text] [doi: [10.2196/29586](https://doi.org/10.2196/29586)] [Medline: [34623315](https://pubmed.ncbi.nlm.nih.gov/34623315/)]
32. Sripadungkul D, Sripadungkul S, Boonmak S, Boonmak P. Differences in perception of online anesthesiology between Thai medical students and teachers during the COVID-19 pandemic. *Korean J Med Educ* 2023 Mar;35(1):45-53 [FREE Full text] [doi: [10.3946/kjme.2023.248](https://doi.org/10.3946/kjme.2023.248)] [Medline: [36858376](https://pubmed.ncbi.nlm.nih.gov/36858376/)]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth

F2F: face-to-face**MCQ:** multiple-choice question**TBL:** team-based learning**WBVE:** web-based virtual environment

Edited by S Brini, T Leung; submitted 04.Jul.2025; peer-reviewed by T Gladman; comments to author 03.Nov.2025; revised version received 23.Dec.2025; accepted 24.Dec.2025; published 15.Jan.2026.

Please cite as:

Sripadungkul D, Boonmak S, Somjit M, Plailaharn N, Sriraj W, Boonmak P

Web-Based Virtual Environment Versus Face-To-Face Delivery for Team-Based Learning of Anesthesia Techniques Among Undergraduate Medical Students: Randomized Controlled Trial

JMIR Med Educ 2026;12:e80097

URL: <https://mededu.jmir.org/2026/1/e80097>

doi:10.2196/80097

PMID:

©Darunee Sripadungkul, Suhattaya Boonmak, Monsicha Somjit, Narin Plailaharn, Wimonrat Sriraj, Polpun Boonmak. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 15.Jan.2026. 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.

Original Paper

Virtual Reality in Training and Assessment Among Clinical Students and Lecturers at a Nigerian University: A Phenomenological Study

Abiola Olubusola Komolafe¹, PhD; Omotade Adebimpe Ijarotimi², MBChB, MPH; Oluseye Ademola Okunola³, PhD; Olufemi Mayowa Adetutu⁴, PhD; Ayodeji Oludola Oluwatope⁵, PhD; Olatunde Abiona⁶, PhD; Ojo Melvin Agunbiade³, PhD; Adeboye Titus Ayinde⁷, PhD; Stephen Babatunde Aregbesola⁸, BChD; Olawale Babatunde Akinwale⁹, PhD; Babatope Ayodeji Kolawole¹⁰, MBChB; Lanre Idowu¹¹, PhD; Alaba Adeyemi Adediwura¹², PhD; Olayinka Donald Otuyemi¹³, BDS, MSc, MPH

¹Department of Nursing Science, Faculty of Basic Medical Sciences, Obafemi Awolowo University, Ile-Ife, Osun, Nigeria

²Department of Obstetrics, Gynecology and Perinatology, Faculty of Clinical Sciences, Obafemi Awolowo University, Ile-Ife, Osun, Nigeria

³Department of Sociology and Anthropology, Faculty of Social Sciences, Obafemi Awolowo University, Ile-Ife, Osun, Nigeria

⁴Department of Demography and Social Statistics, Faculty of Social Sciences, Obafemi Awolowo University, Ile-Ife, Osun, Nigeria

⁵Department of Computer Science and Engineering, Faculty of Technology, Obafemi Awolowo University, Ile-Ife, Osun, Nigeria

⁶Department of Computer Information Services, College of Arts and Sciences, Indiana University Northwest, Gary, IN, United States

⁷Department of Psychology, Faculty of Social Sciences, Obafemi Awolowo University, Ile-Ife, Osun, Nigeria

⁸Department of Oral and Maxillofacial Surgery, Faculty of Dentistry, Obafemi Awolowo University, Ile-Ife, Osun, Nigeria

⁹Department of Electronic and Electrical Engineering, Faculty of Technology, Obafemi Awolowo University, Ile-Ife, Osun, Nigeria

¹⁰Department of Medicine, Faculty of Clinical Sciences, Obafemi Awolowo University, Ile-Ife, Osun, Nigeria

¹¹Department of Educational Technology and Library Studies, Faculty of Education, Obafemi Awolowo University, Ile-Ife, Osun, Nigeria

¹²Department of Educational Foundation and Counselling, Faculty of Education, Obafemi Awolowo University, Ile-Ife, Osun, Nigeria

¹³Department of Child Dental Health, Faculty of Dentistry, Obafemi Awolowo University, College of Health Sciences, Ile-Ife, Osun, Nigeria

Corresponding Author:

Olayinka Donald Otuyemi, BDS, MSc, MPH

Department of Child Dental Health

Faculty of Dentistry

Obafemi Awolowo University, College of Health Sciences

Road 2, OAU Campus

Ile-Ife, Osun, 220001

Nigeria

Phone: 234 08037191734

Email: otuyemi@yahoo.com

Abstract

Background: Virtual reality (VR) technology is increasingly used in health care professionals' education as a novel tool for teaching, learning, and assessment.

Objective: This study explored the experiences of clinical students and lecturers with VR for training and assessment at a Nigerian institution. It also explored students' perceptions of the usefulness of VR in improving their clinical abilities, knowledge retention, engagement, and overall learning experience.

Methods: A qualitative research study was conducted among 24 clinical students and 8 clinical lecturers. A developed Virtual reality model to TRain and Assess Clinical Students (VTRACS) was used to train and assess clinical students using clinical scenarios. Data were collected through 4 focus group discussions conducted among the clinical students and 8 in-depth interviews conducted among the clinical lecturers. Trustworthiness was maintained, and ethical approval for the study was obtained. The focus group discussions and in-depth interviews were audio-recorded, transcribed verbatim, and analyzed using NVivo (version 11; QSR International).

Results: Many of the participants had no previous experience with VR in teaching and learning activities, but judging from their engagement with VTRACS, they defined VR as an alternative learning method (alternative to the traditional physical method). Major themes emerging from the study were expression of excitement, simple and useful innovation, proficiency enhancement, challenges with innovation, and uniformity. The clinical students adjudged VTRACS as an educational supplement with a feeling of unlimited learning access, enhancing clinical abilities while positively impacting their confidence and reducing clinical errors. The participants also described the objectivity and standardization of clinical scenarios as drivers of uniformity in training and assessment of clinical students. The participants were, however, concerned about the loss of empathy with the use of VTRACS, which may negatively impact the affective domain of learning.

Conclusions: The use of VR in the teaching and assessment of clinical students at a Nigerian university is perceived as a complementary method of learning that increases skill acquisition, provides unlimited access to training, and enhances proficiency. While VR is considered to be engaging and beneficial to health care professionals' education, there is a need for its effective incorporation into clinical courses and mitigation of challenges such as cost and technology to ensure the realization of the full potential of VR in health care professionals' education.

(*JMIR Med Educ* 2026;12:e75021) doi:10.2196/75021

KEYWORDS

virtual reality; virtual patient; clinical students; health care professionals' education; digital technology; Nigeria; VR in education; educational technology

Introduction

The challenges posed by the restriction of physical contact during and in the aftermath of the COVID-19 outbreak seem to have suddenly highlighted the need to develop alternative methods of handling teaching and learning activities that require physical contact with a view to attaining desired results [1-3]. In Nigeria, medical training, whose pedagogical design revolves around physical clinical assessment, is no exception.

The increasing adoption of technology, including virtual reality (VR), as an emerging tool in health care professionals' educational training processes, backed by empirical research, reinforces the credibility of this perspective [4]. VR simulations are engaging, interactive, and realistic, and can improve the learning experience and enable new approaches to the training and assessment of clinical trainees. This technology enables students to participate in complicated clinical scenarios in a controlled and safe setting, allowing for the development of important skills without the hazards associated with real-world patient encounters.

VR has enormous potential in medical education, particularly in resource-limited countries, such as Nigeria. Traditional clinical training approaches frequently encounter obstacles, such as restricted access to clinical facilities, insufficient patient interactions, and inconsistent teaching resources [5,6]. Digital technologies have had a tremendous impact on education, resulting in the implementation of integrated information and communication technology (ICT) initiatives. However, concerns about teaching and learning quality persist, particularly during the COVID-19 pandemic and in other situations that limit physical contact. Many educational institutions have limited experience and digital capacity, which results in learning gaps and disparities in the quality of education and training received by clinical students. This has prompted educational institutions to improve their digital capacity and plan for a successful digital transformation. A nonsystematic literature analysis demonstrated that ICT integration has an impact not only on student achievement but also on school-related factors and stakeholders.

Factors associated with digital transformation play an important role in effective and efficient changes in educational institutions [7]. VR presents an opportunity to address these issues by providing standardized, high-fidelity simulations that are accessible to all students regardless of their geographical location.

The adoption of VR in Nigerian health care professionals' education is still in its nascent stage, and there is a need for empirical research to understand its impact and effectiveness. The use of a developed VR model in clinical students' instruction and assessment has the potential to alter health care professionals' education in Nigeria [8]. This study adds to the growing body of literature on the use of VR in education by conducting a contextualized analysis of its impact at a Nigerian university. The findings of this study can help guide future VR implementations in health care professionals' education and ensure that this technology meets the unique challenges faced by Nigerian clinical students. Understanding the experiences may provide actionable information to the educators, policymakers, and technology developers on future VR implementation in health care professionals' education to ensure the technology addresses the uniqueness of health care professionals' education, including the potential benefits and limitations of VR in enhancing health care professionals' education in Nigeria. This study aimed to explore the experiences of clinical students following their engagement with a developed VR technology software, the Virtual reality model to TRain and Assess Clinical Students (VTRACS), as part of their training and assessment at a Nigerian university. The study hypothesized that the use of the VR model (VTRACS) in training and assessment of clinical students would promote objectivity in clinical assessment, improve students' clinical abilities, and overall learning experience.

Methods

Setting

This study adopted a phenomenological design to explore the experiences of students and faculty members using the VTRACS software in the College of Health Sciences. Phenomenology is especially appropriate for this study because it allowed the researchers to uncover the essence of how participants perceived, interpreted, and made meaning of VR in their educational environment. Phenomenology was applied in the purposeful sampling of participants who had direct experience with VR and in exploring their personal experiences and interpretations of the technology in the results and discussion, rather than focusing on how VR was used. The college comprises the Faculty of Basic Medical Sciences with 2 programs, namely Nursing Science and Medical Rehabilitation, the Faculty of Dentistry with 1 program, and the Faculty of Clinical Science with 1 program, although it has 2 departments: Medicine and Surgery.

The study was conducted using a haptic VR simulation platform, VTRACS, with limited haptic feedback, delivering immersive clinical scenarios through standalone VR headsets (Meta Quest 2). The simulations were designed to replicate common patient care scenarios, such as performing vital signs checks, abdominal examinations (for pregnancy and appendicitis), and conducting tooth extraction, focusing on skill acquisition in nursing, medical, and dental education. The software used a learner-centered approach for the training and assessment of clinical students across the 3 faculties within the College of Health Sciences of the university [8]. Participants interacted with the virtual environment using hand-held controllers, allowing for navigation, action selection, and patient engagement. While the system provided a high level of visual and auditory immersion, haptic feedback was limited to vibrations; no tactile sensations or physical resistance were simulated. Each VR session lasted approximately 20-25 minutes and was conducted in a controlled laboratory environment with technical support available throughout. The content was aligned with curricular objectives and developed to reflect realistic clinical workflows and patient interactions.

Recruitment Strategy

A total of 32 participants (8 clinical lecturers and 24 clinical students) were recruited for the study. A purposive sampling technique was used to recruit 6 clinical students and 2 clinical lecturers each from Nursing Science, Dentistry, Medicine, and Surgery, who had previously been involved in training and assessment with the VTRACS model. The choice of 6 students from each department was based on the rationale that, unlike quantitative studies, qualitative research focuses on rich, detailed insights from a smaller, carefully selected sample. Research shows that saturation often occurs within 6-12 interviews per homogeneous group, making 6 students per department a reasonable and efficient choice. It also makes data analysis more manageable, especially when using thematic analysis, as was

done in this study. Equal representation ensures that comparisons across departments are valid and not skewed by unequal sample sizes. This is crucial when exploring discipline-specific experiences with VR tools in clinical training. Equal numbers help maintain a balanced view of the phenomenon across different clinical disciplines.

The inclusion criteria were being a clinical student and having been trained and assessed in clinical examination with VTRACS, while clinical students who had not been trained or assessed with VTRACS were excluded from the study. The provost of the college and head of departments identified and nominated clinical lecturers with a broad range of disciplinary backgrounds and wide experience who would give diverse insights for the in-depth interviews (IDIs). The students and lecturers had 2 supervised exposures to the VTRACS model before the interviews were conducted.

Data Collection

VTRACS was used to teach and assess 24 clinical students on clinical scenarios in their field of study. Thereafter, data were collected from the students and the clinical lecturers using an interview guide with 4 open-ended questions (Table 1). Four focus group discussions (FGDs), with an average duration of 45 minutes, involving 6 students from each of the 4 departments, were conducted to explore their experiences with the VTRACS model. The basis for selecting 6 students per department was to ensure equality in the representation of the students. In addition, 8 IDIs lasting an average of 30 minutes were conducted with clinical lecturers from the 4 departments to explore their experiences with the VTRACS model. Qualitative rigor of the study was maintained. To ensure credibility, the research team built rapport with the participants and engaged them in the use of VR before the interviews were conducted. Credibility was also enhanced through member checking, where interview transcripts were given to a few participants who were willing to review their transcripts to confirm accuracy and resonance. Participants' feedback from the process of member-checking confirmed accuracy and true representation of the interviews. Various strategies contributed to enhancing confirmability in this study. First, data triangulation was ensured by collecting information from multiple sources, such as clinical students in all the faculties in the college and the clinical lecturers. Second, multiple methods, such as FGDs and IDIs, were used to ensure consistency across different datasets. Third, probing followed the flow of the interviews rather than leading, and clarifications were sought where and when necessary. For dependability, documentation of the research process was done transparently through verbatim transcription of all the recorded interviews. All FGDs and IDIs were audio-recorded and transcribed verbatim. An audit trail was maintained, and the researchers kept audio recordings, analysis notes, and coding details. The description of the setting, participants, and context of exploring participants' experiences with the use of VTRACS allows for the transferability of the study. Qualitative data analysis software, NVivo (version 11; QSR International), was used in organizing data into nodes and themes.

Table 1. Sociodemographic characteristics of participants (N=32).

Variable	Faculties (n=8), n (%)	Students (n=24), n (%)
Age (years)		
21-25	— ^a	20 (83.3)
26-30	—	4 (16.7)
31-35	—	—
36-40	—	—
41-45	3 (37.5)	—
36-40	2 (25)	—
≥46	3 (37.5)	—
Sex		
Male	4 (50)	12 (50)
Female	4 (50)	12 (50)
Faculty or department		
Dentistry	2 (25)	6 (25)
Nursing	2	6
Medicine	2	6
Surgery	2	6
Level of study		
Lecturing	8 (100)	—
400	—	12 (50)
500	—	10 (41.7)
600	—	2 (8.3)

^aNot applicable.

Ethical Considerations

Ethical approval was obtained from the Health Research Ethics Committee of the Obafemi Awolowo University Teaching Hospitals Complex, Ile-Ife, Nigeria, with protocol number ERC/2022/11/21. The VTRACS project team contacted the potential participants and explained to them the objectives of the study and the purpose of the interviews. Written informed consent was obtained from all the participants, and they were assured of their confidentiality and their right to withdraw from the study at any time without penalty. None of the participants who gave consent to participate in the FGD and IDI withdrew their consent.

Results

Characteristics of Participants

A total of 32 clinical students and faculty members participated across the 3 faculties in the College of Health Sciences, as shown in Table 1. Twenty-four clinical students, 6 from each of the 4 departments or faculties, took part in 4 homogeneous FGDs, and 8 clinical lecturers, 2 from each department or faculty, participated in IDIs.

Themes and Nodes From Participants' Experience

The results of the study yielded 5 major themes and 13 nodes. The experiences of the students and the lecturers were largely similar, except that lecturers reported more concerns compared to the students, possibly because of generational differences (Gen X and Gen Z) in exposure to and interaction with technology. [Textbox 1](#) provides the themes and their corresponding nodes.

Textbox 1. Themes and nodes from participants' experience with the Virtual reality model to TRain and Assess Clinical Students (VTRACS).

Expression of excitement

- Novice
- Complementary learning method
- Imagination rather than reality

Simple and useful innovation

- Ease of use
- Compatible with the task

Proficiency enhancement

- Unlimited learning access
- Confidence
- Fewer medical errors

Challenges with innovation

- Limited patient complexity and feedback
- Loss of empathy
- Financial burden of innovation

Uniformity

- Standardized clinical scenarios
- Objectivity in clinical assessment

Theme 1: Expression of Excitement

The introduction of VTRACS for learning by clinical students prepared them for real-world situations, allowing learning and continuous assessment in furtherance of academic excellence and technical expertise. This theme, “expression of excitement,” emerged from the clinical students’ description of their experience with VTRACS as being novice, as an alternative learning method, and as imagination rather than reality. Many of the students and clinical lecturers did not have previous experience with VR for teaching and learning activities, but based their understanding on their interaction with VTRACS. The participants said:

I haven't used it (VR) before... [VTRACS DENTISTRY FGD 1]

Prior to this, I've not had...I've only heard about it, I've not taken part in it. [VTRACS IDI MEDICINE 2]

I have not heard about it before...I have not had any experience about it before [VTRACS IDI NURSING 2]

I've had contact...using VR, but not really in terms of education...education, because there are so many games that require the use of VR and all that... [VTRACS IDI DENTISTRY 2]

The definition of VR varies due to the differences in the experiences of the participants with VR. However, a common consensus in the definition of VR is its recognition as an alternative learning method. The participants said:

Virtual reality is more or less an alternative to physical practical where you still have the real-life experience scenario at least to a large extent, and you are able to carry out procedures to an extent like the way you would physically. [VTRACS NURSING FGD 3]

To the best of my knowledge, by virtual reality, it's just like a form of alternative to the usual traditional way of learning...well, I do not really have the details of how it works. I know it has to do with technological modifications [VTRACS IDI MEDICINE 2]

The responses from participants in the FGDs and IDIs describe VTRACS as a virtual tool that enhances learning for users who may not have the opportunity for physical training. While some participants viewed it as an alternative, the overall perspective more accurately reflects VTRACS as a complementary method to the traditional physical methods adopted in teaching, particularly in contexts where hands-on experience with real-life patients is essential and available.

The participants demonstrated unfamiliarity with the usage of VTRACS in their statements, but explained their perception of the versatility of VTRACS in teaching and learning from their first experience with it. The participants described the performance of procedures with VTRACS as imagination rather than reality. The imaginative nature of VTRACS immerses users into a different reality where they can explore different and novel learning components. For instance, the participants said:

Virtual reality is like imagination...someone's imagination like...defined from what you think the reality will be as the name implies – virtual reality, like it's an imagination. [VTRACS NURSING FGD 4]

It's a simulation...of the clinical procedure which can be used to teach students or...especially online without having an actual patient there. So, I think it's a very...it's a very good initiative. [VTRACS DENTISTRY FGD 2]

The participant acknowledged VTRACS as a reliable learning platform in the online environment where face-to-face engagements are not possible. This component of VTRACS coincides with the dire need for a substitute method of learning as a result of the limitations imposed by the pandemic. The statement, "It is a very good initiative," describes the positive perception of VR as an alternative teaching method among clinical students. The new pedagogical method seen in VTRACS and the use of imagination in the performance of clinical procedures, which stems from hands-on experience with VTRACS, gave participants a feeling of excitement. Although participants explained that it did not feel the same as physical practice, this is a limitation in replicating reality, as the use and adoption of VTRACS may not replace traditional physical training in clinical practice, but serves as an alternative.

It was really exciting, comfortable, and actually interesting. [VTRACS NURSING, FGD 1]

I also found the experience interesting and...comfortable [VTRACS SURGERY, FGD 6]

....it was fun, although it wasn't like the physical one. [VTRACS MEDICINE, FGD 4]

I had a...I had a nice experience, and it was interesting, very interesting... [VTRACS IDI MEDICINE 1]

I mean, the experience is something you want to look toward [VTRACS IDI DENTISTRY 1]

Although the participants were excited, there was diversity in excitement. While the majority of both students and lecturers found the experience with VTRACS pleasant, comfortable, and interesting, 2 lecturers expressed unpleasantness in the form of headache and dizziness. One of them said:

It was not that pleasant to me when I placed the... accessories, ...the one that was put on the head...(headset) okay, so, when I...when I wore it, I discovered that I was having a kind of headache during the process [VTRACS IDI NURSING 2]

This experience of discomfort reinforces the existence of individual differences and health disparities in the use of technologies.

Theme 2: Simple and Useful Innovation

As revealed by the participants' testimonies, they experienced VR in various ways. Participants who described their experiences as simple also expressed various uses of VR for games, education, watching films, and video demonstrations. This acquaintance with VR made the learning of VTRACS

easier because they had previous knowledge of the functioning of VR technological innovation.

It (VTRACS) picked my interest,...and when there is interest, learning becomes much easier. It was quite easy and very logical and systematic in a way. ...it's very clear, and it's well arranged and very logical. [VTRACS NURSING, FGD 5]

It was easy for me to navigate.... the operations and then the...the keys to use, and the instructions were very easy to apply and very easy to follow....it was so easy to use for me. [VTRACS SURGERY, FGD 3]

It gives room for multiple practice...it's easier...students are able to practice more. [VTRACS DENTISTRY, FGD 6]

However, some of the participants who had no previous experience with VR found VTRACS interesting but complex at first interaction. For example, one participant said:

it was fun,although it was difficult to learn at first, but with patience to adapt to a new system. It's more difficult than physical one. Yeah, maybe because I've not tried it so many times,if it's physical, I know where to place...I know how to reach... [VTRACS MEDICINE, FGD 4]

Some lecturers expressed difficulty with hand-eye coordination compared to the students. They, however, opined that learning new technology at their age is good for their cognition. In the words of one of them:

They (procedural steps in VTRACS) were logical. The only thing was coordinating between the brain and the eye movement...that took some time, but like I said, practice makes perfect... one will be stimulating his or her brain in doing such things. If there is something you've not been doing before, and you start doing it will help each participant whether old or young, particularly the older ones like me. [VTRACS IDI SURGERY 1]

Theme 3: Proficiency Enhancement

The theme "proficiency enhancement" was generated from 3 nodes, namely: unlimited learning access, confidence, and fewer medical errors. The nodes evolved from the participants' responses to the benefits of VTRACS in clinical students' training. The participants believed it was an avenue for students to learn better and gain more knowledge, even beyond the boundaries of the classroom.

Possible benefits could be that in terms of time flexibility...the students don't always have to be in the hospital settings...they can learn outside hospital settings...it's flexible [VTRACS IDI MEDICINE 2]

This flexibility, in the opinion of the lecturers, would also give them more time to do other things because they do not have to be physically present to supervise the students at all practice sessions.

The participants' responses reflected an opportunity for seamless learning, transcending physical barriers and limitations of availability. For instance, one of the participants said:

It (VTRACS) can be used to teach a large class, when there is a need for physical distancing, when we have to conduct exams, and we don't have the patients on ground, we can use the virtual experience to conduct the clinical exams... [VTRACS IDI SURGERY 2]

Also, the responses from the participants emphasized the benefits of this technology to society. VTRACS, as a virtual platform for learning, equips learners better and helps them replicate these procedures in real-life situations. This further boosts the confidence of clinical professionals through guidance in performing real-life procedures and reduces cases of medical errors. The urge to be in the field of medicine and clinical sciences may be further encouraged through the adoption and use of this technology, thereby producing more medical professionals to cater to the health needs of an ever-evolving society.

...if students have known so many procedures using VR, and they have understood very well, they've learnt it, ...it will reduce some errors that are committed in the hospitals [VTRACS IDI NURSING 2]

I think if we implement this virtual reality now, it will help both the students and the staff at large...the society at large, ..., to be able to ...train more students...more doctors will be available [VTRACS SURGERY, FGD 3]

Theme 4: Challenges With Innovation

This theme captures the experiential and financial concerns expressed by participants regarding the implementation and use of VTRACS. As exciting as most of the participants found VTRACS, some participants described VTRACS as unable to replicate the variability and realism of real patient interactions, limiting opportunities for comprehensive clinical feedback. They also expressed concern that the lack of real human engagement reduced their ability to connect emotionally, potentially diminishing the development of empathy. One of the participants said:

...a shortcoming of this virtual reality program is that no two patients are the same, so there is a way you will treat a live patient and that you get a particular type of feedback that will help you become a better doctor, that you won't get from virtual reality. [VTRACS DENTISTRY, FGD 3]

I think the disadvantage is that you will probably be unable to show empathy [VTRACS DENTISTRY FGD 2]

The challenges I think will be that...there could be a possibility of losing touch with the empathy that comes with.... I mean, the affective aspect [VTRACS IDI MEDICINE 2]

The participants noted that VTRACS may limit the development of empathy due to the absence of real human interaction. In

addition, the lack of comprehensive, responsive feedback in VTRACS may be a potential barrier to professional growth. However, some participants indicated that VTRACS could reduce the anxiety often associated with assessments in traditional physical settings, thereby enhancing performance and proficiency. As one participant explained:

Most times, students make mistakes in the physical method of assessment, not because they don't know it, but because of anxiety. With this VR, the fact that you are wearing it (headset), and you can't see the person assessing you, you are comfortable to put what you already know into practice. [VTRACS NURSING, FGD 3]

Some of the participants, students and lecturers alike, expressed concerns about the financial implications of acquiring the new technology, which may negatively affect sustainability. One student said:

...that means every student will have it, and obviously, the financial burden will be there. [VTRACS NURSING, FGD 6]

One of the lecturers also said:

It might be expensive. I don't know the cost of this one, but I know with the way things are, it will be. [VTRACS IDI SURGERY 1]

Another lecturer said:

I don't know how much one costs (headset)...in terms of cost containment, I don't know if it's that cost-effective in the sense that how many of them will be needed for the number of students that we have. [VTRACS IDI NURSING 2]

These concerns were, however, not sufficient to deter them from embracing VTRACS because they perceived the concerns as challenges that could be overcome.

Theme 5: Uniformity

The theme "uniformity" emerged as a result of participants' acknowledgment that VTRACS ensures standardized clinical scenarios for all clinical students in teaching and examination, thereby promoting objectivity in clinical assessment. The participants said:

...it (VTRACS) is also standardized.... what the patient is presenting with will be standardized for everybody, not that one patient will have an extremely difficult case, and one person will have an easy case, and they will still be score based on that. [VTRACS DENTISTRY, FGD 3]

Discussion

Principal Findings

The use of VR in the training and assessment of clinical students presents transformative potential for health care professionals' education, particularly within the context of a Nigerian university. This qualitative study explored the experiences of clinical students engaged in VR technology (VTRACS) as part of their educational curriculum, highlighting their perceptions

of the innovation, usefulness, and the benefits and challenges associated with this approach. Overall, participants expressed excitement about using VTRACS and recognized it as a valuable educational supplement, particularly in contexts where access to traditional hands-on training is constrained. The objectivity in assessment and standardized clinical scenarios was described as a driver of uniformity in clinical training and assessment by the participants. The flexibility and accessibility of VR-based simulation were seen as key strengths, allowing learners to engage with clinical scenarios in a safe, controlled environment, thus enhancing clinical abilities. However, participants also expressed concern about the limited patient complexity and feedback, and the potential loss of empathy with the use of VTRACS.

Comparison With Previous Work

The findings of this study reveal that VR technology significantly enhances the learning experience of clinical students. The immersive nature of VR allows students to engage in near-realistic clinical scenarios, providing a safe and controlled environment to practice and hone their skills. Richards [9] posited that mixed-reality technologies in health care professionals' education can boost student engagement by combining virtual and augmented reality imagery. Bridge et al [10] also demonstrated that a VR environment improved mindfulness meditation and allowed pseudo-anonymous interactions with peers and tutors. The students in this study reported an increased level of engagement and motivation, noting that VR simulations are more interactive and enjoyable than traditional learning methods. This aligns with the findings of Lau et al [11], who conducted a study involving nurses as participants and reported that the majority similarly found the immersive VR experience engaging and experienced a strong sense of presence. Both studies underscore the potential of VR to enhance learner motivation through immersive, interactive environments, thereby supporting improved learning outcomes by providing a dynamic and stimulating educational experience.

One of the primary benefits identified by the participants in this study was improvement in clinical skill acquisition and confidence. VR simulations offer repetitive practice opportunities in a safe, controlled environment, without the risk of harm to patients. This repetitive practice is crucial for mastering clinical procedures and building confidence. Supporting this, Kennedy et al [12], who compared traditional and interactive VR instructions for clinical students, found that those who received VR training made 40% fewer errors in a simulated practical setting compared to those who received traditional training, highlighting VR's effectiveness in reducing human error and enhancing procedural knowledge. Similarly, Lau et al [11] reported that VR can improve nurses' knowledge and skills, though they emphasized the need for further refinement of VR prototypes to enhance the user experience. In this study, students also noted that the immediate feedback provided by the VR system enabled them to promptly identify and correct mistakes, contributing to more effective and self-directed learning. This finding aligns with a systematic review of 31 studies, which demonstrated that VR-based education significantly improves knowledge and skill outcomes among health care professionals when compared to traditional

learning methods, reinforcing VR's positive impact on clinical competence and learner confidence [13].

The integration of VR into health care professionals' education has the potential to address accessibility and equity issues. Lakshminarayanan et al [14] affirmed that affordability and integrated experience creators, when combined with augmented reality (AR) or VR, can facilitate low-cost data collection in remote and rural locations. Evidence has shown that the use of VR in health care, particularly in surgical education, is gaining popularity owing to possible advantages such as enhanced preoperative resident exposure and improved patient safety [15]; however, VR may not completely imitate organic tissues or generate realistic scenarios, and persistent challenges, such as inconsistent internet connections and implementation costs, may impede equitable adoption among academic centers. Ibrahim et al [16] investigated the incorporation of equity, diversity, and inclusion (EDI) into clinical simulation instruction for health care trainees and clinicians. They found that the integration of EDI in clinical simulation education improves self-awareness, communication, insight, knowledge, self-efficacy, and competence. The recommendations include a systematic approach to incorporating EDI, creating a digital repository of EDI-focused scenarios, cocreating simulations with individuals from varied backgrounds, ensuring a secure learning environment, and conducting more rigorous research to enhance the science of clinical simulations.

Participants in this study generally perceived VTRACS as a valuable tool for enhancing learning, particularly in contexts where access to traditional hands-on or physical training is limited. In resource-limited settings such as Nigeria, access to high-quality clinical training facilities and real patient interactions can be limited [17]. VR provides a means of bridging this access gap by offering standardized and high-fidelity simulations accessible to all students, regardless of their geographical location or institutional resources. This democratization of training resources can contribute to leveling the playing field and ensuring that all students receive comprehensive and equitable training. Andigema et al [18] emphasized the importance of action and collaboration in implementing and scaling up artificial intelligence advances to improve health outcomes and ensure that African populations receive the health care they deserve. As immersive technologies such as VR continue to evolve, their strategic integration into health care professionals' education has the potential not only to transform learning experiences but also to drive equitable health care delivery on a global scale.

The ability to engage in safe, repeatable, and immersive clinical simulations was highlighted as a key benefit of VTRACS. While some participants initially described VTRACS as an alternative to conventional training, their overall reflections highlighted important challenges with the innovation, such as limited realism and feedback and a reduced capacity to foster empathy, which are typically addressed more effectively in hands-on, physical settings. These challenges underscore the importance of positioning VTRACS not as an alternative but as a complementary training method. This perspective aligns with evidence from previous studies, which support the use of VR in simulation-based learning primarily as a supplement to

traditional physical methods, unless realistic haptic feedback is integrated to replicate tactile experiences [19-21]. For example, a systematic review by Rangarajan et al [19] found that haptic-enabled VR simulations significantly enhanced surgical performance compared to nonhaptic VR, highlighting the critical role of tactile realism in effectively substituting physical training. Although the VR platform used in this study includes limited haptic feedback through controller vibrations, it lacks tactile sensations or physical resistance, which constrains its ability to fully replicate real-world interactions or convey empathy. Despite its strong visual and auditory immersion, this limitation reduces the system's overall realism.

In addition to pedagogical concerns, this study also identified logistical and technical barriers that could impact the effective implementation of VR tools such as VTRACS. Participants reported challenges, including unfamiliarity with the VR learning format, difficulty operating hand controllers, and occasional physical discomfort. These findings are consistent with those reported by Lau et al [11], who observed similar usability issues in their study. These challenges underscore the need to address both ergonomic design and user onboarding when integrating VR-based training into educational settings.

While virtual and augmented reality technology is gaining traction in higher education due to its immersive learning potential, the high cost of acquiring and maintaining VR hardware and software remains a major constraint, particularly for institutions operating under limited budgets. For instance, Marks and Thomas [22] described a purpose-built VR laboratory that housed 26 Oculus Rift headset units. Over 5 teaching sessions, 4833 students were taught in the laboratory, with notable increases in student engagement, especially in the Faculty of Engineering, and a 250% increase (n=1016) in student enrollment. Although 71.5% (n=211) of the students reported improved learning outcomes, the initiative was supported by subsidized funding, highlighting potential cost barriers to wider adoption. Conversely, the shift toward remote education during the COVID-19 pandemic prompted the exploration of VR as a more scalable and cost-effective alternative. A retrospective case study comparing high-fidelity manikin-based simulation and VR-based learning found that VR was 22% less time-consuming and 40% more cost-efficient, supporting its feasibility as a pedagogical solution in resource-constrained environments [23].

Furthermore, technical issues, such as software glitches and hardware malfunctions, were noted by some students, which can disrupt the learning process and cause frustration. A learning curve was also noted, with some users requiring additional training to engage confidently with the platform. These findings echo previous studies that stress the importance of robust institutional support, reliable technology infrastructure, and user training when integrating VR into educational settings. Collectively, these insights reinforce the idea that while VR can meaningfully augment health care professionals' education, its role should be carefully framed to complement rather than replace traditional hands-on methods.

Cultural and contextual factors also play crucial roles in the implementation of VR in health care professionals' education.

In the Nigerian context, factors such as technological infrastructure, institutional readiness, and cultural attitudes toward technology adoption can influence the success of VR integration. The COVID-19 pandemic has highlighted the crucial role of educational technology in the global education system. However, probably because of a lack of resources, Nigeria has faced challenges in sustaining educational services, resulting in a shift in the school calendar, longer graduation dates, and lower research output. To maintain sustainability, the Nigerian education system must adopt a pragmatic approach to innovation by using the diffusion of innovation theory. However, many practitioners lack the requisite skills and attitudes, require in-service training, and face challenges in having a methodical approach to purchasing and installing ICT infrastructure [24].

Following the use of the headset, some participants raised the issue of headache and dizziness. VR sickness, also known as visually induced motion sickness, can be described as a physiological, unpleasant experience resulting from a mismatch between users' vestibular and visual stimuli, which can negatively impact the VR user experience [25]. However, certain factors, such as fear, previous experience of motion sickness, and previous exposure to VR, can also influence participants' experience of VR sickness [26]. Evidence shows that the number of sessions and the intervals between sessions are important factors in increasing adaptation effects and minimizing motion sickness [27-29]. It is usually advisable to start with short sessions, adjust headset settings for comfort, take breaks, avoid intense content, and consult a doctor if the VR sickness persists.

It is essential to tailor the VR content to reflect specific clinical scenarios and cultural nuances relevant to the Nigerian health care system to ensure its effectiveness and acceptance among students and educators. The positive experiences reported by the students in this study suggest a promising future for the integration of VR into clinical education. To maximize the potential of VR, institutions should consider investing in robust technical infrastructure, providing ongoing training and support for both students and faculty, and continuously evaluating and updating VR content to ensure its relevance and accuracy. This is supported by a study that assessed the educational applications of AR and VR across a variety of educational disciplines, concentrating on their impact on student motivation, learning results, engagement, and overall learning experiences [30]. Through simulations, this study investigated how AR and VR can improve information retention and skill acquisition while also encouraging active learning, collaboration, and critical thinking. The study also examined the potential of AR and VR in remote education and made practical recommendations for instructors to effectively incorporate these technologies into their teaching practices [30].

Strengths and Limitations

The qualitative research design used in this study was well-suited, as it enabled an in-depth exploration of students' experiences, perceptions, and contextual interpretations related to the use of VTRACS in their learning processes. One of the primary advantages of qualitative research in educational and media didactic contexts is its ability to capture the complexity and nuance of learner engagement with emerging technologies.

Through rich, narrative data, qualitative methods used in this study provide insight into how students interpret the effectiveness, usability, and emotional impact of digital learning tools, which are often not easily measurable through quantitative approaches. This is especially relevant when evaluating immersive technologies such as VR, where user experience, perceived realism, and pedagogical value can vary significantly across individuals and contexts.

However, qualitative approaches also have inherent limitations. First, the findings of this study are based on a relatively smaller, nonrepresentative sample, which limit generalizability. Second, the subjective nature of data collection and analysis introduces potential biases from both participants and researchers, despite efforts to ensure reflexivity and transparency. Third, in the context of media didactics, where technological tools are often applied across diverse educational settings, the findings should be interpreted with caution and viewed as context-dependent rather than universally applicable. Nevertheless, the insights gained from qualitative studies can serve as a foundation for further research and development.

Future Directions

Future work could build on these findings through mixed methods or quantitative study designs to validate trends observed in the qualitative data. Moreover, the findings can inform the iterative design of educational technologies, training programs for lecturers, and institutional implementation strategies, ensuring that pedagogical innovation remains grounded in the lived experiences of learners.

Conclusion

The use of VR in the teaching and assessment of clinical students is an emerging development in medical education in Nigeria. The use of VTRACS in the training and assessment of clinical students in a Nigerian university was perceived as a complementary learning method that provides continuous access to training and enhances proficiency. While there are hurdles such as cost and technological issues, students' overall favorable feedback demonstrates VR's promise as a valuable tool in health care professionals' education to promote objectivity and improve clinical abilities. Tackling these problems may unlock the full potential of VR in the health care professionals' education and the development of competent and confident health care professionals in Nigeria.

Acknowledgments

We thank all the students and faculty members who participated in the FGDs and in-depth interviews. The authors also acknowledge the financial support of the Tertiary Education Trust Fund under the National Research Fund Grant scheme (reference no TETF/DR&D/CE/NRF2020/CC/24/VOL.1). The authors declare that the funder did not influence the study design; data collection, analysis, or interpretation; writing of the paper; and decision to submit for publication.

Data Availability

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

Funding

This study was funded by the Tertiary Education Trust Fund through the National Research Fund Grant scheme (reference no TETF/DR&D/CE/NRF2020/CC/24/VOL.1). The funder had no role in study design, data collection, analysis, interpretation, or manuscript writing.

Authors' Contributions

AOK contributed to the conceptualization, funding acquisition, methodology, resources, investigation, formal analysis, validation, visualization, writing of the original draft, and review and editing of the manuscript. OAI contributed to the conceptualization, funding acquisition, methodology, resources, investigation, validation, and writing, review, and editing. OAO contributed to the conceptualization, funding acquisition, methodology, and resources. OM Adetutu contributed to the conceptualization, funding acquisition, and methodology. AOO contributed to the conceptualization, funding acquisition, software, and resources. OA contributed to the conceptualization, funding acquisition, software, and resources. OM Agunbiade contributed to the conceptualization, funding acquisition, methodology, formal analysis, and validation. ATA contributed to the conceptualization, funding acquisition, and methodology. SBA contributed to the conceptualization, funding acquisition, methodology, resources, and investigation. OBA contributed to the conceptualization, funding acquisition, and software. BAK contributed to the conceptualization, funding acquisition, methodology, and resources. LI contributed to the conceptualization, funding acquisition, and methodology. AAA contributed to the conceptualization, funding acquisition, and methodology. ODO contributed to the conceptualization, funding acquisition, methodology, resources, investigation, validation, project administration, and supervision.

Conflicts of Interest

None declared.

References

1. Ossai EN. Impact of COVID-19 on medical education and the challenges: how prepared is Nigeria? *Pan Afr Med J* 2020;37(Suppl 1):45 [FREE Full text] [doi: [10.11604/pamj.supp.2020.37.45.24915](https://doi.org/10.11604/pamj.supp.2020.37.45.24915)] [Medline: [33552373](https://pubmed.ncbi.nlm.nih.gov/33552373/)]
2. Komolafe AO, Ijarotimi OA, Adetutu OM, Okunola OA, Ojo TO, Oyinlola FF, et al. Acceptability, appropriateness and feasibility of webinar in strengthening research capacity in COVID-19 era in Nigeria. *Niger Postgrad Med J* 2022;29(4):288-295 [FREE Full text] [doi: [10.4103/npmj.npmj_167_22](https://doi.org/10.4103/npmj.npmj_167_22)] [Medline: [36308257](https://pubmed.ncbi.nlm.nih.gov/36308257/)]
3. Otuyemi OD, Jebose OJ. Experience and perceptions of Nigerian undergraduate dental students to virtual learning during the COVID-19 pandemic. *Niger J Clin Pract* 2023 May;26(5):566-573 [FREE Full text] [doi: [10.4103/njcp.njcp_265_22](https://doi.org/10.4103/njcp.njcp_265_22)] [Medline: [37357471](https://pubmed.ncbi.nlm.nih.gov/37357471/)]
4. Barteit S, Lanfermann L, Bärnighausen T, Neumann 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](https://doi.org/10.2196/29080)] [Medline: [34255668](https://pubmed.ncbi.nlm.nih.gov/34255668/)]
5. Muhamuza A, Naijuma JN, MacIntosh H, Sharma N, Singhal N, Hollaar GL, et al. Understanding the barriers and enablers for postgraduate medical trainees becoming simulation educators: a qualitative study. *BMC Med Educ* 2023 Jan 14;23(1):28 [FREE Full text] [doi: [10.1186/s12909-022-03995-3](https://doi.org/10.1186/s12909-022-03995-3)] [Medline: [36641526](https://pubmed.ncbi.nlm.nih.gov/36641526/)]
6. Elendu C, Amaechi DC, Okatta AU, Amaechi EC, Elendu TC, Ezech CP, et al. The impact of simulation-based training in medical education: A review. *Medicine (Baltimore)* 2024 Jul 05;103(27):e38813 [FREE Full text] [doi: [10.1097/MD.00000000000038813](https://doi.org/10.1097/MD.00000000000038813)] [Medline: [38968472](https://pubmed.ncbi.nlm.nih.gov/38968472/)]
7. Timotheou S, Miliou O, Dimitriadis Y, Sobrino SV, Giannoutsou N, Cachia R, et al. Impacts of digital technologies on education and factors influencing schools' digital capacity and transformation: A literature review. *Educ Inf Technol* (Dordr) 2023;28(6):6695-6726 [FREE Full text] [doi: [10.1007/s10639-022-11431-8](https://doi.org/10.1007/s10639-022-11431-8)] [Medline: [36465416](https://pubmed.ncbi.nlm.nih.gov/36465416/)]
8. Akinwale O, Abiona O, Oluwatope A, Otuyemi O, Ijarotimi O, Olubusola Komolafe A, et al. Designing a virtual reality system for clinical education and examination. *Computers & Education: X Reality* 2024 Dec 01;5(17):100083-100041 [FREE Full text] [doi: [10.1016/j.cexr.2024.100083](https://doi.org/10.1016/j.cexr.2024.100083)]
9. Richards S. Student Engagement Using HoloLens Mixed-Reality Technology in Human Anatomy Laboratories for Osteopathic Medical Students: an Instructional Model. *Med Sci Educ* 2023 Feb;33(1):223-231 [FREE Full text] [doi: [10.1007/s40670-023-01728-9](https://doi.org/10.1007/s40670-023-01728-9)] [Medline: [36691419](https://pubmed.ncbi.nlm.nih.gov/36691419/)]
10. Bridge P, Mehta J, Keane P, El-Sayed O, Mackay S, Ketterer S, et al. A virtual reality environment for supporting mental wellbeing of students on remote clinical placement: A multi-methods evaluation. *Nurse Educ Today* 2024 Jul;138:106184 [FREE Full text] [doi: [10.1016/j.nedt.2024.106184](https://doi.org/10.1016/j.nedt.2024.106184)] [Medline: [38537594](https://pubmed.ncbi.nlm.nih.gov/38537594/)]
11. Lau ST, Siah RCJ, Dzakirin Bin Rusli K, Loh WL, Yap JYG, Ang E, et al. Design and Evaluation of Using Head-Mounted Virtual Reality for Learning Clinical Procedures: Mixed Methods Study. *JMIR Serious Games* 2023 Aug 30;11:e46398 [FREE Full text] [doi: [10.2196/46398](https://doi.org/10.2196/46398)] [Medline: [37647108](https://pubmed.ncbi.nlm.nih.gov/37647108/)]
12. Kennedy G, Pedram S, Sanzone S. Improving safety outcomes through medical error reduction via virtual reality-based clinical skills training. *Safety Science* 2023 Sep 01;165(17):106200-106241 [FREE Full text] [doi: [10.1016/j.ssci.2023.106200](https://doi.org/10.1016/j.ssci.2023.106200)]
13. Kyaw BM, Saxena N, Posadzki P, Vseteckova J, Nikolaou CK, George PP, 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 [FREE Full text] [doi: [10.2196/12959](https://doi.org/10.2196/12959)] [Medline: [30668519](https://pubmed.ncbi.nlm.nih.gov/30668519/)]
14. Lakshminarayanan V, Ravikumar A, Sriraman H, Alla S, Chatterjee VK. Health Care Equity Through Intelligent Edge Computing and Augmented Reality/Virtual Reality: A Systematic Review. *J Multidiscip Healthc* 2023;16:2839-2859 [FREE Full text] [doi: [10.2147/JMDH.S419923](https://doi.org/10.2147/JMDH.S419923)] [Medline: [37753339](https://pubmed.ncbi.nlm.nih.gov/37753339/)]
15. Laspro M, Groysman L, Verzella A, Kimberly L, Flores R. The Use of Virtual Reality in Surgical Training: Implications for Education, Patient Safety, and Global Health Equity. *Surgeries* 2023 Nov 29;4(4):635-646 [FREE Full text] [doi: [10.3390/surgeries4040061](https://doi.org/10.3390/surgeries4040061)]
16. Mutch J, Golden S, Purdy E, Chang CHX, Oliver N, Tallentire VR. Equity, diversity and inclusion in simulation-based education: constructing a developmental framework for medical educators. *Adv Simul (Lond)* 2024 May 16;9(1):20 [FREE Full text] [doi: [10.1186/s41077-024-00292-5](https://doi.org/10.1186/s41077-024-00292-5)] [Medline: [38750552](https://pubmed.ncbi.nlm.nih.gov/38750552/)]
17. Afolabi OF, Ugwu CN, Eze CO, Serges-Anthanase DF, Onwuchekwa CO, Ude MA. Quality of Health Care in Resource Limited Settings-Patients' Satisfaction Survey at a Mission Hospital in South East, Nigeria. *Open J Intern Med* 2021;11(04):201-209 [FREE Full text] [doi: [10.4236/ojim.2021.114016](https://doi.org/10.4236/ojim.2021.114016)]
18. Andigema AS, Ngnokam N, Cyrielle T, Kamga N, Danaëlle L, Ekwelle E. Transforming African Healthcare with AI: Paving the Way for Improved Health Outcomes. *J Transl Med Epidemiol* 2024 Sep 01;7(1):1046. [doi: [10.47739/2333-7125/1046](https://doi.org/10.47739/2333-7125/1046)]
19. Rangarajan K, Davis H, Pucher PH. Systematic Review of Virtual Haptics in Surgical Simulation: A Valid Educational Tool? *J Surg Educ* 2020;77(2):337-347. [doi: [10.1016/j.jssurg.2019.09.006](https://doi.org/10.1016/j.jssurg.2019.09.006)] [Medline: [31564519](https://pubmed.ncbi.nlm.nih.gov/31564519/)]
20. Tan J, Karim M, Tamanna R, Kim S, Patel B. Comparing Learning Outcomes of Virtual Reality (VR) Simulators Using Haptic Feedback Versus Box Trainer (BT) in Laparoscopic Training: A Systematic Review and Meta-Analysis. *Cureus* 2025 Feb;17(2):e78910. [doi: [10.7759/cureus.78910](https://doi.org/10.7759/cureus.78910)] [Medline: [40091919](https://pubmed.ncbi.nlm.nih.gov/40091919/)]

21. Chahal B, Aydin A, Ahmed K. Virtual reality vs. physical models in surgical skills training. An update of the evidence. *Curr Opin Urol* 2024 Jan 01;34(1):32-36 [FREE Full text] [doi: [10.1097/MOU.0000000000001145](https://doi.org/10.1097/MOU.0000000000001145)] [Medline: [37962178](https://pubmed.ncbi.nlm.nih.gov/37962178/)]
22. Marks B, Thomas J. Adoption of virtual reality technology in higher education: An evaluation of five teaching semesters in a purpose-designed laboratory. *Educ Inf Technol* (Dordr) 2022;27(1):1287-1305 [FREE Full text] [doi: [10.1007/s10639-021-10653-6](https://doi.org/10.1007/s10639-021-10653-6)] [Medline: [34257511](https://pubmed.ncbi.nlm.nih.gov/34257511/)]
23. Bumbach MD, Culross BA, Datta SK. Assessing the Financial Sustainability of High-Fidelity and Virtual Reality Simulation for Nursing Education: A Retrospective Case Analysis. *Comput Inform Nurs* 2022 Sep 01;40(9):615-623. [doi: [10.1097/CIN.0000000000000916](https://doi.org/10.1097/CIN.0000000000000916)] [Medline: [35524349](https://pubmed.ncbi.nlm.nih.gov/35524349/)]
24. Ogwu EN, Emelogo NU, Azor RO, Okwo FA. Educational Technology Adoption in Instructional Delivery in the New Global reality. *Educ Inf Technol* (Dordr) 2023;28(1):1065-1080 [FREE Full text] [doi: [10.1007/s10639-022-11203-4](https://doi.org/10.1007/s10639-022-11203-4)] [Medline: [35875827](https://pubmed.ncbi.nlm.nih.gov/35875827/)]
25. Salimi Z, Ferguson-Pell MW. Motion sickness and sense of presence in a virtual reality environment developed for manual wheelchair users, with three different approaches. *PLoS One* 2021;16(8):e0255898 [FREE Full text] [doi: [10.1371/journal.pone.0255898](https://doi.org/10.1371/journal.pone.0255898)] [Medline: [34411151](https://pubmed.ncbi.nlm.nih.gov/34411151/)]
26. Chattha U, Janjua U, Anwar F, Madni T, Cheema M, Janjua S. Motion Sickness in Virtual Reality: An Empirical Evaluation. *IEEE Access* 2020 Dec;8(6):130486-130499. [doi: [10.1109/ACCESS.2020.3007076](https://doi.org/10.1109/ACCESS.2020.3007076)]
27. Domeyer JE, Cassavaugh ND, Backs RW. The use of adaptation to reduce simulator sickness in driving assessment and research. *Accid Anal Prev* 2013 Apr;53:127-132. [doi: [10.1016/j.aap.2012.12.039](https://doi.org/10.1016/j.aap.2012.12.039)] [Medline: [23416680](https://pubmed.ncbi.nlm.nih.gov/23416680/)]
28. Mackrouss I, Lavallière M, Teasdale N. Adaptation to simulator sickness in older drivers following multiple sessions in a driving simulator. *Gerontechnology* 2014 Sep 01;12(2):101-111. [doi: [10.4017/gt.2013.12.2.004.00](https://doi.org/10.4017/gt.2013.12.2.004.00)]
29. Kasegawa C, Itaguchi Y, Yamawaki Y, Miki M, Hayashi M, Miyazaki M. Effects of within-day intervals on adaptation to visually induced motion sickness in a virtual-reality motorcycling simulator. *Sci Rep* 2024 Sep 22;14(1):21302 [FREE Full text] [doi: [10.1038/s41598-024-71526-9](https://doi.org/10.1038/s41598-024-71526-9)] [Medline: [39307847](https://pubmed.ncbi.nlm.nih.gov/39307847/)]
30. AlGerafi M, Zhou Y, Oubibi M, Wijaya T. Unlocking the Potential: A Comprehensive Evaluation of Augmented Reality and Virtual Reality in Education. *Electronics* 2023 Sep 20;12(18):3953-3941. [doi: [10.3390/electronics12183953](https://doi.org/10.3390/electronics12183953)]

Abbreviations

AR: augmented reality

EDI: equity, diversity, and inclusion

FGD: focus group discussion

ICT: information and communication technology

IDI: in-depth interview

VR: virtual reality

VTRACS: Virtual reality model to TRain and Assess Clinical Students

Edited by T Leung; submitted 27.Mar.2025; peer-reviewed by D Lerner, R Omari; comments to author 02.Jul.2025; revised version received 14.Aug.2025; accepted 30.Nov.2025; published 26.Jan.2026.

Please cite as:

Komolafe AO, Ijarotimi OA, Okunola OA, Adetutu OM, Oluwatope AO, Abiona O, Agunbiade OM, Ayinde AT, Aregbesola SB, Akinwale OB, Kolawole BA, Idowu L, Adediwura AA, Otuyemi OD

Virtual Reality in Training and Assessment Among Clinical Students and Lecturers at a Nigerian University: A Phenomenological Study

JMIR Med Educ 2026;12:e75021

URL: <https://mededu.jmir.org/2026/1/e75021>

doi: [10.2196/75021](https://doi.org/10.2196/75021)

PMID:

©Abiola Olubusola Komolafe, Omotade Adebimpe Ijarotimi, Oluseye Ademola Okunola, Olufemi Mayowa Adetutu, Ayodeji Oludola Oluwatope, Olatunde Abiona, Ojo Melvin Agunbiade, Adeboye Titus Ayinde, Stephen Babatunde Aregbesola, Olawale Babatunde Akinwale, Babatope Ayodeji Kolawole, Lanre Idowu, Alaba Adeyemi Adediwura, Olayinka Donald Otuyemi. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 26.Jan.2026. 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.

Live Podcasting as an Educational Intervention in Dentomaxillofacial Radiology: Controlled Cohort Study

Anna-Lena Hillebrecht¹, Dr Med Dent; Daniel Fritzsche², Med Dent; Thamar Voss³, Dipl Psych, Prof Dr phil; Anne Kruse², Dr Med Dent, Priv Doz; Andreas Keßler^{1,4}, Dr Med Dent, Priv Doz, MSc; Kirstin Vach⁵, Dipl Math, PhD; Markus Jörg Altenburger⁶, Prof Dr Med Dent; Rainer Schmelzeisen⁷, Prof Dr Med, Dr Med Dent; Wiebke Semper-Hogg⁷, Dr Med Dent, Priv Doz

¹Department of Prosthetic Dentistry, Center for Dental Medicine, Faculty of Medicine, Medical Center, University of Freiburg, Hugstetter Strasse 55, Freiburg, Germany

²Department of Operative Dentistry and Periodontology, Center for Dental Medicine, Faculty of Medicine, Medical Center, University of Freiburg, Freiburg, Germany

³Department of Educational Science, University of Freiburg, Freiburg, Germany

⁴Department of Conservative Dentistry and Periodontology, University Hospital, Ludwig-Maximilians University of Munich, Munich, Germany

⁵Institute of Medical Biometry and Statistics, Medical Center, University of Freiburg, Freiburg, Germany

⁶Department of Orthodontics, Center for Dental Medicine, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany

⁷Department of Oral and Maxillofacial Surgery, Center for Dental Medicine, Faculty of Medicine, Medical Center, University of Freiburg, Freiburg, Germany

Corresponding Author:

Anna-Lena Hillebrecht, Dr Med Dent

Department of Prosthetic Dentistry, Center for Dental Medicine, Faculty of Medicine, Medical Center, University of Freiburg, Hugstetter Strasse 55, Freiburg, Germany

Abstract

Background: Podcasts are increasingly used in health professions education; however, most formats are asynchronous and noninteractive. Didactically grounded, synchronous implementations in dental curricula are scarce.

Objective: This study aims to design, implement, and evaluate a synchronous, case-based live podcast (LP) as a didactic teaching format in dentomaxillofacial radiology.

Methods: In a controlled cohort study with 2 third-year cohorts (N=41), the intervention group (IG; n=21, 51%) received weekly case-based LP sessions in addition to standard teaching, while the control group (CG; n=20, 49%) received standard teaching only. Acceptability was evaluated 6 months postcourse using the 27-item student evaluation questionnaire and open-text responses. Knowledge was assessed immediately after the course with a 21-item radiology knowledge test, and after 6 months, with a 15-item interdisciplinary clinical application test.

Results: The primary outcome was student-reported acceptability of the LP format. It was rated highly by students in the Student Evaluation Questionnaire (mean out of 10: structure 9.76, interactivity 9.62, interdisciplinary relevance 9.55). Qualitative feedback was assessed highlighting motivation, authenticity, and discussion quality. In the radiology knowledge test, no group differences were observed (IG: n=21, 51% vs CG: n=20, 49%; $P=.37$). In the interdisciplinary clinical application test, the IG outperformed the CG in restorative dentistry (median 5, IQR 4-5 vs median 4, IQR 3-5; $P=.02$; $r=0.38$) and in item-level analysis (15/21, 71% vs 40%; $P=.04$; $\phi=0.64$).

Conclusions: The LP format represents a feasible, scalable, and low-threshold approach to fostering clinical reasoning in dental curricula, particularly at the transition to clinical training. While radiology-specific theoretical competencies did not differ between the groups, students consistently rated the LP as more engaging and motivating compared to standard lectures.

(*JMIR Med Educ* 2026;12:e77980) doi:[10.2196/77980](https://doi.org/10.2196/77980)

KEYWORDS

case-based learning; clinical reasoning; dental education; digital learning; health professions education; interdisciplinary teaching; podcast; student engagement; synchronous learning

Introduction

Overview

In the landscape of health professions education, digital audio formats, particularly podcasts, have gained substantial momentum [1-5]. Their flexibility, accessibility, and conversational tone have made them an increasingly attractive medium for knowledge dissemination and learner engagement [4,6]. While podcasts are now widely accepted as a complementary teaching method in undergraduate medical education, their use remains predominantly asynchronous and prerecorded, with limited interactivity and integration into curricular strategies [1]. Despite their potential, podcasts have rarely been explored as intentionally designed didactic tools grounded in contemporary educational theory. While the demand for flexible and digital teaching formats in dental education continues to rise, there is a striking lack of pedagogically grounded concepts specifically addressing the use of podcasts within the dental curriculum [5,7]. To better understand students' needs in modern education, a preliminary needs assessment was conducted among undergraduate dental students at the University of Freiburg. The results indicated that while students appreciated classical media such as lectures and textbooks for foundational learning, they strongly favored digital formats for more advanced clinical content. Podcasts, although not yet widely used for learning, were perceived as a promising format for flexible, self-directed studies. These findings highlighted a clear interest in diverse, interactive, and clinically relevant digital learning formats [8].

Rationale and Objectives

Recent initiatives such as the student-led technology committee highlight the growing recognition of students as active stakeholders in shaping digital education, demonstrating how learner involvement can drive innovation and relevance in curricular technologies [9]. At the same time, in dentistry, clinical reasoning (CR) is gaining increasing importance, particularly as dental treatment needs are rising among vulnerable populations, including older adults and patients with complex medical conditions [10,11]. CR is particularly evident in dentomaxillofacial radiology (DMFR), where diagnostic decisions based on radiographic interpretation significantly influence treatment planning. CR in this context includes not only the technical analysis of imaging data but also interdisciplinary judgment, risk assessment, and patient-centered decision-making. CR refers to the cognitive process by which health care professionals collect, interpret, and synthesize patient information to make informed diagnostic and therapeutic decisions. CR remains a central yet pedagogically challenging competence in medical and dental education. Developing CR requires authentic, situated learning experiences that bridge disciplines and promote critical discussion, an area in which traditional lecture-based formats often fall short.

In recent years, traditional lecture-based formats have increasingly lost their appeal among students, raising the question of whether innovative and more interactive approaches can reengage learners and potentially revitalize attendance in academic teaching. This study introduces and evaluates the live

podcast (LP) as an innovative digital teaching format that combines the accessibility of podcasting with the didactic strengths of interactive, case-based learning. To our knowledge, this is the first systematic implementation of an LP format within the dental curriculum. The LP was not conceived as a mere content delivery tool but as a didactically grounded educational intervention designed to foster clinical reasoning, interdisciplinary thinking, and student engagement through authentic expert dialogue. The objective was to introduce a didactically designed LP format within undergraduate DMFR and evaluate its feasibility and student-reported acceptability and to explore potential effects on theoretical and interdisciplinary knowledge compared with standard teaching.

The LP format was hypothesized to be feasible, well accepted by students, and to improve knowledge scores in domains requiring case-based reasoning compared with standard instruction.

Methods

This study primarily presents and evaluates the LP as a novel teaching format in DMFR education, focusing on its instructional design, implementation, and initial effectiveness.

Curricular Context

Undergraduate dental education in Germany is structured as a 5-year program, consisting of a 2-year preclinical phase (basic sciences such as anatomy, physiology, biochemistry, and histology, along with introductory dental subjects), followed by a 3-year clinical phase. The curriculum is nationally regulated and guided by the National Competence-Based Learning Objectives Catalogue in Dentistry. DMFR is introduced at the beginning of the clinical training (fifth semester) in close connection with basic dental instruction and is subsequently applied throughout the clinical phase. The course includes radiographic techniques, diagnostic interpretation, radiation protection, and ethical considerations, integrating theoretical instruction with practical training (intraoral radiographs, including periapical radiographs, bitewing radiographs, and occlusal radiographs; panoramic radiographs such as orthopantomogram; and cone-beam computed tomography). The LP sessions were designed based on National Competence-Based Learning Objectives Catalogue in Dentistry objectives, particularly Z 21 (clinical diagnostics, radiographic imaging, and radiation protection) and Z 5 (dentist as medical expert), thereby ensuring consistency with national competency standards.

Didactic and Conceptual Design

The LP was developed as a synchronous, case-based learning format and implemented as a mandatory component of the DMFR course for third-year undergraduate dental students at the University of Freiburg. Each weekly 45-minute session followed a structured design: anonymized real-life cases were discussed in moderated dialogue between a dental radiologist and a specialist from a complementary dental or medical discipline (eg, pediatric dentistry, prosthodontics, oral surgery, and radiology). Discussions focused on radiographic diagnosis, differential diagnoses, and interdisciplinary treatment planning.

All episodes were designed with explicit learning goals that line with the interdisciplinary nature of DMFR (Tables 1 and emphasized diagnostic competence and clinical reasoning, in 2).

Table . Curriculum framework for the live podcast: episodes, topics, disciplines, imaging modalities (periapical radiographs [PA], bitewing radiographs [BW], occlusal radiographs [OC], panoramic radiographs such as orthopantomogram [OPG], and cone-beam computed tomography [CBCT]), and radiographic diagnoses.

Episode	Topic	Discipline	Imaging modalities in this episode	Radiographic diagnoses
1	Pediatric dentistry	Pediatric dentistry	PA; BW; OPG	Approximal caries and apical periodontitis
2	Endodontontology	Endodontics	PA; CBCT for complex cases	Irreversible pulpitis and apical periodontitis
3	Caries diagnostics	Restorative dentistry	BW; PA	Bitewing caries; secondary caries versus artifacts
4	General prosthodontic planning	Prosthodontics	OPG; PA	Tooth loss, apical periodontitis, bone loss, and TMJ ^a findings
5	Special care dentistry	Special care dentistry	OPG; PA	Tooth loss, caries, bone loss, and radiation protection issues
6	General radiology	Oral and maxillofacial radiology	PA; BW; OPG; CBCT; MRI ^b	Modality characteristics (intraoral, BW, CBCT, and MRI)
7	History of dentistry	History of medicine or dentistry	Historical imaging methods	Early radiographs and OPG
8	Artificial intelligence in dentistry	Informatics; oral radiology	PA; OPG; CBCT	AI ^c -assisted detection of caries and periodontal bone loss
9	Orthodontics	Orthodontics	OPG; lateral cephalogram; PA; OC	Tooth development; retained teeth
10	Periodontology	Periodontology	PA; BW; OPG	Horizontal and vertical bone loss
11	Surgical removal of wisdom teeth	Oral and maxillofacial surgery	OPG; CBCT for complex cases	Impactions, cysts, and nerve injury risk

^aTMJ: temporomandibular joint.

^bMRI: magnetic resonance imaging.

^cAI: artificial intelligence.

Table . Curriculum framework for the live podcast: episodes, topics, covered domains, interdisciplinary aspects, and learning goals.

Episode	Topic	Covered domains	Interdisciplinary aspects	Learning goals
1	Pediatric dentistry	Imaging and diagnostics in primary and mixed dentition, trauma management, prevention, and developmental diagnostics	Pediatrics, and oral and maxillofacial surgery (fracture management)	Apply pediatric-specific diagnostic strategies, and select imaging in pediatric trauma and explain informed consent in minors
2	Endodontontology	Pain diagnostics, pulp vitality assessment, and decision-making in endodontic emergencies	Oral and maxillofacial surgery and restorative dentistry	Strengthen case-based reasoning and interpret findings in irreversible pulpitis and differentiate periapical pathologies
3	Caries diagnostics	Primary and secondary caries, diagnostics, and prevention	Prevention, pediatrics, prosthodontics, and endodontics	Enhance diagnostic accuracy and apply imaging in caries diagnostics and distinguish true lesions from imaging artifacts
4	General prosthodontic planning	Prosthetic concepts, treatment planning, and radiographic considerations	Geriatrics, orthodontics, oral and maxillofacial surgery, and restorative dentistry	Integrate diagnostics into treatment planning, and use OPG ^a in planning and develop interdisciplinary prosthodontic strategies
5	Special care dentistry	Prosthetic care for people with disabilities, barriers to care, and ethics	Medicine, nursing, and ethics (focus on vulnerable and medically complex patients)	Link ethics with radiology, and apply radiation protection in vulnerable groups and critically reflect on imaging indications
6	General radiology	Physics; ALARA ^b principle; imaging modalities	Cross-disciplinary relevance of imaging decisions	Build safety and modality competence, and apply ALARA and differentiate indications for imaging modalities
7	History of dentistry	Evolution of diagnostics and imaging	History of medicine	Promote scientific reflection, and describe historical developments in dental diagnostics and radiology
8	Artificial intelligence in dentistry	Digital diagnostics and decision support	Ethics and data science	Foster critical, future-oriented thinking; evaluate benefits and limitations of AI ^c -assisted diagnostics
9	Orthodontics	Malocclusions and growth diagnostics	Pediatrics and prosthodontics	Relate radiology to growth analysis, and explain OPG or CBCT ^d and model analysis in orthodontics and assess impacted teeth
10	Periodontology	Periodontal diagnostics, disease progression, and treatment planning	Prosthodontics and implantology	Strengthen diagnostic integration, and identify periodontal bone loss radiographically and link findings to implant and prosthodontic planning
11	Surgical removal of wisdom teeth	Indications, surgical techniques, and preoperative planning	Radiology, anesthesiology, and prosthodontics	Link diagnostics with surgery, and select OPG versus CBCT appropriately and assess surgical risks radiographically

^aOPG: orthopantomogram.^bALARA: As Low As Reasonably Achievable.^cAI: artificial intelligence.^dCBCT: cone-beam computed tomography.

Technical Setup

Sessions were held in a lecture hall with professional audiovisual infrastructure. At least 2 microphones transmitted the voices of the podcasters via the hall sound system, and an additional microphone was available for student interaction. A projector displayed radiographs and clinical photographs. Each session was recorded live with sound levels adjusted by a technician. Recordings were postprocessed in Adobe Audition (version 23.11; Adobe Inc) and uploaded via Panopto (version 14.0; Panopto Inc) to the official university learning platform, ensuring high audio quality and asynchronous access for enrolled students ([Multimedia Appendix 1](#)).

Participants and Study Design

All third-year dental students enrolled in the mandatory DMFR course at the University of Freiburg were eligible. In total, 41 students participated in 2 consecutive cohorts: a control group (CG; n=20, 49%; with n=12, 60% women and n=8, 40% men) that attended traditional lectures covering DMFR principles, radiographic techniques, and clinical relevance, and an intervention group (IG; n=21, 51%; with n=11, 52% women and n=10, 48% men) that attended the LP sessions. No exclusion criteria were applied because all students were required to complete the course as part of the curriculum. The cohorts received comparable curricular content from the same teaching staff, ensuring baseline comparability of exposure. No pretest was conducted because both groups had identical prior curricular exposure. All students completed the subsequent knowledge assessments. Item-level correctness and point-based domain scores were analyzed across restorative dentistry, oral surgery, and orthodontics. The 3 assessment instruments used in this study were the interdisciplinary clinical application test (ICAT; [Multimedia Appendix 2](#)), the radiology knowledge test (RKT; [Multimedia Appendix 3](#)), and the Student Evaluation Questionnaire (SEQ; [Multimedia Appendix 4](#)).

Assessment of Student-Reported Acceptability

To determine the primary outcome, student-reported acceptability was evaluated 6 months after the LP using an anonymous questionnaire with 27 rating items (SEQ), on a 10-point unipolar scale (1=I strongly disagree, 10=I strongly agree). The questionnaire covered organizational aspects, atmosphere, interactivity, memorability, long-term benefit, integration of theory and practice, reflective engagement, social learning, and overall evaluation. Selected domains were stratified by gender. An optional open-text field captured qualitative impressions that were analyzed descriptively.

Assessment of Learning Outcomes

Knowledge served as the secondary outcome and was evaluated at 2 time points to capture both immediate and longer-term learning effects.

Immediately after the course: IG and CG completed a DMFR radiology-specific competence test (RKT), with 21 single-choice items (F1-F21) covering radiation protection, diagnostic imaging, and interpretation.

Six months after the course, IG and CG participated in an interdisciplinary multiple-choice test (ICAT) with 15 case-based

items (5 items each for restorative dentistry, oral surgery, and orthodontics), addressing radiographic interpretation, interdisciplinary treatment planning, and clinical application of radiological findings.

Statistical Analysis

No formal sample size calculation was performed. As a feasibility study embedded in routine teaching, the available cohort size (N=41) defined the sample. This aligns with the exploratory aim to evaluate acceptability and didactic value rather than to demonstrate definitive knowledge effects. Descriptive statistics (mean, SD, median, and IQR) were calculated for all outcomes. Chi-square tests were applied for item-specific correctness, and Wilcoxon rank-sum tests were used for group comparisons of point-based scores and gender differences. The Fisher exact test was applied in the case of single-choice items. Effect sizes (*r*) were reported where appropriate. Given the exploratory design, no adjustment for multiple testing was applied. Statistical analyses were performed in Stata (version 19; StataCorp) with a significance level of $\alpha=0.05$.

Ethical Considerations

The study involved human participants and was conducted in accordance with institutional and national research regulations and the Declaration of Helsinki. The study was reviewed informally by the ethics committee of the University of Freiburg (reference number: 25-1479). As it served internal educational quality assurance without collection of personal or health-related data, formal approval was not required. Informed consent was obtained from all participants during their voluntary enrollment in the course after they had received detailed information about the study purpose, procedures, data use, and their right to withdraw at any time without negative consequences. Data were collected and analyzed anonymously and treated confidentially in accordance with data protection regulations. Participation was voluntary and no financial or other compensation was provided.

Results

Overview

A total of 41 third-year dental students participated in the study (IG: n=21, 51%; CG: n=20, 49%). All students completed the RKT immediately after the course, and all participated in the ICAT interdisciplinary follow-up test 6 months later. Student-reported acceptability of the LP format was assessed 6 months postcourse; 21 students (100% of the IG) completed the SEQ.

Primary Outcome: Student-Reported Acceptability

To determine student-perceived acceptability of the LP format, 21 students completed the SEQ 6 months after the course. Overall, the LP received consistently high ratings across all evaluated dimensions, including atmosphere, organization, perceived usefulness, and interdisciplinary relevance. The mean scores were above 9/10 in nearly all categories, indicating very high acceptance and perceived educational value. No meaningful gender-related differences were observed ([Table 3](#)).

Table . Student Evaluation Questionnaire (SEQ): evaluation of the live podcast in different categories stratified by gender.

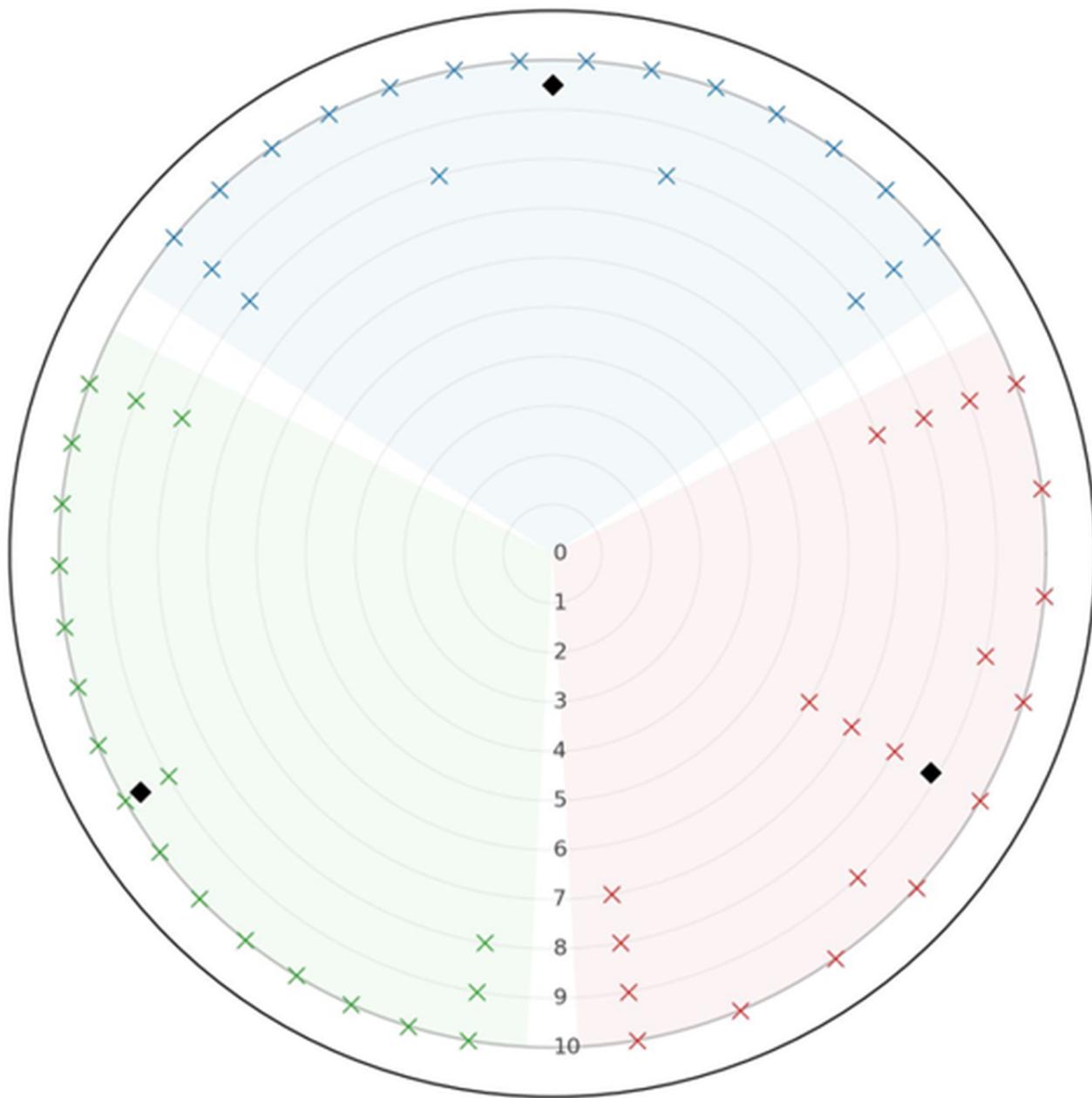
Category	Women (n=11; score: 1-10), mean (SD)	Men (n=10; score: 1-10), mean (SD)
Atmosphere	9.91 (0.30)	9.40 (0.97)
Organization	9.82 (0.40)	9.70 (0.48)
Useful teaching method	9.73 (0.65)	9.56 (0.73)
Scope of knowledge	9.18 (1.32)	9.56 (1.01)
Interdisciplinary connections	9.45 (0.82)	9.67 (0.50)
Personal perception	6.64 (1.21)	7.67 (1.94)
Addition to the lecture	8.64 (1.75)	8.78 (1.39)
Interest in topics	9.91 (0.30)	9.30 (0.95)

No meaningful gender-related differences were identified for the total evaluation score ($P=0.18$) or for any subscale ($P=0.32-0.79$).

Student ratings for the 3 selected evaluation items as illustrative examples are shown in [Figure 1](#). Overall, the majority of participants provided high ratings, indicating strong agreement across all 3 items. Specifically, for “learning from the

experiences of others,” 20% (n=4) of participants rated 8, 10% (n=2) rated 9, and 70% (n=14) rated 10. For “encouraged reflection,” 5% (n=1) rated 6, 15% (n=3) rated 7, 15% (n=3) rated 8, 20% (n=4) rated 9, and 45% (n=9) rated 10. For “useful teaching method,” 10% (n=2) rated 8, 15% (n=3) rated 9, and 75% (n=15) rated 10. All other rating categories received no responses.

Figure 1. Distribution of participant ratings for the Student Evaluation Questionnaire (SEQ) items. The radial plot visualizes the distribution of all ratings (1-10) for the items “learning from the experiences of others” (blue), “encouraged reflection” (red), and “useful teaching method” (green; each $n=20$). Each cross represents 1 individual response, while black diamonds (◆) indicate mean values. No ratings were given for scores 1-5. For “learning from the experiences of others,” 20% ($n=4$) of participants rated 8, 10% ($n=2$) rated 9, and 70% ($n=14$) rated 10. For “encouraged reflection,” 5% ($n=1$) rated 6, 15% ($n=3$) rated 7, 15% ($n=3$) rated 8, 20% ($n=4$) rated 9, and 45% ($n=9$) rated 10. For “useful teaching method,” 10% ($n=2$) rated 8, 15% ($n=3$) rated 9, and 75% ($n=15$) rated 10. All other rating categories received no responses. A unipolar rating scale from 1 (I strongly disagree) to 10 (I strongly agree) was used, with higher values indicating a more favorable evaluation.



Qualitative Findings

In addition to the quantitative results, students provided written feedback at the end of the course. Selected anonymized comments (translated from German) illustrate how the LP was perceived:

The podcasts were scheduled at a good time before lunch, which encouraged further discussion of the topics in the cafeteria afterwards. The content was very stimulating.

In some sessions, time was even too short, as the discussions were highly engaging. A second microphone for students would sometimes have been helpful.

I sincerely appreciate the effort that went into these sessions. It was a great experience to learn the material in such an interesting way through contributions from many different lecturers.

Through the podcast I realized how much dentistry has to offer. The podcast, with its many insights into different specialties, gave me renewed motivation and interest in my studies. After four semesters of rote

learning, there could not have been a better introduction to the clinical phase.

I really enjoyed attending the podcast lectures. They were varied and never felt like dry teaching, yet I learned a great deal. The relaxed atmosphere, such as sitting on the sofa, created an environment that did not feel like studying, although one gained a lot of knowledge.

These student voices indicate that the LP sessions were experienced as innovative, motivating, and discussion-stimulating, creating a positive learning atmosphere at the transition from preclinical to clinical studies.

Secondary Outcomes: Knowledge Acquisition

In the ICAT, item-level analysis revealed a significantly higher proportion of correct responses in the IG compared with the CG

in restorative dentistry ($P=.04$; $\phi=0.64$). No statistically significant differences were observed in oral surgery ($P=.85$; $\phi=0.005$) or orthodontics ($P=.14$; $\phi=0.34$).

Across all ICAT domains combined, no statistically significant difference in overall performance was observed between groups ($P=.37$; $\phi=0.49$). A distribution-based analysis showed that in the IG, 7 students answered 3 items correctly, 13 students answered 2 items correctly, and 1 student answered 1 item correctly, with no student scoring 0 correct answers. In contrast, in the CG, 4 students answered 3 items correctly, 12 students answered 2 items correctly, 2 students answered 1 item correctly, and 2 students answered no items correctly. This distribution indicates a trend toward higher overall performance in the IG, although the difference did not reach statistical significance (Table 4).

Table 4. Interdisciplinary clinical application test (ICAT): point-based domain scores by group (Wilcoxon rank-sum test; $\alpha=.05$)^a.

Domain	IG ^b (points), median (IQR)	IG (points), mean (SD)	CG ^c (points), median (IQR)	CG (points), mean (SD)	Effect size r
Restorative dentistry	5 (4-5)	4.62 (0.67)	4 (3-5)	3.75 (1.25)	0.38
Oral surgery	5 (3-5)	3.95 (1.46)	5 (3.5-5)	4.10 (1.37)	0.05
Orthodontics	5 (5-5)	5.00 (0.00)	5 (5-5)	4.60 (1.23)	0.23

^aMedian, IQR, mean, and SD values are presented for the intervention and control groups. The effect size (r) is shown for each domain. Only the comparison in restorative dentistry reached statistical significance ($P=.02$). No significant differences were found for oral surgery or orthodontics.

^bIG: intervention group.

^cCG: control group.

No statistically significant differences were observed in the RKT between IG and CG for the overall score or any subdomain (all $P>.05$). The total RKT score was numerically higher in the IG (median 14, IQR 13-15) than in the CG (median 13, IQR 11-14), but this difference did not reach statistical significance ($P=.17$; $r=0.22$), indicating comparable short-term acquisition of radiology-specific theoretical competences in both groups.

In the ICAT, analysis of item-level correctness revealed a significantly higher performance in the IG (75/105, 71%) than in the CG (40/100, 40%) in restorative dentistry ($P=.04$; $\phi=0.64$). No significant differences were observed in oral surgery ($P=.85$; $\phi=0.005$) or orthodontics ($P=.14$; $\phi=0.34$). Overall correctness across all domains was slightly higher in the IG (161/315, 51%) compared with the CG (147/300, 49%), though this difference was not statistically significant ($P=.37$; $\phi=0.49$; Table 4).

Discussion

Principal Findings

This study demonstrates the development and feasibility of a novel teaching format in undergraduate dental education: the live podcast. To our knowledge, this is the first systematically designed and evaluated synchronous podcast format embedded within a dental curriculum. Unlike traditional podcasts, the LP was conceived as a didactically grounded intervention that combines digital accessibility with case-based, interactive learning. Our results provide initial evidence that LPs are feasible and highly accepted. The consistently high SEQ ratings

in structure, clarity, and interactivity suggest that LPs foster an engaging learning environment, particularly at the transition from preclinical to clinical training.

Students in the intervention group performed significantly better in the ICAT in the restorative dentistry domain, which indicates that authentic case discussions supported by expert dialogue may enhance diagnostic reasoning. Although no significant differences were observed across all domains in ICAT, descriptive trends point to positive effects on interdisciplinary understanding. Beyond knowledge outcomes, very high acceptance ratings and open-ended comments show that students perceived the LP as motivating, discussion-enhancing, and well aligned with their learning preferences. In line with contemporary student-voice literature, the LP addressed the cognitive dimension of clinical reasoning and also fostered engagement, reflexivity, and ownership through its interactive and dialogic structure [12]. Goh and Lim [13] argue that dental learning is increasingly conceptualized as situated and relational. The LP exemplifies this by embedding reasoning in authentic, interdisciplinary practice.

The RKT did not reveal significant differences between the 2 groups. For theoretical radiology competencies, both formats appear equally effective. Taken together with the very high acceptance ratings in the SEQ, this suggests that the added value of the LP lies less in measurable short-term knowledge gains and more in providing a motivating, engaging, and student-centered learning environment with an interdisciplinary focus [1].

Evidence from this study suggests that synchronous, audio-based teaching formats such as LPs can act as effective didactic tools in complex, interdisciplinary areas of dental education, including DMFR. By fostering active engagement, real-time dialogue, and case-based reflection, LPs extend the potential of both traditional lectures and asynchronous e-learning. The positive reception aligns with prior literature on the educational benefits of podcasts. Kaplan et al [14] emphasize brevity, narrative structure, and authentic expert dialogue as advantages aligned with adult learning principles and conducive to learner motivation and retention. Consistent with these findings, students in our study reported in the SEQ high levels of didactic clarity, structure, and perceived longer-term benefit, dimensions repeatedly highlighted in podcast-based education [15]. The conversational and interdisciplinary character of the LP mirrors the informal, yet focused tone described by Kaplan et al [14] as central to engagement and professional connectedness. Narrative and storytelling, which support sense-making, empathy, and professional identity formation [16], were intentionally embedded through authentic case discussions. Qualitative feedback indicated that these elements were experienced as engaging and professionally meaningful, resonating with sociocultural learning theory [14,17]. In line with work that identifies podcasts as vehicles for democratizing discourse and fostering reflection [18], live podcasting may strengthen interdisciplinary awareness, reflective reasoning, and inclusive dialogue in dental education.

Core disciplinary knowledge was tested in the RKT immediately after the course to capture short-term acquisition of radiology-specific competences in line with constructive alignment principles. Interdisciplinary case-based knowledge in the ICAT, by contrast, was assessed after 6 months because durable learning and knowledge transfer typically manifest after a consolidation phase. By separating immediate mastery from delayed application, the design aimed to minimize test contamination and to capture distinct learning outcomes targeted by the LP. The combination of synchronous interaction and asynchronous access corresponds to evolving learner expectations for hybrid, customizable formats [19]. The ability to revisit discussions, catch up on missed sessions, and selectively review content increases long-term educational use and accommodates diverse learning needs. These findings are consistent with prior institutional needs assessments showing strong preferences for flexible, multimodal tools [8].

Didactic Implications for Teaching Practice

LPs are a cost-effective, low-threshold format that integrates readily into blended learning environments. The synchronous,

case-based design suits the development of clinical reasoning and interdisciplinary thinking. Opportunities for learner questions and moderated dialogue promote interaction, reflection, and active engagement. High ratings in structure, relevance, and perceived benefit are known drivers of acceptance in higher education [20]. Embedding LPs as curriculum-aligned supplements, supported by preparatory and follow-up materials, can deepen learning and support knowledge retention. LPs are intended to complement, not replace, lectures so that students benefit from interactive digital formats while still using the lecture hall for consolidation and further integration of knowledge. Because LPs rely on existing audiovisual infrastructures and can be easily streamed, recorded, and integrated into learning management systems, this teaching format is scalable and transferable to other health professions programs.

Limitations

The sample was small and restricted to a single center, which limits generalizability. Group allocation used consecutive cohorts rather than randomization, which may introduce selection bias. No pretest of baseline knowledge was conducted, and individual prior knowledge could not be controlled. The ICAT comprised 15 multiple-choice items and assessed midterm retention only, so effects on long-term learning and clinical performance remain unclear. The evaluation of student-reported acceptability in the SEQ relied on self-report, which may be affected by social desirability.

Future Directions

Future research should use randomized, adequately powered, multicenter designs with baseline testing and correction for multiple comparisons. A longer follow-up is needed to examine durability. Performance-based outcomes such as objective structured clinical examinations and workplace-based assessments should be included to evaluate transfer to clinical decision-making and patient care. Studies could also compare LP design variants, for example, the degree of interactivity, scaffolding materials, and image integration.

Conclusions

The LP was feasible and well accepted in the undergraduate DMFR course. It was associated with a significant knowledge gain in restorative dentistry and yielded results comparable to lectures in radiology-specific competences, while students rated it as more engaging and motivating. LPs, therefore, represent a scalable complement to standard instruction with potential to strengthen clinical reasoning and student-centered learning.

Acknowledgments

The authors thank Jonas Wagner, Sebastian Ollinger, Finn Weisser, Cumhur Yörük, and Robin Gommel for advice and technical support, the interdisciplinary experts who contributed as podcast guests, and the students for their active participation in the live podcast and the assessments.

Funding

This work was funded entirely by internal resources of the University of Freiburg with no external financial support, and we acknowledge the support by the Open Access Publication Fund of the University of Freiburg.

Data Availability

All relevant data are included in this article.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Live podcast episode 2—endodontontology (case-based teaching material) PDF of the case-based teaching slides. The material includes learning objectives, a trauma-related case (teeth 21/22), sequential clinical and radiographic images (periapical radiographs [PA] and cone-beam computed tomography [CBCT]), key reasoning steps for pain diagnostics and pulp vitality assessment, take-home messages, and an optional self-test multiple-choice question (MCQ). All radiographs and clinical photographs are shown in anonymized form with separate patient consent for publication.

[[PDF File, 624 KB - mededu_v12i1e77980_app1.pdf](#)]

Multimedia Appendix 2

Interdisciplinary clinical application test (ICAT).

[[PDF File, 82 KB - mededu_v12i1e77980_app2.pdf](#)]

Multimedia Appendix 3

Radiology knowledge test (RKT).

[[PDF File, 441 KB - mededu_v12i1e77980_app3.pdf](#)]

Multimedia Appendix 4

Student Evaluation Questionnaire (SEQ).

[[PDF File, 139 KB - mededu_v12i1e77980_app4.pdf](#)]

References

1. 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](https://doi.org/10.1097/ACM.0000000000004565)] [Medline: [34935729](#)]
2. Wang LKP, Paidisetti PS, Sathyaranarayanan S. Using podcasts to study efficiently and increase exposure to medical specialties. *Acad Med* 2023 Feb 1;98(2):157. [doi: [10.1097/ACM.0000000000005075](https://doi.org/10.1097/ACM.0000000000005075)] [Medline: [36696295](#)]
3. Cho D, Cosimini M, Espinoza J. Podcasting in medical education: a review of the literature. *Korean J Med Educ* 2017 Dec;29(4):229-239. [doi: [10.3946/kjme.2017.69](https://doi.org/10.3946/kjme.2017.69)] [Medline: [29207454](#)]
4. Cook DA, Triola MM. What is the role of e-learning? Looking past the hype. *Med Educ* 2014 Sep;48(9):930-937. [doi: [10.1111/medu.12484](https://doi.org/10.1111/medu.12484)] [Medline: [25113119](#)]
5. Condo M, McKay J, Dassinger J, Nolan B. Podcasting as an innovative tool for National Physical Therapy Examination preparation. *Med Sci Educ* 2023 Jun;33(3):631-633. [doi: [10.1007/s40670-023-01792-1](https://doi.org/10.1007/s40670-023-01792-1)] [Medline: [37501798](#)]
6. Surapaneni KM. Implications of using podcasts as handy tools for continuing professional development in health professions. *Med Sci Educ* 2023 Aug;33(4):1021. [doi: [10.1007/s40670-023-01830-y](https://doi.org/10.1007/s40670-023-01830-y)] [Medline: [37546187](#)]
7. Apel Z, Fagundes NCF, Sharmin N, et al. Social media in oral health education: a scoping review. *Eur J Dent Educ* 2025 Feb;29(1):50-63. [doi: [10.1111/eje.13053](https://doi.org/10.1111/eje.13053)] [Medline: [39462438](#)]
8. Kruse AB, Isailov-Schöchlin M, Giesler M, Ratka-Krüger P. Which digital learning strategies do undergraduate dentistry students favor? A questionnaire survey at a German university. *GMS J Med Educ* 2023;40(4):Doc49. [doi: [10.3205/zma001631](https://doi.org/10.3205/zma001631)] [Medline: [37560041](#)]
9. Shenson JA, Adams RC, Ahmed ST, Spickard A. Formation of a new entity to support effective use of technology in medical education: the student technology committee. *JMIR Med Educ* 2015 Sep 17;1(2):e9. [doi: [10.2196/mededu.4676](https://doi.org/10.2196/mededu.4676)] [Medline: [27731843](#)]
10. Annamma LM, Varma SR, Abuttayem H, et al. Current challenges in dental education—a scoping review. *BMC Med Educ* 2024 Dec 23;24(1):1523. [doi: [10.1186/s12909-024-06545-1](https://doi.org/10.1186/s12909-024-06545-1)] [Medline: [39716191](#)]
11. Prosser GM, Louca C, Radford DR. Potential educational and workforce strategies to meet the oral health challenges of an increasingly older population: a qualitative study. *BDJ Open* 2022 Mar 10;8(1):6. [doi: [10.1038/s41405-022-00098-5](https://doi.org/10.1038/s41405-022-00098-5)] [Medline: [35273157](#)]

12. Subramanian J, Anderson VR, Morgaine KC, Thomson WM. The importance of “student voice” in dental education. *Eur J Dent Educ* 2013 Feb;17(1):e136-e141. [doi: [10.1111/j.1600-0579.2012.00773.x](https://doi.org/10.1111/j.1600-0579.2012.00773.x)] [Medline: [23279401](https://pubmed.ncbi.nlm.nih.gov/23279401/)]
13. Goh AYS, Lim AD. Reconceptualising dentist learning—a commentary. *Eur J Dent Educ* 2022 Aug;26(3):639-642. [doi: [10.1111/eje.12741](https://doi.org/10.1111/eje.12741)] [Medline: [34954843](https://pubmed.ncbi.nlm.nih.gov/34954843/)]
14. Kaplan H, Verma D, Sargsyan Z. What traditional lectures can learn from podcasts. *J Grad Med Educ* 2020 Jun;12(3):250-253. [doi: [10.4300/JGME-D-19-00619.1](https://doi.org/10.4300/JGME-D-19-00619.1)] [Medline: [32595838](https://pubmed.ncbi.nlm.nih.gov/32595838/)]
15. Back DA, von Malotky J, Sostmann K, Hube R, Peters H, Hoff E. Superior gain in knowledge by podcasts versus text-based learning in teaching orthopedics: a randomized controlled trial. *J Surg Educ* 2017;74(1):154-160. [doi: [10.1016/j.jsurg.2016.07.008](https://doi.org/10.1016/j.jsurg.2016.07.008)] [Medline: [27651055](https://pubmed.ncbi.nlm.nih.gov/27651055/)]
16. Dahlstrom MF. Using narratives and storytelling to communicate science with nonexpert audiences. *Proc Natl Acad Sci U S A* 2014 Sep 16;111(Suppl 4):13614-13620. [doi: [10.1073/pnas.1320645111](https://doi.org/10.1073/pnas.1320645111)] [Medline: [25225368](https://pubmed.ncbi.nlm.nih.gov/25225368/)]
17. Kahlke R, Bates J, Nimmon L. When I say ... sociocultural learning theory. *Med Educ* 2019 Feb;53(2):117-118. [doi: [10.1111/medu.13626](https://doi.org/10.1111/medu.13626)] [Medline: [29943464](https://pubmed.ncbi.nlm.nih.gov/29943464/)]
18. Kamal S, Trivedi SP, Essien UR, Nematollahi S. Podcasting: a medium for amplifying racial justice discourse, reflection, and representation within graduate medical education. *J Grad Med Educ* 2021 Feb;13(1):29-32. [doi: [10.4300/JGME-D-20-00990.1](https://doi.org/10.4300/JGME-D-20-00990.1)] [Medline: [33680297](https://pubmed.ncbi.nlm.nih.gov/33680297/)]
19. Riddell J, Robins L, Brown A, Sherbino J, Lin M, Ilgen JS. Independent and interwoven: a qualitative exploration of residents' experiences with educational podcasts. *Acad Med* 2020 Jan;95(1):89-96. [doi: [10.1097/ACM.0000000000002984](https://doi.org/10.1097/ACM.0000000000002984)] [Medline: [31517682](https://pubmed.ncbi.nlm.nih.gov/31517682/)]
20. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. *MIS Q* 2003 Sep 1;27(3):425-478. [doi: [10.2307/30036540](https://doi.org/10.2307/30036540)]

Abbreviations

CG: control group

CR: clinical reasoning

DMFR: dentomaxillofacial radiology

ICAT: interdisciplinary clinical application test

IG: intervention group

LP: live podcast

RKT: radiology knowledge test

SEQ: Student Evaluation Questionnaire

Edited by P Kanzow; submitted 29.May.2025; peer-reviewed by AV Mahuli, DP Valenzuela; accepted 19.Nov.2025; published 05.Jan.2026.

Please cite as:

Hillebrecht AL, Fritzsche D, Voss T, Kruse A, Keßler A, Vach K, Altenburger MJ, Schmelzeisen R, Semper-Hogg W. *Live Podcasting as an Educational Intervention in Dentomaxillofacial Radiology: Controlled Cohort Study* *JMIR Med Educ* 2026;12:e77980

URL: <https://mededu.jmir.org/2026/1/e77980>

doi: [10.2196/77980](https://doi.org/10.2196/77980)

© Anna-Lena Hillebrecht, Daniel Fritzsche, Thamar Voss, Anne Kruse, Andreas Keßler, Kirstin Vach, Markus Jörg Altenburger, Rainer Schmelzeisen, Wiebke Semper-Hogg. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 5.Jan.2026. 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.

Original Paper

Development of Therapeutic Competencies in Health Care Students: Qualitative Focus Group Study Using 360-Degree Video and Virtual Reality Technology

Camilla Lauritzen¹, PhD; Charlotte Reedtz¹, PhD; Kjersti Bergum Kristensen¹, PhD; Vår Mathisen², PhD; Eva Therese Næss³, ClinPsy; Rigmor Furu⁴, MD; Hege Nermo⁵, PhD; Rita Jentoft², MD

¹Regional Center for Child and Youth Mental Health, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

²Department of Health and Care Sciences, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

³Department of Psychology, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

⁴Centre for Health Education Development, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

⁵Department of Dentistry, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

Corresponding Author:

Camilla Lauritzen, PhD

Regional Center for Child and Youth Mental Health

Faculty of Health Sciences

UiT The Arctic University of Norway

Postboks 6050 Stakkevollan

Tromsø, 9037

Norway

Phone: 47 77645871

Email: camilla.lauritzen@uit.no

Abstract

Background: Therapeutic competence is a critical skill for health care professionals, encompassing communication, interaction, and guidance in vulnerable situations. Virtual reality (VR) and 360-degree video technologies have emerged as innovative tools in health care education, offering immersive and interactive learning experiences. However, there is limited research on their effectiveness in developing therapeutic competencies among health care students.

Objective: This pilot study aimed to explore the feasibility, usability, and perceived educational value of a virtual learning resource using VR and 360-degree video to enhance therapeutic competence in health care students.

Methods: A virtual learning resource was developed, consisting of three modules: (1) a virtual home visit, (2) observation of therapeutic conversations using a 360-degree video, and (3) practice of therapeutic conversations in a simulated environment using VR. The resource was tested with students (n=12) from occupational therapy, psychology, and dentistry programs. Data were collected through focus group interviews after the students completed the modules. Thematic analysis was conducted to identify key themes related to the educational value and learning outcomes of the resource.

Results: The analysis revealed four key themes: (1) active exploration, where students engaged deeply with the material and contextualized theoretical knowledge; (2) observation, which provided practical insights into therapeutic conversations; (3) practice and reflection, which allowed students to refine their skills and build confidence; and (4) translation of theoretical knowledge into practical skills. Students reported that the resource was engaging, immersive, and effective in promoting learning compared to traditional teaching methods. Some students found the VR experience intense but valuable for skill development.

Conclusions: This pilot study demonstrates the feasibility and potential educational value of integrating VR and 360-degree video into health care education. The findings provide preliminary insights into the resource's ability to enhance therapeutic competence and student engagement. Future research should focus on larger, multi-institutional studies to validate these findings and assess the resource's impact on measurable learning outcomes.

(*JMIR Med Educ* 2026;12:e75776) doi:[10.2196/75776](https://doi.org/10.2196/75776)

KEYWORDS

health care education; virtual learning resources; virtual reality; therapeutic competence; immersive learning

Introduction

Background

Students pursuing health professional education are preparing for professions where they will interact with people in vulnerable life situations. Therefore, it is crucial that these students develop knowledge and skills in interaction, communication, and guidance during their education. In the context of health professions, this can be described as developing therapeutic competencies. Therapeutic competence can be defined as specific skills in conversations and interactions with patients and their families. These skills can be acquired through interactions between students and teachers [1].

Therapeutic competence encompasses a range of skills, including empathy, active listening, and the ability to adapt communication to the patient's needs and context. It also involves understanding the broader social and familial network of the patient, as well as the impact of the patient's condition on their family. In this paper, we explore how the development of therapeutic competence can be enhanced through an interactive virtual learning resource, using 360-degree video and virtual reality (VR) technology in teaching. VR and 360-degree videos have been increasingly recognized as effective tools in health care education. These technologies provide immersive and interactive learning experiences, allowing students to practice skills in a safe and controlled environment [2,3].

The learning resource is centered around supporting a family where one of the parents is experiencing mental illness. However, the methodology and therapeutic approach are quite generic and can be applied to several other severe conditions affecting a family, such as parental cancer, injuries, or substance use disorders.

Therapeutic competence also refers to the ability to assess and understand the patient's overall situation as part of a larger social and familial network. This holistic assessment includes the patient's roles and responsibilities, which may involve caring for children. An important aspect of this competence is understanding how the patient's illness affects their family, particularly their children, and adapting health care to the patient's life situation. Children whose parents experience significant illnesses, such as mental illness or substance use disorders, may require support, information, and follow-up. This perspective on therapeutic responsibilities expands the unit of care provision from a narrow focus on the patient to the wider family and caregiving system and is commonly referred to as a "family focus" in health services [4].

In Norway, health care personnel have a legal obligation to identify whether patients have children, following the revision of the Health Personnel Act in 2010 [5]. This legal provision requires health care personnel providing care to adult patients not only to clarify whether the patient has minor children but also to assess what information or support the family might need. The legislation applies to all illnesses that might impact everyday functioning and parenting, not just mental illnesses [5]. However, extensive national evaluations show that efforts

to change practices in hospitals so that all patients with children are assessed and offered support are progressing very slowly [6-9]. It has also been revealed that children are rarely included in conversations with health care personnel when parents are severely ill [10]. The challenges of successfully including children are related to obstacles both within families themselves and within the health care system [10,11]. Internationally, several studies document similar issues. Health care personnel across professions struggle to establish systematic routines that encompass the family situation of the patient [12-14].

Lack of time, competence, and competing tasks are among the reasons why many children do not receive the support and information they need when their parents are severely ill [11]. Part of the explanation may be that health care personnel perceive this work as additional rather than an integrated part of health care [9,11]. Additionally, health care personnel find it challenging to talk to children about difficult topics, such as substance abuse and mental illness in their parents [15].

Acquiring skills in communication, interaction, and guidance for families is thus central to the education of health care personnel. Health care professionals frequently encounter situations that demand therapeutic competence [16]. This requires an understanding of the situation and the ability to interact with the patient in an appropriate manner [17]. Therefore, it is essential that education focuses on developing therapeutic competencies, particularly skills in communication and interaction [16].

Approaches to Learning Therapeutic Competencies

Emphasizing the active role of students in knowledge development and applying educational approaches that encompass interactivity is an important starting point for developing therapeutic competence [1]. Chi and Wylie [18] describe 4 learning modes, namely passive, active, constructive, and interactive, which were tested in a 5-year project. The results showed that constructive and interactive activities, involving interaction, increased learning outcomes. The students who participated learned significantly more through interactive learning forms compared to traditional 1-way communication from teacher to student [2,3].

Observation can also be an important learning arena for students training to become health care professionals. Studies have shown that health care personnel lack competence in conducting professional conversations with children [4,5], and many are unsure of how to proceed in conversations with families [6,7]. However, by observing family and child conversations, students can gain knowledge about how such conversations can be conducted, although this is challenging to achieve in practice with families in highly vulnerable life situations. Learning through observation is often referred to as the apprenticeship model, which primarily involves the master first explaining or demonstrating how to perform various tasks, followed by beginners observing and trying the tasks themselves [8,9].

Interactive learning approaches and observation should be followed by a component of reflection to optimize the learning of new skills. Reflection is considered an essential component of the learning process, especially when it comes to acquiring

new skills in practice [10]. Reflection involves a conscious and systematic assessment of one's own experiences, actions, and results.

Another aspect to consider is that in the field of children of parents with severe illness, encounters with families where one or both parents face challenges due to illness may reveal children who are not receiving adequate care at home. In such situations, health care professionals may have a legal obligation to report these cases to agencies, such as social services. However, studies have shown that health care personnel largely underreport neglect [11]. To fulfill their reporting responsibilities, it is important that students training to become health care professionals are also trained to recognize risk factors and signs of neglect when interacting with families where serious illness is involved. Reflection promotes a deeper understanding, as it develops the ability to think analytically and solve problems creatively [12]. Additionally, reflection can help build confidence in the role of a health care professional, as one becomes more aware of one's own competence and potential for development.

Health care personnel report that it can be difficult to ask about sensitive topics, and they are often unsure of what to do with the information they receive [13]. A final aspect that should be emphasized when students are acquiring therapeutic competence and competence in family-focused treatment is the opportunity to practice. To develop skills in asking difficult questions and responding appropriately, it is necessary to obtain practice.

It is ethically challenging to allow students to practice on real patients as interns, especially when it involves complicated life challenges and illness profiles in patients. Practicing through role-play is common, and role-play can be a good way to train professionals for situations that will arise in real life. However, it can be challenging to get all students to participate in role-play in a way that provides learning outcomes because many may find it demanding to be in focus [14]. Furthermore, topics addressed through role-play can trigger negative emotions that may lead the learning process in an undesirable direction [15]. Practicing in simulated and virtual situations can thus be a good alternative. By using virtual learning resources, health science students can practice, among other things, establishing contact and building alliances in an arena where the students' skills have no consequences for the patient, as they might in an internship situation. Additionally, the virtual situation provides greater opportunities for repeated practice.

Study Objectives and Research Questions

This pilot study explores the feasibility, usability, and perceived educational value of a virtual learning resource designed to enhance therapeutic competence. The study aims to provide preliminary insights that can inform the development of larger-scale evaluations and more robust research designs in the future.

The research question in the study is "How can a virtual learning resource promote knowledge and skills in therapeutic competencies when interacting with patients who have minor children?"

Methods

Overview

The current project aimed to strengthen therapeutic competence in students training to become health professionals by developing virtual and interactive learning resources. The project received funding from the Norwegian Directorate of Higher Education and Skills to develop and test the learning resource. Alongside the development and testing of the resource, an approach inspired by action research was adapted, in the sense that researchers were closely involved in the development and testing process [19]. An important prerequisite was to facilitate systematic development and improvement based on students' experiences with using the learning resource.

While the study was inspired by action research principles, it does not fully align with the iterative cycles typically associated with action research. Instead, the study focused on piloting and refining the learning resource based on student feedback.

The Learning Resource: Design, Structure, and Learning Objectives

Learning Objectives

To strengthen therapeutic competence development in assessing and offering support to patients who have children, we developed a virtual learning resource for students training to become health care professionals. The objective of the learning resource is to give students and health care personnel insight into family-focused practice and to provide knowledge about how children are affected by their parents' illness. Furthermore, the learning objective is to teach students what kind of information they need to understand the parents' illness and its consequences. Additionally, the resource provides an example of how to conduct conversations with parents and children of different ages. The learning resource facilitates an arena for students to practice and reflect on their own performance and encompasses student-active learning forms, observational learning, reflection, and practice. The learning resource consists of 3 modules, which are described below.

Modules

Module 1

In this module, students go on a virtual home visit to an example family, the Hansen family's house. This is done as preparatory work through a flipped-classroom approach. Students follow a learning path where they gradually gain more insight into the family's risk and protective factors. The learning resource provides room for practice and reflection, with interactive tasks and exploration of the case (the family) that encourages participants to assess and develop their own practice. The virtual home visit is easily done using a computer and does not require special equipment. Students walk through the family's home and acquire knowledge about both the topic area of children as relatives and the example family, the Hansens. This is done by exploring various clickable interactive elements placed in different rooms in the home.

Module 2

In this module, students observe conversations with families using 360-degree video and VR headsets. VR refers to a technology that allows the user to interact in a computer-generated environment that mimics reality. VR for skill training has proven to be as effective in teaching as traditional exercises [3]. One of the unique strengths of using VR is that it can provide experiences that would otherwise not be available, such as observation in virtual rooms and practicing specific skills. In this module, the use of VR offers auditory and visual experiences that create a sense of presence in situations that would otherwise not be accessible and can thus enrich the learning process through realistic and engaging simulations.

Figure 1. Overview of the modules in the learning resource. VR: virtual reality.

Module 1	Module 2	Module 3
<ul style="list-style-type: none"> • Virtual home visit at the case family • Flipped classroom approach • Addressing the knowledge base for therapeutic competence in regard to patients who have minor children 	<ul style="list-style-type: none"> • Observation of conversations with the patient and his family • VR scenarios illustrating family focus in practice 	<ul style="list-style-type: none"> • Skills training in communication • Adapting information to different groups of patients or family members

Participants and Procedure in Testing the Learning Resource

The testing of the learning resource was conducted with 2 groups of students recruited from a faculty of health sciences. A total of 12 students participated, including 2 students each from the occupational therapy, psychology, and dentistry programs in each group. The students had limited knowledge of family-focused treatment before the testing. However, they had received some instruction in therapeutic competence and communication, mainly theoretical with elements of role-play. Module 1 was completed at home as preparation, while modules 2 and 3 were completed in a learning lab at the faculty. Based on the experiences from the first group, adjustments were made to the setup for test group 2.

The students used VR headsets for modules 2 and 3. The researchers were present during the entire testing period and took notes consecutively. After finishing the third module, 2 researchers interviewed the students in focus groups. The interviews were transcribed by a transcription service and quality assured by the researchers who conducted the interviews. Analyses of the initial interview revealed minor errors and weaknesses in the learning resource, which were then adjusted

Module 3

The final module opens opportunities for practicing conversations with the family and the children. With VR headsets on, students sit in the therapy room and interact with the family themselves. The 360-degree videos provide the experience of the family coming into the room to the student as a therapist. There is no dialogue in this module, but the family enters the room and looks directly at the student. It is then up to the student to engage in conversations with the family and the children of different ages. Specific tasks have been created for the module to help the students practice how to express themselves. The exercises for module 3 can be done alone or in groups of students. [Figure 1](#) provides an overview of the entire learning resource.

before test group 2 tried the resource. The results from test group 2 led to further fine-tuning of the learning resource, and a teacher guide was developed for pedagogical support. The learning resource and guide were then made available in Canvas Commons. Canvas Commons is an integrated content repository for Canvas that lets educators discover, share, and import course materials—like modules, assignments, and quizzes—across courses and institutions. The learning resource is published with open access in line with the principles of open education. We are currently in the process of translating the resource to English, making it freely available for use in international education.

Analyses

The text data in this study were analyzed thematically. Thematic analysis is based on selected themes by comparing information from all informants in a study [20]. Text data were systematically reviewed by the first author, and the content was sorted into categories by theme. The thematic analysis followed Braun and Clarke's 6-step framework, which includes familiarization with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report [21]. Codes were developed inductively based on recurring patterns in the focus group transcripts and

were iteratively refined by 2 researchers to ensure consistency and reliability.

The overarching thematic basis for the analysis has been the educational value and learning outcomes of the learning resource. Through the analysis of the interviews, it was identified how the learning resource contributed to learning by (1) exploration; (2) observation; (3) practicing, reflecting, and discussing; and (4) translation of theoretical knowledge into practical skills.

Ethical Considerations

The project was exempt from ethics review approval by the regional ethical committee, as it does not involve the collection of health information. Privacy considerations have been assessed and approved by SIKT (Norwegian Agency for Shared Services in Education and Research). All informants in the study have been presented with written information material and have signed informed consent forms. The participating students received a small compensation fee for taking part in the testing of the resource.

Results

Overview

In the analysis of the focus group interviews, several themes were identified that illustrate how a virtual learning resource can promote knowledge and skills in therapeutic competencies when interacting with patients who have minor children. Overall, the results showed that the learning resource engaged the students and contributed to an active exploration of the learning material, leading to a deeper understanding and better retention of information. Another overarching finding was that the learning resource was perceived by students as a more interactive and engaging learning experience than traditional teaching. Below, we delve into the 4 categories that emerged in the analysis. These were (1) learning through exploration, (2) learning through observation, (3) learning through practicing, reflection, and discussion, and (4) translating what has been learned into practical skills.

Learning Through Exploration

Students found that the virtual resources facilitated an active exploration process where they could engage directly with the material in an interactive way. Students described how the preparatory work in module 1 allowed them to explore various aspects of the family's situation, not just by passively reading, but by investigating and assessing the family situation and viewing the situation from different perspectives. The interactive exploration led to a deeper understanding of the topic and made it easier to remember what they had learned compared to traditional methods, for example, receiving a lecture with Microsoft PowerPoint on the subject matter. The dialogue below illustrates this:

Did it make it a bit easier to grasp the health legislation this way, compared to having something presented on a PowerPoint? You're nodding, yes.
[Interviewer]

Yes, that's a very good example, with the legislation... I hate when it's introduced on a slide. It's something I can't manage to learn when it comes on PowerPoint, whereas here I could almost remember the paragraph word by word afterward. [Student]

I found it very helpful when the legislation was very connected to the situations and the examples. They were tailored to the situations you had either gone through or were going to go through, so it was a bit like: "okay, that applies here, and that applies there." It made much more sense. Because when you're reading it in a textbook, it's like: "Yes, it probably applies there." But when you could see why, it was way better. [Student 2]

So, it was easier to remember it? [Interviewer]

Yes, really. [Student 2]

Speaking of what he said about retaining the information, because in our educational program, we constantly ask for clinical examples. Put it into context, give us something to attach it to, and don't just list an 80-page PowerPoint with legislation. We can't grasp that. But receiving something concrete, I think, can be used in many other contexts too, and it is so much easier to learn when you have something to attach it to. [Student 1]

In both focus groups, there was broad agreement that being able to explore the topic themselves, in this case through a virtual home visit, was perceived as promoting learning. It made abstract theory about the consequences of being a child of parents with various life challenges, as well as training in health legislation, more concrete and comprehensible.

Some students found it somewhat intense to work with the topic in this way. The quote below illustrates this:

Before I came here, I had an expectation that it would be interesting and educational. I'm generally interested in therapeutic competence, communication forms, and such things. But I didn't think it would be as all-encompassing as it actually was. When I did module 1 on the PC at home yesterday, I even found it more educational than a PowerPoint. So, now in VR mode, I almost felt my brain boiling, because there was so much info. I started thinking like: "Who should I look at? Should I look at her or him? Who am I most curious about in this family?" And that's a good thing, but yes, very intensive learning. So, it's a great resource, but I noticed I was a bit caught off guard. I didn't know it would be so many stimuli and impressions. [Student]

This finding shows that it can also be overwhelming to work with such topics in VR, and that teachers therefore must consider that various reactions may arise in the student group. Discomfort can also be a source of development and lead to students experiencing a high learning outcome. Being able to explore the case on their own and become familiar with the challenges the example family had was perceived as relevant and educational.

Another student expressed this as follows:

I think that often when we do case-based learning, the information that is important in the case is pointed out to you. But in this exploration, you get a lot of information to assess yourself. With being in such an interactive resource, you get the opportunity to choose for yourself what you want to focus your attention on. You practice much more on noticing the important things, compared to when you just get presented with something. When you have to look for it yourself, you also practice focusing your attention. So, I thought that was very cool. [Student]

In the conversation about exploring the topic area on their own using the interactive, virtual home visit in module 1, the students reported that it had a high educational value. The students said that it resembled reality. When students are in internships or start practicing as healthcare personnel after their education is finished, they don't necessarily know what the patients' or families' issues are. They have to use their developing therapeutic competence and knowledge to assess risks and problem areas to work on moving forward. The overarching finding in this category was that the learning resource was perceived as a very practice-oriented and effective way to learn these skills.

Learning Through Observation

The students reported that observing therapeutic conversations and communication between the therapist, patient, and the patient's family was very educational. The observations allowed them to see the acquired theory applied in practice. Students discussed how they observed different communication styles in conversations with children versus adults and how the therapist adapted their language and methods according to the client's age and level of understanding. This provided students with a practical understanding of how to tailor therapeutic techniques to different client groups.

Observing these conversations can be used as a starting point for skills training and give us the opportunity to practice therapeutic competence aimed at children, adolescents, adults, and older adults. It allows us to see what we need to do as therapists to bring out the best possible information about their life situation, and in what way we can use ourselves as tools to acquire information. [Student]

Another student expressed it as follows:

It was very educational because you get to see a real situation of how both the therapist and those coming to the therapist express themselves, and how their body language is. In my case, I learn a lot by viewing examples, like in an internship, by watching how my supervisor does it, and this was similar. [Student]

The overarching finding from this category showed that students experienced high learning outcomes from observing a conversation between a therapist and a family.

Learning Through Practicing, Reflection, and Discussion

VR offers a unique opportunity not only to observe but also to actively engage through simulated therapeutic sessions. This leads to reflection and discussion, which are crucial for learning and developing therapeutic competence. Participants appreciated the opportunity to discuss and reflect on their actions and decisions in the VR scenarios, as shown in the dialogue below:

How did you find reflecting along the way? Useful or not useful? [Interviewer]

I generally think just seeing that you have different ways of perceiving things, and that someone can make you aware of something you absolutely weren't aware of. We discovered that we disagreed on how we perceived the situation, and I think it's great to see different viewpoints. [Student 1]

Yes, and it triggers a much longer thought process too. [Student 2]

I thought it was very educational, even though I felt tired at the end. [Student 3]

I thought the questions were good. The questions allowed us to reflect on several possibilities and perspectives instead of giving us the answer immediately. [Student 4]

This excerpt from one of the focus group interviews highlights the value of questions that stimulate reflection. This helps them understand and internalize the learning material better.

Another important pedagogical value that emerged in the focus groups was the opportunities the VR resource provided for practice. The dialogue below illustrates this:

I noticed that for me to practice my therapeutic competence, I need a certain calm. Because I noticed that I didn't quite know where to start at first, because I was stressed by them [the family] staring at me. It takes practice, I think, because you have to be able to be in a certain situation several times to be able to do it well. So, I felt that I got into it the more I practiced. [Student 1]

To get into a kind of flow. I started to think about my body language eventually. After a while, I started to think: "How am I sitting now, actually? Is it okay to sit like this?" But at first, I couldn't even think about what to say. [Student 2]

The overarching result in this category is that students experienced good learning outcomes from trying out, practicing, reflecting, and discussing with each other using this learning resource.

Translating What Has Been Learned Into Practical Skills

Finally, the students discussed how they can apply what they have learned through VR in their professional practice. This includes applying knowledge and skills in real therapeutic situations, such as handling sensitive topics with clients or adapting communication styles based on the client's needs and

situation. Students reflected on how they can use insights from the learning resource to improve their practice, especially in how they interact with clients and handle complex and sensitive issues. The dialogue below illustrates this:

So, what do you think about conducting such a conversation now? What would it be like? [Interviewer]

One of the things I'm a bit nervous about is exactly that. But it feels a bit easier now. I feel like I know more. [Student]

Students also reported that they have become more aware of what the term therapeutic competence entails. They also mentioned that they had become more attentive to the whole family and the importance of involving relatives. The dialogue below illustrates this:

I'm also a bit curious about what you now consider to be therapeutic competence. What do you think now? Is it clear to you what it is? [Interviewer]

It's a whole bag of things, with listening and hearing, tips, tricks, what works, what doesn't, reflecting, getting them to become aware of things, and not just like: "Pull yourself together, get well!" [Student 1]

I became much more aware that maybe we should try to involve families more often. [Student 2]

The overarching result in this category is that, in terms of translating theoretical knowledge into actual therapeutic competence, the students had positive learning outcomes by using the learning resource.

Discussion

Principal Findings

This pilot study explored the feasibility, usability, and perceived educational value of a virtual learning resource designed to enhance therapeutic competence in healthcare students. The findings suggest that the resource has the potential to promote active exploration, observation, and practice in a safe and controlled environment. Students reported that the resource was engaging and effective in promoting learning compared to traditional teaching methods. The study's findings align with existing literature on the benefits of interactive and immersive learning approaches in healthcare education [2,18].

The results also showed that the learning resource provided students with the opportunity to observe and practice therapeutic conversations in a safe environment. This aligns with the need to develop skills in adapting communication and interaction to different patient groups. These skills are crucial for responsible professional practice. VR technology offers a unique opportunity to learn through observing simulated, yet realistic, therapeutic conversations. VR technology can simulate complex interactions with patients and their families, providing students with some experience, yet avoiding the ethical concerns that may arise from practicing directly on patients. This is particularly relevant in education, where direct practice with real patients can be challenging or difficult to achieve, such as in the early stages of educating health professionals. Compared to role-play, where

students themselves play the different roles, using films of relevant scenarios where students practice being a therapist who observes, listens, and interacts through VR technology appears more direct and thus professionally relevant than inventing thoughts, feelings, and behaviors in people they play themselves. Furthermore, the students developed new concepts and understanding through reflection and discussion of what they had observed in the scenarios. They could relate legislation and therapeutic techniques to practical situations, showing that they moved from concrete experience to abstract conceptualization [22]. Translating what they had observed into concrete skills seemed, therefore, easier using this learning resource.

The learning resource is designed for students to reflect on their observations, often referred to as reflective observation [22]. Using reflection as a key component in the learning process, the opportunity to reflect on their observations was provided, and the students appreciated this option. The students found that the questions in the learning resource stimulated deeper thinking, reflection, and discussion, which is crucial for developing a deeper understanding and ability to adapt to new challenges. By discussing how the VR experience allowed them to see theory applied in practice, they felt they gained a deeper understanding of the material. Reflection on what they observed, such as communication styles and the therapist's adaptation to the client's age, thus helped strengthen their learning.

The learning resource gave the students an opportunity to practice and refine their interaction skills in a controlled environment before applying them in real situations. Practice is essential for developing skills. Students reported that they experienced increased confidence and flow in communication after repeated exercises in the VR environment. This underscores the importance of practice for achieving mastery and confidence in therapeutic practice. Trying out new skills in a safe environment can be essential for active experimentation. The students reported that practice in VR helped them achieve flow and confidence in therapeutic conversations, which is a critical part of the learning process. Practicing in VR was perceived as activating and engaging, and the students found that the resource provided a deeper understanding as well as promoted the development of skills.

Translating theoretical knowledge into practical skills can be difficult to achieve with traditional learning methods. Students reported that the VR experience provided a more vivid and realistic understanding of the subject matter, contributing to more effective learning. Students described the VR experience as particularly valuable because it allowed them to "step into" scenarios relevant to their field of study. This feeling of being part of the scenario contributed to a more authentic and engaging learning experience. The students felt that visualization contributed to a better understanding of complex subject matter. The learning resource created a unique opportunity to explore, observe, and interact in complex situations in a way that text or images cannot. In light of Kolb's experiential learning cycle [22], one can say that the learning resource provides a concrete experience that is essential to start the learning cycle. Students could see body language, hear tone of voice, and observe interactions in a controlled but realistic environment. This helped them understand complex subject matter and integrate theoretical

and practical aspects of the therapeutic competence they will need in practice. Through exploration, observation, reflection, and discussion, the learning resource thus contributes to the acquisition of experience, laying the foundation for the development of therapeutic competence.

VR-based practice offers significant advantages for developing foundational therapeutic communication skills, such as active listening, empathy, and adapting communication to different client groups. The immersive nature of VR allows students to observe and practice interactions in realistic scenarios, which can enhance their confidence and readiness for real-world practice [23]. However, certain aspects of authentic therapeutic interaction, such as emotional attunement, relational pacing, and the influence of cultural and familial contexts, may be more challenging to replicate in a virtual environment. These elements often require nuanced, real-time adjustments that are best developed through direct interaction with patients and guided supervision. To address these limitations, the learning resource was complemented with structured reflection activities that allowed students to explore the subtleties of therapeutic relationships in greater depth.

Therapeutic communication is a complex skill that develops progressively throughout a health care professional's education and career [1,16]. For students in the early stages of their training, foundational skills, such as active listening, empathy, and basic conversational techniques, are critical [18,22]. The virtual learning resource aligns with these developmental needs by providing a safe and controlled environment for students to practice these foundational skills. Unlike residents or more experienced practitioners, students may have limited exposure to real-life therapeutic interactions, making simulated environments particularly valuable for bridging the gap between theoretical knowledge and practical application. Future studies could explore how the resource might be adapted for learners at different stages of their professional development, including residents and practicing clinicians, to address more advanced aspects of therapeutic communication.

Equity and fairness are critical considerations in the implementation of VR-based learning environments. Learners come from diverse cultural, linguistic, and socioeconomic backgrounds, which can influence their communication styles and access to technology. Similarly, patients and families bring unique cultural and contextual factors that shape their communication needs and expectations. To support equitable outcomes, the learning resource could be adapted to include scenarios that reflect a wide range of cultural and familial contexts. Additionally, structured reflection activities and supervisory guidance could help learners develop cultural competence and inclusive communication practices. Future iterations of the resource should also consider accessibility features, such as subtitles or alternative formats, to ensure that all learners can fully engage with the material.

Practical Implications

Although VR can have significant educational benefits, several challenges must be addressed to maximize educational value and learning outcomes. It is important to consider and assess the ethical aspects of using VR in education, because simulations

of sensitive topics can have unforeseen consequences. Teachers using such learning resources should be prepared for emotional reactions and have a plan for how to handle such reactions. As with all teaching on sensitive topics that are perceived as challenging for students, it requires educators to help students navigate these challenges by providing support and guidance throughout the learning process and through supervision.

User-friendliness and technical support are important prerequisites when applying virtual resources to teaching situations [23]. Participants in the focus groups reported the need for technical support and guidance when using VR, emphasizing the importance of accessible and effective technical assistance. Another challenge is the cost and availability of VR technology. Although the prices of VR equipment have fallen in recent years, costs can still be a barrier for many educational institutions. Furthermore, effective use of VR requires appropriate software and updates, which can also incur significant costs. It also requires some logistics to ensure that equipment is charged and updated with new software so that it can be easily used.

It is also important to consider how VR can be integrated and systematically implemented in educational programs to maximize learning outcomes. To ensure that VR technology becomes an integrated part of educational programs, educational institutions must consider both infrastructural and pedagogical adjustments. This requires anchoring with management, investment in necessary technology, organization of training, use, and guidance at the institution. By using VR technology, however, one can achieve more targeted competence development. In many health professions, VR scenarios can be tailored to meet the need for developing specific clinical and therapeutic skills.

Limitations

This pilot study involved a small sample size (n=12) and a single-case context, which limits the generalizability of the findings. The study was designed to provide preliminary insights into the feasibility and perceived educational value of the virtual learning resource, rather than to draw definitive conclusions about its effectiveness. Future research should include larger, more diverse student groups and multi-institutional studies to validate the findings and assess the resource's impact on measurable learning outcomes. Additionally, comparative studies with control groups using traditional teaching methods would provide a more robust evaluation of the resource's relative advantages and limitations.

While the study was inspired by action research principles, it did not fully adhere to the iterative cycles typically associated with action research. Instead, the study focused on piloting and refining the learning resource based on student feedback. Future research could adopt a more comprehensive action research approach, incorporating multiple iterations and cycles of feedback to further enhance the resource and its implementation.

Conclusion

This pilot study demonstrates the feasibility and potential educational value of integrating VR and 360-degree video into health care education. The participating students experienced

that the virtual learning resource promoted the learning of therapeutic competence. Interactive learning, practice, and reflection were important elements that contributed to the students' development of skills necessary to meet the demands of health professions. The findings provide preliminary insights

into the resource's ability to enhance therapeutic competence and student engagement. Future research should focus on larger, multi-institutional studies to validate these findings and assess the resource's impact on measurable learning outcomes.

Acknowledgments

The authors would like to acknowledge The Resource Center for Teaching, Learning, and Technology (RESULT) at UiT - The Arctic University of Norway. Furthermore, we would like to acknowledge the students who participated in testing and developing the learning resource. We used a locally adapted large language model for translation and language editing purposes. The model used is named chat.uit.no and is based on OpenAI ChatGPT-4.

Funding

The project received funding from the Norwegian Directorate for Higher Education and Skills: 2019/10223.

Data Availability

The datasets generated or analyzed during this study are not publicly available because they consist of qualitative interview data that cannot be adequately anonymized, and written consent to store the data in an open repository was not obtained prior to the study.

Authors' Contributions

Conceptualization: CL Data curation: CL, ETN, RF, RJ Formal analysis: CL Funding acquisition: CL, RF, RJ Project administration: CL, RF, RJ Writing – original draft: CL Writing – review & editing: CL, CR, KBK, VM, ETN, RF, HN, RJ

Conflicts of Interest

None declared.

References

1. Jentoft R, Mathisen V., & Mathisen, V. In: Arntzen C, Gramstad A, Jentoft R, editors. *Ergoterapi - mangfold og muligheter (Occupational Therapy: Diversity and Opportunities)*. Oslo: Orcana Academic Publisher; 2015.
2. Blair C, Walsh C, Best P. Immersive 360° videos in health and social care education: a scoping review. *BMC Med Educ* 2021;21(1):590 [[FREE Full text](#)] [doi: [10.1186/s12909-021-03013-y](https://doi.org/10.1186/s12909-021-03013-y)] [Medline: [34819063](#)]
3. Kaplan AD, Crut J, Endsley M, Beers SM, Sawyer BD, Hancock PA. The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: a meta-analysis. *Hum Factors* 2021;63(4):706-726. [doi: [10.1177/0018720820904229](https://doi.org/10.1177/0018720820904229)] [Medline: [32091937](#)]
4. Foster K, O'Brien L, Korhonen T. Developing resilient children and families when parents have mental illness: a family-focused approach. *Int J Ment Health Nurs* 2012;21(1):3-11. [doi: [10.1111/j.1447-0349.2011.00754.x](https://doi.org/10.1111/j.1447-0349.2011.00754.x)] [Medline: [21692961](#)]
5. Parr J. Om Lovdata. *Lov og Rett* 1989;28(7):330-331 [[FREE Full text](#)] [doi: [10.18261/issn1504-3061-1989-07-06](https://doi.org/10.18261/issn1504-3061-1989-07-06)]
6. Lauritzen C, Reedtz C, Rognmo K, Nilsen MA, Walstad A. Identification of and support for children of mentally ill parents: a 5 year follow-up study of adult mental health services. *Front Psychiatry* 2018;9:507 [[FREE Full text](#)] [doi: [10.3389/fpsyg.2018.00507](https://doi.org/10.3389/fpsyg.2018.00507)] [Medline: [30386268](#)]
7. Lauritzen C, Kolmannskog AB, Iversen AC. Family assessment conversations as a tool to support families affected by parental mental illness: a retrospective review of electronic patient journals. *Int J Ment Health Syst* 2018;12:18 [[FREE Full text](#)] [doi: [10.1186/s13033-018-0199-x](https://doi.org/10.1186/s13033-018-0199-x)] [Medline: [29760769](#)]
8. Reedtz C, Jensaas E, Storjord T, Kristensen KB, Lauritzen C. Identification of children of mentally ill patients and provision of support according to the norwegian health legislation: a 11-year review. *Front Psychiatry* 2021;12:815526 [[FREE Full text](#)] [doi: [10.3389/fpsyg.2021.815526](https://doi.org/10.3389/fpsyg.2021.815526)] [Medline: [35095621](#)]
9. Ruud T, Birkeland B, Faugli A, et al. Barn som pårørende: resultater fra en multisenterstudie [Children of ill parents: results from a multi-center study]. Fafo. 2015. URL: <https://www.fafo.no/publikasjoner/eksterne-rapporter/barn-som-parrende> [accessed 2025-12-29]
10. Kristensen KB, Lauritzen C, Reedtz C. Support for children of parents with mental illness: an analysis of patients' health records. *Front Psychiatry* 2022;13:778236 [[FREE Full text](#)] [doi: [10.3389/fpsyg.2022.778236](https://doi.org/10.3389/fpsyg.2022.778236)] [Medline: [35356710](#)]

11. Lauritzen C, Reedtz C, Van Doesum K, Martinussen M. Factors that may facilitate or hinder a family-focus in the treatment of parents with a mental illness. *J Child Fam Stud* 2015;24(4):864-871 [FREE Full text] [doi: [10.1007/s10826-013-9895-y](https://doi.org/10.1007/s10826-013-9895-y)] [Medline: [25814823](https://pubmed.ncbi.nlm.nih.gov/25814823/)]
12. Laletas S, Reupert A, Goodyear M, Morgan B. Pathways of care: targeting the early childhood sector for early intervention. *Adv Ment Health* 2015;13(2):139-152. [doi: [10.1080/18387357.2015.1066293](https://doi.org/10.1080/18387357.2015.1066293)]
13. Laletas S, Goodyear M, Reupert A. Parental mental illness: cross-sectional analysis of family focused practice within the early childhood sector. *J Child Fam Stud* 2017;27(5):1650-1660. [doi: [10.1007/s10826-017-0992-1](https://doi.org/10.1007/s10826-017-0992-1)]
14. Grant A, Reupert A, Maybery D, Goodyear M. Predictors and enablers of mental health nurses' family-focused practice. *Int J Ment Health Nurs* 2019;28(1):140-151. [doi: [10.1111/inm.12503](https://doi.org/10.1111/inm.12503)] [Medline: [29949231](https://pubmed.ncbi.nlm.nih.gov/29949231/)]
15. Lauritzen C, Reedtz C. Support for children of service users in Norway. *Mental Health Practice* 2013;16(10):12-18. [doi: [10.7748/mhp2013.07.16.10.12.e875](https://doi.org/10.7748/mhp2013.07.16.10.12.e875)]
16. Mota P, Selby K, Gouveia A, Tzartzas K, Staeger P, Marion-Veyron R, et al. Difficult patient-doctor encounters in a Swiss university outpatient clinic: cross-sectional study. *BMJ Open* 2019;9(1):e025569 [FREE Full text] [doi: [10.1136/bmjopen-2018-025569](https://doi.org/10.1136/bmjopen-2018-025569)] [Medline: [30782753](https://pubmed.ncbi.nlm.nih.gov/30782753/)]
17. Weingarten M, Guttman N, Abramovitch H, Margalit RS, Roter D, Ziv A, et al. An anatomy of conflicts in primary care encounters: a multi-method study. *Fam Pract* 2010;27(1):93-100. [doi: [10.1093/fampra/cmp082](https://doi.org/10.1093/fampra/cmp082)] [Medline: [19948564](https://pubmed.ncbi.nlm.nih.gov/19948564/)]
18. Chi M, Wylie R. The ICAP framework: linking cognitive engagement to active learning outcomes. *Educ Psychol* 2014;49(4):219-243 [FREE Full text] [doi: [10.1080/00461520.2014.965823](https://doi.org/10.1080/00461520.2014.965823)]
19. Ulvik M, Riese H, Roness D. *Å Forske På Egen Praksis. Aksjonsforskning Og Andre Tilnærmingar Til Profesjonell Utvikling I Utdanningsfeltet* [Researching One's Own Practice: Action Research and Other Approaches to Professional Development in the Field of Education]. Bergen: Fagbokforlaget; 2016.
20. Thagaard T. Systematikk og innlevelse. In: *En Innføring I Kvalitative Metoder* [Systematics and Empathy: an Introduction to Qualitative Methods]. Bergen: Fagbokforlaget; 2018.
21. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2008;3(2):77-101 [FREE Full text] [doi: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)]
22. Kolb DA. *Experiential Learning: Experience as the Source of Learning and Development*. New Jersey: Pearson Education; 2015.
23. Lie SS, Helle N, Sletteland NV, Vikman MD, Bonsaksen T. Implementation of virtual reality in health professions education: scoping review. *JMIR Med Educ* 2023;9:e41589 [FREE Full text] [doi: [10.2196/41589](https://doi.org/10.2196/41589)] [Medline: [36692934](https://pubmed.ncbi.nlm.nih.gov/36692934/)]

Abbreviations

SIKT: Norwegian Agency for Shared Services in Education and Research

VR: virtual reality

Edited by AH Sapci, MD; submitted 10.Apr.2025; peer-reviewed by A Ugbor, R Ayub; comments to author 15.Jun.2025; revised version received 20.Nov.2025; accepted 14.Dec.2025; published 02.Feb.2026.

Please cite as:

Lauritzen C, Reedtz C, Kristensen KB, Mathisen V, Næss ET, Furu R, Nermo H, Jentoft R
Development of Therapeutic Competencies in Health Care Students: Qualitative Focus Group Study Using 360-Degree Video and Virtual Reality Technology

JMIR Med Educ 2026;12:e75776

URL: <https://mededu.jmir.org/2026/1/e75776>

doi: [10.2196/75776](https://doi.org/10.2196/75776)

PMID:

©Camilla Lauritzen, Charlotte Reedtz, Kjersti Bergum Kristensen, Vår Mathisen, Eva Therese Næss, Rigmor Furu, Hege Nermo, Rita Jentoft. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 02.Feb.2026. 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.

Implementation and Evaluation of a Cancer Immunotherapy Elective for Medical Students: Mixed Methods Descriptive Study

Mark Raynor¹, MD; Rivers Hock¹, MD; Brandon Godinich¹, MBA, MHA; Satish Maharaj^{2,3}, MD; Houriya Ayoubieh¹, MD; Cynthia Perry¹, PhD; Jessica Chacon¹, PhD

¹Department of Medical Education, Foster School of Medicine, Texas Tech University Health Sciences Center El Paso, 130 Rick Francis, El Paso, TX, United States

²Department of Hematology/Oncology, University of South Florida, Tampa, FL, United States

³Moffitt Cancer Center, Tampa, FL, United States

Corresponding Author:

Jessica Chacon, PhD

Department of Medical Education, Foster School of Medicine, Texas Tech University Health Sciences Center El Paso, 130 Rick Francis, El Paso, TX, United States

Abstract

Background: Cancer immunotherapy represents a transformative advancement in oncology, offering new avenues for treating malignancies by harnessing the immune system. Despite its growing clinical relevance, immunotherapy remains underrepresented in undergraduate medical education, particularly in curricula integrating foundational immunology with clinical application. To address this gap, we developed and implemented a fully online elective for fourth-year medical students focused on core immunology concepts, immunotherapy mechanisms, Food and Drug Administration–approved treatments, immune-related adverse events, and patient-centered therapeutic decision-making.

Objective: This study aimed to evaluate the effectiveness of an asynchronous-synchronous online cancer immunotherapy elective in improving medical student knowledge, engagement, and critical-thinking skills. We hypothesized that participation in the elective would be associated with perceived improvements in knowledge and clinical preparedness and inform future strategies for integrating cancer immunotherapy into medical curricula.

Methods: We conducted a mixed methods study with fourth-year medical students enrolled in a 2-week elective at a US medical school. The curriculum included a self-paced foundational module, online discussion board, and a capstone oral presentation requiring students to propose a novel immunotherapy approach. Participants completed pre- and postcourse quizzes assessing immunotherapy knowledge and an anonymous postcourse Likert-scale survey. Quantitative data were summarized descriptively, and Likert responses were reported using medians and IQRs. Due to the small sample size, 2-tailed unpaired *t* tests comparing pre- and postcourse quiz averages were underpowered to detect statistically significant differences. Qualitative data were analyzed using inductive thematic analysis with investigator triangulation.

Results: A total of 35 students completed the elective, and 20 submitted the postcourse survey (response rate: 57%). Across all Likert-scale items, students reported a median response of 5 (Strongly Agree), with IQR values ranging from 0 to 1, indicating uniformly positive perceptions and minimal variability in their evaluation of the course. Descriptively, average postcourse quiz scores were higher than precourse scores, suggesting improved conceptual understanding. Qualitative thematic analysis revealed three major themes: (1) increased confidence engaging with complex immunotherapy mechanisms, (2) appreciation for the flexibility and interactivity afforded by the hybrid asynchronous-synchronous model, and (3) enhanced understanding of the real-world clinical application of immunotherapy across interdisciplinary settings.

Conclusions: Descriptive quantitative and qualitative findings suggest that a targeted online cancer immunotherapy elective may enhance learners' perceived knowledge and critical-thinking capacity related to emerging cancer therapies. The course's hybrid structure offered flexibility, accessibility, and potential scalability. As immunotherapy continues to expand in clinical practice, this model provides a promising framework for integration into medical curricula. Future work should include larger cohorts and longitudinal follow-up into residency to more rigorously assess educational impact.

(*JMIR Med Educ* 2026;12:e71628) doi:10.2196/71628

KEYWORDS

medical education; immunotherapy; cancer; undergraduate medical education; cancer immunotherapy

Introduction

Immunotherapy, heralded as a beacon of hope in the oncological landscape, stands at the forefront of medical innovation, offering novel avenues for combating cancer [1-3]. As immune checkpoint blockade, cellular therapies, and cancer vaccinology assume increasing prominence, proficiency in cancer immunotherapy is becoming an essential competency for future health care practitioners [1,2,4,5]. Yet, despite its transformative clinical impact, formal education on cancer immunotherapy remains fragmented or absent in most undergraduate medical curricula. Prior studies have highlighted that health care trainees often feel underprepared to discuss or apply immunotherapy principles in clinical settings, reflecting a broader disconnect between rapid scientific advances and medical education [6,7].

Existing educational work in this area has largely centered on postgraduate oncology programs, continuing medical education, or highly specialized workshops. There remains a critical gap in evidence-based, integrative, and educational models that can effectively prepare medical students, before residency, for the clinical realities of modern cancer care. This lack of structured exposure creates downstream challenges: learners enter rotations or residency without foundational immunotherapy literacy, clinicians report variability in trainee preparedness, and the rapidly evolving therapeutic landscape continues to outpace traditional curricula.

To address this gap, we developed the *Basics and Implementation of Cancer Immunotherapy* elective at the Texas Tech University Health Sciences Center (TTUHSC) El Paso Paul L. Foster School of Medicine. This fully online elective was intentionally designed to provide accessible foundational knowledge while fostering higher-order thinking, clinical reasoning, and innovation in the context of rapidly evolving immunotherapy treatments.

The goal of the course was to deliver content and create a sustainable, adaptable, and evaluable model for broader implementation. Therefore, we grounded the curriculum in 2 complementary frameworks: RE-AIM (Reach, Effectiveness, Adoption, Implementation, and Maintenance) and Kirkpatrick's model. RE-AIM is an implementation science framework typically used in public health and program dissemination research [8,9]. It guided the structural design of the elective by prompting us to consider who the elective could reach, how well it might function across learner groups, what resources and faculty expertise were required for adoption, how consistently it could be delivered in an online environment, and how it could be maintained or scaled over time.

In contrast, Kirkpatrick's model, which focuses on Reaction, Learning, Behavior, and Results, provided a pedagogical framework for evaluating educational impact at the learner level. This dual framework approach allowed us to simultaneously examine (1) learner satisfaction, knowledge, and perceived skill development, and (2) broader programmatic considerations such as feasibility, reproducibility, and sustainability within an already crowded medical curriculum [10-14]. By integrating these 2 models, we aimed to ensure both a high-quality

educational experience and a design that could be reliably implemented or adapted by other institutions.

This study aimed to evaluate the effectiveness of the elective in enhancing student knowledge and critical thinking using a mixed methods approach. Specifically, we sought to answer the research question: Does participation in this elective improve students' knowledge, perceptions of course relevance, and critical thinking in cancer immunotherapy? We hypothesized that students completing the elective would demonstrate measurable knowledge gains, report a positive educational experience, and show enhanced critical thinking related to immunotherapy. Findings from this study can guide future curricular efforts aimed at integrating contemporary immunotherapy concepts into undergraduate medical education in a structured and measurable way.

Methods

Rationale for Elective Development

Medical students at Foster School of Medicine receive limited exposure to cancer immunotherapy during the preclerkship phase because the integrated curriculum and density of required content constrain opportunities for in-depth coverage [15-17]. To address this gap, faculty designed a fourth-year elective grounded in a spiral learning model that builds on students' foundational immunology knowledge from the first 2 preclerkship years and reinforces concepts introduced during their clinical rotations.

Curricular Design

We selected a virtual elective to match student preferences for greater fourth-year flexibility and to leverage the institution's expanded capacity for high-quality online learning primarily developed during the COVID-19 era [18-21]. The elective directly addressed institutional needs identified through annual curriculum review and supported the School of Medicine's priority to prepare graduates for rapidly evolving biomedical innovations, such as immunotherapy. Within the Department of Medical Education, the elective advanced key curricular priorities: (1) integrating foundational basic sciences, such as immunology, with clinical reasoning; (2) broadening elective options that promote scholarly inquiry; and (3) expanding technology-enhanced and competency-based instruction.

Faculty with expertise in cancer immunotherapy, biochemistry and cell biology, clinical genetics, oncology, educational design, and online pedagogy led curriculum development. The team followed a staged design process: we mapped core immunotherapy competencies, identified gaps across the preclinical and clerkship curriculum, and iterated content through consultation with our clinical oncology faculty. Medical students actively contributed to early course development. A small group of students co-developed the immunology primer by drawing on their research experience and peer and faculty feedback to ensure clarity, relevance, and appropriate depth. These students also conducted an informal needs assessment through conversations with classmates, confirming that many trainees felt underprepared to discuss contemporary cancer immunotherapies during clinical rotations.

The faculty built the final course using interactive synchronous and asynchronous components, including the student-developed primer, structured discussion boards, and a culminating student presentation. These elements aimed to deepen engagement with the scientific literature and support inquiry–driven learning [18–20]. Constructivist learning theory, interdisciplinary integration [22], and reflective practice guided the instructional approach. Expert consensus, national immunotherapy competency frameworks, and student-identified needs informed topic selection. Basic science and clinical faculty peer-reviewed all content before launch to ensure scientific rigor, clinical relevance, and instructional quality.

Pedagogical Framework, Principles, and Competencies and Standards

Developing a fourth-year elective for medical students on cancer immunotherapy required deliberate alignment with pedagogical frameworks, instructional principles, established competencies, and educational standards to meet the learning objectives. Basic science faculty (immunology, biochemistry, and cell biology) and clinical faculty (internal medicine, oncology, and clinical genetics) jointly oversaw the course.

Constructivist and Experiential Foundations

The design of the elective drew on constructivist and experiential learning theories to promote knowledge construction through active engagement, inquiry, and reflective practice. We incorporated Kirkpatrick’s Four-Level Model and the RE-AIM framework to guide course design and evaluation. These frameworks ensured that learning activities targeted measurable outcomes, real-world application, and iterative improvement based on faculty and student feedback. The course combined asynchronous modules, faculty-led synchronous discussions, and a final oral presentation to support both individual learning and collaborative inquiry.

Kirkpatrick’s Four-Level Model

We applied the Kirkpatrick’s Model [10] to evaluate student learning and course impact.

- Level 1. Reaction: We collected postcourse survey feedback to assess engagement and satisfaction with course structure and content.
- Level 2. Learning: We measured knowledge acquisition using asynchronous pre- and postquizzes focused on immunology basics and the interplay between the immune system and the tumor microenvironment.
- Level 3. Behavior: We evaluated critical thinking by assigning discussion prompts requiring students to propose novel therapies and apply theoretical concepts in innovative ways.
- Level 4. Results: We examined how effectively students identified and proposed solutions to gaps in current immunotherapy practices during their oral presentations.

RE-AIM Framework

In parallel, we used the RE-AIM framework to evaluate implementation and scalability.

- *Reach*: We monitored enrollment numbers and demographic diversity.

- *Effectiveness*: We analyzed improvements in quiz scores and qualitative feedback on learning outcomes.
- *Adoption*: We noted the course’s integration into the fourth-year elective catalog and student uptake across multiple intended specialties.
- *Implementation*: We assessed delivery fidelity, faculty coordination, and alignment of synchronous and asynchronous components.
- *Maintenance*: We planned annual updates to incorporate emerging immunotherapy advances and maintain curricular relevance.

This dual framework approach enabled a comprehensive evaluation of learner experience and overall educational impact.

Pedagogical Framework

We provided a pedagogical framework for evaluating educational impact at the learner level:

- *Constructivist approach* [23]: Active learning was emphasized where students constructed their understanding of cancer immunotherapy through exploration, inquiry, and problem-solving.
- *Interdisciplinary integration* [24]: Principles from basic sciences, such as immunology, cancer biology, pharmacology, and molecular biology, were integrated to provide a holistic understanding of cancer immunotherapy.
- *Reflective practice* [25,26]: Students were encouraged to reflect on their learning, clinical experiences, and possible immune adverse events. Students were additionally challenged to consider their patient population, particularly those underrepresented in medicine when considering immunotherapy for cancer treatment.

Principles

Principles were integrated as follows:

- *Evidence-based* [27]: All content and teaching methods were grounded in current scientific evidence and clinical practice guidelines. A consultation with experts in the cancer immunotherapy field at the basic science and clinical settings was conducted.
- *Patient-centered* [28,29]: The importance of patient perspectives and personalized treatment approaches in cancer care was emphasized.
- *Ethical awareness* [30]: Ethical dilemmas, such as access to treatment, clinical trial participation, and the implications and importance of genetic testing in cancer immunotherapy, were discussed.

Competencies and Standards [31–34]

We developed the competencies and standards to meet the learning objectives:

- *Knowledge base*: Ensure that students understand the fundamental principles of cancer immunology, including the role of immune checkpoints, innate and adaptive immunity, cytokines, and monoclonal antibodies in cancer treatment.
- *Clinical skills*: Ensure that students develop a foundational understanding of how clinicians evaluate candidates for

- immunotherapy, monitor treatment responses, and identify common immune-related adverse events.
- *Critical thinking:* Oversee student ability to analyze current immunotherapy options and formulate evidence-based innovative treatment.
- *Communication skills:* Provide feedback on student ability to effectively communicate with faculty and peers regarding basic science concepts, cancer immunotherapy treatment, and potential side effects.
- *Professionalism:* Provide student feedback on demonstrating empathy, cultural competence, and ethical behavior in all aspects of immunotherapy care and professional interactions.

Elective Structure

We developed the elective structure for enhancing student knowledge and critical thinking:

- *Pre-elective quiz:* We developed the pre-elective quiz to measure students' baseline understanding of basic immunology and foundational cancer immunotherapy concepts. Faculty selected quiz topics by mapping them to the elective's core learning objectives and to national immunotherapy competency frameworks [35,36]. The prequiz intentionally emphasized basic foundational knowledge, including innate and adaptive immunity, immune cell function, antigen recognition, tumor and immune system interactions, and introductory immunotherapy mechanisms, to ensure that all learners entered the elective with a minimum level of readiness. To determine the length of the quiz, faculty reviewed comparable formative and summative assessments used in preclerkship immunology courses at TTUHSC El Paso and agreed on a concise 15-item structure that would prevent test fatigue while allowing coverage of core concepts. Before administering the quiz, basic science faculty and clinical faculty members conducted a peer review to evaluate question clarity, clinical accuracy, and alignment with learning outcomes. The team performed an informal bias and sensitivity review to ensure that items did not disadvantage students based on specialty interest or prior research experience. Faculty also performed a preliminary item difficulty appraisal using standard immunology benchmarks to ensure that the quiz appropriately captured foundational, not advanced, knowledge. Students completed the pre-elective quiz electronically on the first day of the course. Faculty used group-level response patterns to identify shared conceptual gaps and to tailor emphasis during synchronous discussions.
- *Immunology primer review:* Students reviewed a self-paced immunology primer that reinforced foundational concepts needed to analyze immunotherapy mechanisms and clinical applications. The primer served as the common baseline for all subsequent discussions.
- *Introduction to cancer immunotherapy assigned readings:* Assigned readings introduced the historical evolution of immunotherapy, early discoveries in tumor immunology, and the scientific rationale behind modern immune-based cancer treatments. Students were expected to extract key

mechanistic principles and discuss them during faculty-facilitated sessions.

- *Interactive discussion board (collaborative inquiry):* Over 2 weeks, students participated in a moderated online discussion board (Canvas LMS). Faculty posted weekly prompts connecting mechanistic immunology to clinical controversies (eg, "Should combination immunotherapy be used upfront in all patients with high tumor burden?"). Students responded to peers, synthesized evidence from the literature, and reflected on the ethical and clinical implications of their arguments. Faculty monitored discussions, asked probing questions, and clarified misconceptions in real time.

Faculty facilitators led structured online discussions using a consistent set of prompts, such as the following:

- What immunologic mechanism explains this patient's response or nonresponse?
- What biomarkers would guide treatment selection?
- How would you monitor for immune-related adverse events, and which patient population is at most risk?
- What therapeutic alternatives exist, and what is the evidence supporting them?

This structure ensured clinically rigorous, evidence-driven dialogue while encouraging students to integrate mechanistic immunology with real-world oncology decision-making.

- *Ethical considerations, patient perspectives, and social determinants of health:* To ensure robust and meaningful discussion of ethics and equity, the faculty developed structured ethical prompts and probes related to access, affordability, and geographic disparities in immunotherapy delivery. Examples included the following: (1) How do insurance status and socioeconomic factors influence a patient's ability to access checkpoint inhibitors? (2) What ethical considerations arise when offering chimeric antigen receptor (CAR) T-cell therapy to patients in rural or underserved regions? (3) How should clinicians navigate treatment decisions when cultural beliefs or health literacy affects understanding of immunotherapy risks?
- *Future directions and research (student-led inquiry):* Students explored emerging topics such as next-generation CAR T-cell constructs, tumor microbiome effects on immunotherapy, predictive biomarkers, and novel combination strategies. Faculty provided a curated list of high-impact reviews and seminal papers. Each student selected a topic, analyzed current evidence, evaluated limitations of existing studies, and proposed future research directions and a novel therapy. Students presented their findings during a final oral presentation and received faculty feedback grounded in scientific rigor and clinical relevance.
- *Postelective quiz:* We designed the postelective quiz to parallel the prequiz in structure, length, and difficulty so that performance changes reflected student learning rather than differences in test format. Like the prequiz, the postquiz included 15 multiple-choice questions aligned with the same foundational and applied immunotherapy concepts. To ensure content validity, basic science and

clinical faculty mapped each postquiz item to the learning objectives addressed during the elective, including immune mechanisms of action, indications for Food and Drug Administration (FDA)-approved immunotherapies, and immune-related adverse events. Faculty conducted the same peer review, clarity check, and bias review procedures used for the prequiz. Student anonymity prevented individual-level tracking; therefore, we analyzed quiz results at the aggregate group level, comparing pre- and postcourse means to identify trends in knowledge acquisition. Students completed the postelective quiz immediately after the final oral presentations. Faculty used the results to evaluate learning outcomes, assess overall instructional effectiveness, and identify content areas requiring refinement in future course iterations.

Primer Development and Foundational Preparation

To strengthen students' readiness for advanced immunotherapy concepts, we provided a self-paced immunology primer at the start of the course. A basic scientist with immunology expertise and a clinical oncologist collaboratively selected the primer topics to ensure alignment with learning objectives and relevance to both foundational and clinical content. Medical students then drafted and refined the primer content, drawing on their experience in the preclerkship curriculum to identify which concepts were most challenging and most critical for clinical application during the clerkship curriculum. This student-driven development process (1) incorporated learners' perspectives to better target foundational gaps and (2) supported student contributors' professional growth through scholarly authorship and early medical education experience. Faculty

reviewed and approved all content to ensure accuracy, clarity, and educational value. Integrating student voices added a peer-informed dimension to the primer, helping bridge the gap between didactic instruction and practical understanding.

Knowledge Assessment

We created pre- and postelective quizzes ([Multimedia Appendix 1](#)) to gauge students' baseline knowledge, measure learning gains, and evaluate the effectiveness of course content and teaching methods. On the first day of the elective, students completed the pre-elective quiz. Afterward, they received the immunology primer ([Multimedia Appendix 2](#)) as a foundational resource. The primer introduced key immunology topics, including innate and adaptive immunity, immune-cell functions, and mechanisms of immune activation. It emphasized how these principles govern the detection and elimination of cancer cells and highlighted the interplay between the immune system and the tumor microenvironment. The primer also incorporated recent advances such as checkpoint inhibitors and CAR T-cell therapy, enabling students to connect core immunological principles to cutting-edge cancer treatments. By reinforcing these concepts, the primer equipped students with the background needed to understand the complexities of cancer immunotherapy.

Table 1 outlines the key topics and learning objectives that guided students through essential features of the tumor microenvironment and the application of immunotherapy in cancer treatment. Students engaged with these objectives through an online discussion board (Canvas LMS) and delivered an oral presentation on the final day of the elective.

Table 1. List of topics and specific objectives for the online discussion and final presentation.

Topic	Objectives
Discuss tumor microenvironment and the basis for using immunotherapy.	<ul style="list-style-type: none"> Outline the mechanism of how cancer cells lead to T-cell exhaustion. Know the basics of other tumor microenvironment effects such as angiogenesis and extracellular matrix changes.
Describe the mechanism of action of immunotherapy techniques.	<ul style="list-style-type: none"> Identify the role of PD-1^a, PD-L1^b, CTLA-4^c inhibitors, and CAR^d T-cells in cancer therapy.
What are the most common combinations of immunotherapy drugs used at the present time?	<ul style="list-style-type: none"> Discuss current effective treatments available such as for melanoma and lung cancer. Identify promising clinical trials and new developments in immunotherapy.
What are the adverse events commonly associated with immunotherapy?	<ul style="list-style-type: none"> Identify common immunotherapy-related adverse events. Know the current terminology used to classify immunotherapy adverse events.
What patient population is more at risk for adverse events while taking immunotherapy?	<ul style="list-style-type: none"> Identify risk factors based on patient demographics, underlying medical conditions, and cancer-type specific reactions. Determine whether a patient is a good candidate for immunotherapy given risk factors.
Novel therapies: Research current treatment modalities as well as developments in the field. Propose a new therapy protocol, given your new knowledge of immunotherapy.	<ul style="list-style-type: none"> Propose a new therapy based on the understanding of immunotherapy mechanisms based on literature review. Differentiate your innovative treatment from what is currently being used or investigated.

^aPD-1: programmed death-1.

^bPD-L1: programmed death ligand-1.

^cCTLA-4: cytotoxic T-lymphocyte antigen 4.

^dCAR: chimeric antigen receptor.

Participants

Participants in this study consisted of fourth-year medical students enrolled at TTUHSC El Paso. We used purposive sampling to target students in their final year who were eligible and available to participate in a 2-week online elective. Students qualified for inclusion in the research study if they voluntarily enrolled in the elective and provided informed consent to participate in the research study by completing postelective surveys. Participation required students to agree to the study procedures and complete all associated assessments. A total of 35 students enrolled in the elective, and 20 of those students completed the postcourse survey.

We focused on fourth-year students because they typically select electives that align with their intended specialty or postgraduate training plans (Table 2). We chose an online format for this elective based on prior institutional success with virtual coursework and to provide maximal flexibility for students who may travel for residency interviews during this period [18].

Table 2 shows the residency programs that students who completed the cancer immunotherapy elective applied to. The distribution illustrates their specialty preferences, with a notably higher number pursuing internal medicine compared with other fields.

Table . Residency programs medical students completing the cancer immunotherapy elective applied to.

Residency program	Students, n
Internal medicine	5
Pediatrics	3
Anesthesiology	3
Emergency medicine	2
Family medicine	2
Med/Peds	2
Otolaryngology	1
Pathology	1
Psychiatry	1
Total number of students	20

Assessment Instruments

We assessed students' knowledge using pre- and postcourse quizzes that immunology and oncology faculty collaboratively designed. Faculty with expertise in immunotherapy mechanisms, curriculum design, and clinical applications created quiz items and aligned them with the elective's core learning objectives. They also drew on national competency frameworks and relevant literature to support construct and content validity. We evaluated student discussion responses and oral presentations for critical thinking, integration of knowledge, and innovation, providing a complementary behavioral measure aligned with Kirkpatrick level 3 outcomes.

Survey Instruments

Oral Presentation Rubric Design and Validation

We evaluated student presentations using a structured rubric intentionally designed to measure four core constructs aligned with the elective's learning objectives:

1. *Conceptual understanding*: Accuracy and depth of immunology and immunotherapy knowledge, including mechanisms of action and tumor-immune interactions.
2. *Critical thinking and innovation*: Ability to synthesize literature, evaluate evidence, identify therapeutic gaps, and propose novel or evidence-based strategies.
3. *Clinical application*: Integration of clinical relevance, patient considerations, and implications for treatment decision-making.
4. *Communication and professionalism*: Clarity of oral communication, logical organization, visual design, and professional engagement with peers and faculty.

Rubric Development

A team of basic science and clinical faculty collaboratively drafted the rubric. To ensure content validity, faculty mapped each rubric domain directly to the course's stated competencies and to national guidelines for cancer immunotherapy education. The team reviewed representative presentation topics from prior years to ensure that the rubric captured the range of cognitive skills expected of fourth-year medical students.

Validation and Quality Assurance

We strengthened the rubric through a multistep validation process:

- *Peer review*: Three faculty with expertise in immunology, internal medicine, and oncology independently reviewed the rubric for accuracy, alignment with competencies, and clarity of scoring criteria.
- *Bias review*: Educational specialists examined the rubric for ambiguous language, construct-irrelevant variance, and potential bias related to topic selection or communication style.
- *Pilot testing*: Faculty applied the rubric to sample presentation summaries from a previous elective offering. This process allowed the team to refine wording, add performance anchors, and calibrate expectations.
- *Calibration session*: Before evaluation began, faculty reviewers participated in a group-norming session. They scored 2 anonymized sample presentations, compared ratings, and discussed scoring discrepancies to promote interrater consistency.

Scoring and Reliability

Two faculty evaluators independently scored each oral presentation using the finalized rubric. We measured interrater reliability by comparing paired scores across all rubric domains. Reviewers resolved discrepancies through discussion until reaching consensus. This dual-review system strengthened score reliability and helped ensure that the rubric consistently captured higher-order reasoning, conceptual accuracy, and clinical integration.

Postcourse Survey Instrument

We developed a postcourse survey to evaluate students' perceptions of the elective and capture qualitative feedback to inform iterative course improvements. We did not use a previously published instrument because no validated tool exists that specifically measures medical students' perceptions of a cancer immunotherapy elective or assesses learning experiences aligned with RE-AIM and Kirkpatrick evaluation frameworks for a cancer immunotherapy course. Instead, we designed a context-specific instrument grounded in established principles of educational assessment and program evaluation.

The postcourse survey consisted of Likert-scale items, open-ended questions, and demographic prompts and was administered anonymously through Qualtrics. To ensure content validity, the course faculty collaboratively generated the survey items. We mapped each item to the course objectives, the targeted competencies, and the relevant Kirkpatrick levels (Reaction and Learning) to confirm alignment. We refined the survey through faculty peer review, focusing on clarity, item relevance, cognitive load, and potential bias in item phrasing.

To enhance face validity, the draft survey was reviewed by 2 educational researchers unaffiliated with the course who provided feedback on structure, scale selection, and representativeness of domains (usefulness, relevance, comprehension, engagement, and areas for improvement). We revised the instrument accordingly. Due to the small sample size and anonymity requirements, we did not calculate reliability coefficients; however, reviewers assessed the consistency of underlying constructs and the appropriateness of response options.

The final instrument assessed the following domains:

- perception of elective usefulness and relevance;
- comprehension of immunotherapy concepts and techniques;
- engagement with course materials and faculty-led discussions;
- identification of areas requiring improvement; and
- perceived closing of knowledge gaps.

This approach allowed us to capture meaningful student feedback while ensuring that the instrument was appropriate for the elective's goals and context.

Data Collection

We administered pre- and postcourse quizzes electronically to assess group-level knowledge acquisition related to basic immunology and cancer immunotherapy. To maintain full anonymity in accordance with institutional review board (IRB) requirements, we collected all quiz responses without any identifiers or linkage markers. No individual-level identifiers were used; therefore, we analyzed all pre- and postcourse quiz data strictly in aggregate and did not attempt to match responses across time points. This approach allowed us to compare group performance before and after the elective while fully protecting student privacy.

Following course completion, we distributed an optional, anonymous postcourse survey through the Qualtrics platform. The survey included Likert-scale items and open-ended prompts assessing students' perceptions of course effectiveness, relevance of the content, confidence in applying immunotherapy concepts, and suggestions for improvement. Course faculty, including educational researchers, developed the survey to ensure alignment with course objectives and with Kirkpatrick level 1 (Reaction) and level 2 (Learning) outcomes.

Ethical Considerations

This study was reviewed by the TTUHSC El Paso IRB and determined to meet the criteria for exemption from full board review as a minimal-risk study (protocol no. E23116). In accordance with the approved IRB protocol, the requirement

for written informed consent was waived. Participants were provided with an information statement describing the study purpose, planned use of data, and the voluntary nature of participation prior to study involvement. No identifiable information was collected, no compensation was provided, and participant privacy and confidentiality were maintained throughout data collection, analysis, and reporting.

Quantitative Data Analysis

We used a mixed methods approach to evaluate course outcomes. For the quantitative analysis, we compared group-level pre- and postcourse quiz scores using a 2-tailed unpaired *t* test. We did not link responses at the individual level, so we analyzed all scores in aggregate. The unpaired *t* test allowed us to assess differences in group means while preserving student anonymity.

When designing the knowledge assessment, we did not generate formal power calculations because this elective enrolled a small cohort (n=35), and the study design prioritized anonymity over paired testing, which limited our ability to estimate individual-level effect sizes. Instead, faculty experts in medical education, immunology, internal medicine, and oncology developed the quiz with the expectation, based on content coverage and alignment with the learning objectives, that students would show modest but measurable improvement in foundational immunotherapy knowledge. Given the cohort size and the use of unpaired aggregate scores, the study was underpowered to detect small differences between pre- and postcourse means.

Qualitative Data Analysis

For qualitative analysis, we examined open-ended survey responses using inductive thematic analysis, following Braun and Clarke's 6-step methodology [37]. We used Braun and Clarke's 6-step approach [37] to thematic analysis, a widely applied framework for identifying and interpreting patterns in qualitative data. This process involves becoming familiar with the data, generating initial codes, organizing codes into preliminary themes, refining and reviewing those themes, defining and naming them, and producing the final analytic narrative. We did not apply any a priori theoretical framework; instead, we allowed themes to emerge directly from the data. Two investigators (JC and HA) independently reviewed all responses and generated initial, descriptive codes that reflected recurring ideas in student comments (eg, "clarity of content," "workload concerns," "value of clinical discussions," "ethical considerations," "confidence applying immunotherapy concepts," and "suggestions for improvement"). The investigators met to compare and refine these codes, resolve discrepancies through consensus, and organize related codes into preliminary categories (eg, "perceived learning gains," "course structure," "engagement and interaction," and "areas needing clarification"). We iteratively refined a codebook as additional patterns emerged. Thematic saturation occurred when no new codes or concepts appeared in subsequent rounds of review, ensuring that the final themes accurately captured the breadth of student feedback and strengthened the trustworthiness of the analysis.

Implementing Student Feedback

As part of the postcourse survey, we asked students to identify areas for improvement and propose refinements to the elective. We systematically reviewed all feedback and used it to guide iterative course modifications. Student suggestions directly informed several enhancements implemented in subsequent offerings. For example, many students requested clearer guidance before the oral presentation. In response, we added a detailed grading rubric, sample presentation topics, and a faculty-led question and answer session to clarify expectations. Students also recommended streamlining the weekly workload. To address this, we adjusted reading assignments and aligned asynchronous modules more intentionally with live discussion topics to reduce redundancy and cognitive load.

These changes demonstrated our commitment to continuous quality improvement and student-centered course design. The iterative refinement process, informed by both qualitative and quantitative feedback, reflects the practical application of the RE-AIM and Kirkpatrick frameworks to enhance course delivery and educational effectiveness.

Study Duration

The study timeline encompassed data collection, analysis, and dissemination of findings over an estimated 1-2 months following the conclusion of the elective. Timely dissemination of findings ensures prompt feedback and informed iterative elective design and delivery improvements, thereby fostering continuous quality enhancement in medical education at TTUHSC El Paso.

Results

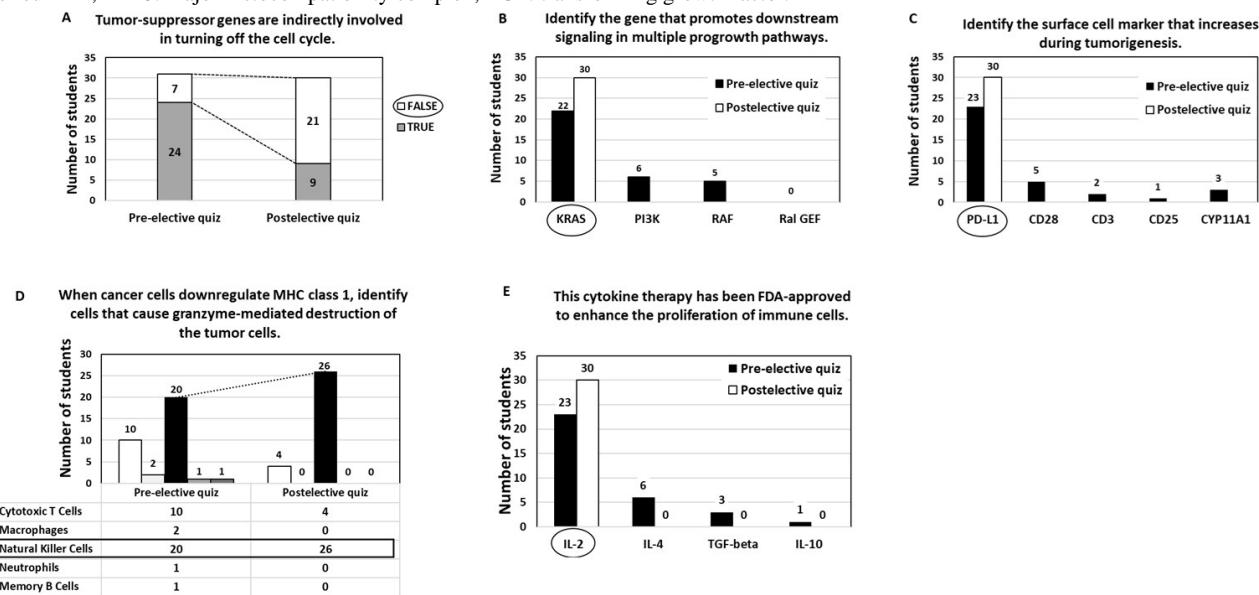
Knowledge Assessment (Pre- and Postcourse Quiz)

A total of 35 students completed the elective, and all quiz data were analyzed in aggregate. Once the students completed the course, they took a postelective quiz. Compared with their performance on the pre-elective quiz, fourth-year medical students demonstrated an improvement in their understanding of basic immunology and cancer immunotherapy topics after completing the elective. For example, students were able to accurately identify the role of tumor-suppressor genes (Figure 1A), the function of KRAS (Figure 1B), and the roles of PD-L1 (Figure 1C) and Natural Killer cells (Figure 1D) within the tumor microenvironment. In addition, students could correctly pinpoint which cytokine therapy was FDA-approved for cancer treatment (Figure 1E).

Figure 1 illustrates the enhanced understanding and knowledge gained by students in these critical areas of cancer biology and immunotherapy after completing the elective. Although the 2-tailed unpaired *t* test did not show statistical significance, postelective quiz scores trended higher than pre-elective scores, suggesting a potential improvement in knowledge. This trend supports the educational value of the elective and highlights the need for further study with paired data.

The study design did not link responses at the individual level, and no power calculation was conducted to estimate detectable effect size; therefore, the analysis remained exploratory rather than inferential. The aggregate data nevertheless provided useful information about baseline and end-of-course performance and will guide future refinement of assessment methods, including the use of paired data and sample-size estimation.

Figure 1. Student proficiency in key concepts of tumor biology and immunotherapy. Compared with pre-elective performance, postelective assessments demonstrated improved student ability to identify and apply foundational concepts in cancer biology and immunotherapy. Specifically, students more accurately recognized the role of tumor-suppressor genes in regulating cell cycle control and preventing malignant transformation (A), as well as the oncogenic function of KRAS in driving aberrant intracellular signaling pathways involved in tumor growth and survival (B). In addition, students showed increased understanding of immune regulatory mechanisms within the tumor microenvironment, including the role of PD-L1 in mediating immune checkpoint inhibition (C) and the contribution of natural killer cells to innate antitumor immune responses (D). Students also correctly identified the Food and Drug Administration-approved cytokine therapy used clinically to enhance immune cell proliferation and activity, reflecting improved recognition of immunomodulatory treatment strategies (E). Together, these results indicate enhanced conceptual understanding across molecular, cellular, and immunologic domains central to contemporary cancer immunotherapy following completion of the elective. FDA: Food and Drug Administration; IL: interleukin-2; MHC: major histocompatibility complex; TGF: transforming growth factor.



Postcourse Survey

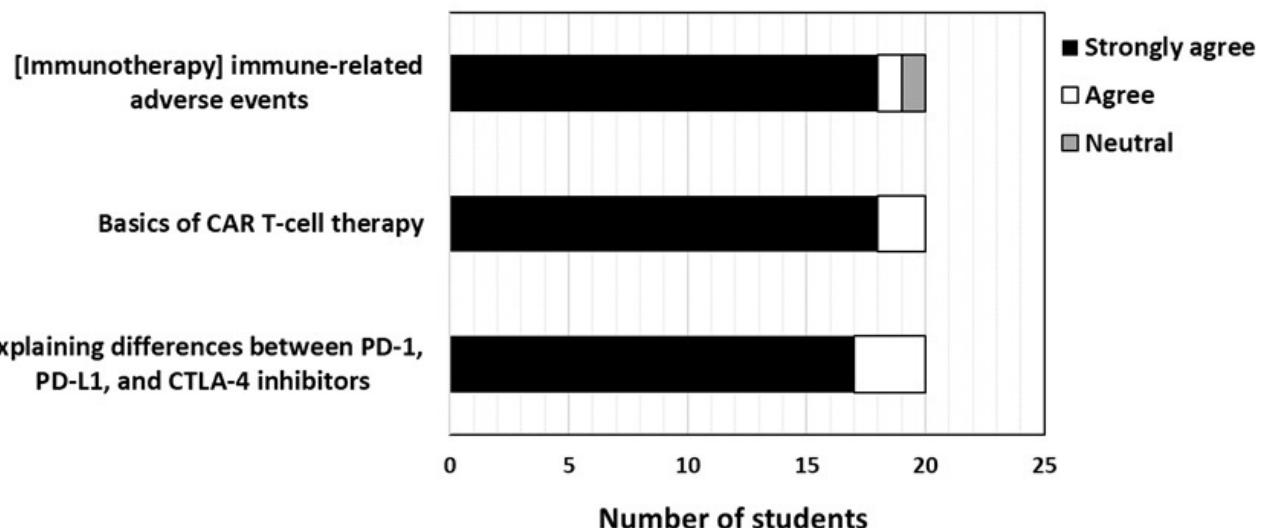
A total of 35 students completed the elective, and 20 students completed the postcourse survey. [Figure 2](#) and [Multimedia Appendix 3](#) summarize the distribution of responses. Students consistently rated the elective favorably across all surveyed domains.

Evaluation of student perceptions demonstrated meaningful gains in their understanding of immune-related adverse events

and core immunotherapy mechanisms. Students expressed increased confidence in identifying and describing the types of immune-related adverse events observed in patients undergoing immunotherapy ([Figure 2](#), top). They also reported greater mastery of fundamental CAR T-cell principles ([Figure 2](#), middle). Importantly, every student agreed that the course improved their ability to distinguish between programmed death-1 (PD-1), programmed death ligand-1 (PD-L1), and cytotoxic T-lymphocyte antigen 4 (CTLA-4) ([Figure 2](#) bottom).

Figure 2. Exploring medical students' understanding of immune-related adverse events and cancer immunotherapy treatments. Students felt knowledgeable about various types of immune-related adverse events observed in some patients receiving immunotherapy (top). Students felt more knowledgeable about the basics of CAR T-cell therapy (middle). All students agreed that the cancer immunotherapy course enabled them to explain the differences between PD-1, PD-L1, and CTLA-4 (bottom). CAR: chimeric antigen receptor; CTLA-4, cytotoxic T-lymphocyte antigen 4; PD-1: programmed death-1; PD-L1, programmed death ligand-1.

Upon completion of this elective, students felt more knowledgeable about:



Students agreed or strongly agreed that the course was instrumental in learning current immunotherapy techniques ([Multimedia Appendix 3A](#)). Students highlighted that the course provided a solid foundation and deeper understanding of cutting-edge immunotherapy practices. Overall, students felt well prepared to apply these techniques in real-world settings.

In addition, the course significantly enhanced students' ability to think critically about developing novel cancer immunotherapies. Nineteen out of twenty students agreed or strongly agreed that the course's comprehensive curriculum and emphasis on innovative thinking fostered their analytical skills and creativity ([Multimedia Appendix 3B](#)). Through understanding basic science concepts and reviewing the current literature, students were encouraged to explore and devise new strategies for cancer treatment. This critical thinking approach not only deepened their understanding of current immunotherapy techniques but also inspired them to contribute to future advancements in the field.

Finally, in a unanimous consensus, medical students have expressed overwhelming satisfaction with the comprehensive coverage of cancer immunotherapy in their recent course. Each student acknowledged that the course effectively addressed and filled significant gaps in their understanding of immunotherapy ([Multimedia Appendix 3C](#)). From foundational concepts to advanced applications, the material provided clarity on complex mechanisms and cutting-edge research in the field. This collective sentiment underscores the course's success in equipping future health care professionals with essential knowledge and skills crucial for addressing the challenges of modern cancer treatment.

Twenty students completed the postelective survey. [Table 3](#) shows median scores with IQRs for all Likert-scale items.

Students reported overwhelmingly positive perceptions of the cancer immunotherapy elective, with consistently high ratings across all survey items. For every item, the median response was 5 (Strongly agree), indicating strong endorsement of the course's effectiveness. Students agreed that the elective was useful for their learning regarding current immunotherapy techniques (median 5, IQR 0) and that the introductory immunology overview supported their understanding of later content (median 5, IQR 0.25).

Students also felt that the course enhanced their ability to think critically about developing novel immunotherapy approaches (median 5, IQR 0) and improved their ability to explain differences among PD-1, PD-L1, and CTLA-4 inhibitors (median 5, IQR 0). Similarly, learners reported increased knowledge of CAR T-cell therapy (median 5, IQR 0) and felt that the amount of course material was appropriate to address gaps in their prior knowledge (median 5, IQR 0).

Students further indicated that their understanding of immune-related adverse events improved as a result of the elective (median 5, IQR 0). The only item with notable response variability was students' confidence in identifying steps needed to manage patients receiving immunotherapy during residency (median 5, IQR 1.0), suggesting greater heterogeneity in perceived readiness for clinical application.

Overall, these results demonstrate consistently strong satisfaction with the elective and meaningful perceived gains in foundational and applied cancer immunotherapy knowledge. Collectively, these findings suggest high perceived educational value and alignment between course objectives and learner-reported outcomes. A summary of the highest-rated domains is demonstrated in [Table 3](#).

Table . Twenty students completed the postelective survey^a.

Item	Median	IQR
Q1. Course usefulness for learning immunotherapy techniques.	5	0
Q2. Immunology overview helped understanding.	5	0.25
Q3. Course enabled critical thinking about novel immunotherapy approaches.	5	0
Q4. Ability to explain PD-1 ^b , PD-L1 ^c , and CTLA-4 ^d differences.	5	0
Q5. Increased knowledge of CAR ^e T-cell therapy.	5	0
Q6. Amount of information was appropriate.	5	0
Q7. Increased knowledge of immune-related adverse events.	5	0
Q8. Ability to identify steps to manage patients receiving immunotherapy.	5	1.0

^aMedian scores and IQRs indicate overwhelmingly positive perceptions of the elective, with all items receiving a median of 5 (Strongly agree). These findings reflect strong alignment between the course objectives and students' reported gains in cancer immunotherapy knowledge.

^bPD-1: programmed death-1.

^cPD-L1: programmed death ligand-1.

^dCTLA-4: cytotoxic T-lymphocyte antigen 4.

^eCAR: chimeric antigen receptor.

Qualitative Themes From Open-Ended Responses

Inductive thematic analysis of open-ended survey responses revealed 4 dominant themes that collectively illustrate how students experienced the elective:

- Theme 1: Increased Conceptual Clarity reflected students' reports of gaining a clearer understanding of immune checkpoint pathways, mechanisms of T-cell activation, and distinctions among PD-1, PD-L1, and CTLA-4.
- Theme 2: Strengthened Ability to Apply Concepts Clinically captured students' increased confidence in connecting mechanistic immunology to therapeutic decision-making, recognition of immune-related adverse events, and understanding FDA-approved therapies such as CAR T-cell therapy.
- Theme 3: Value of Structured, Faculty-Guided Discussion highlighted the importance of guided prompts and faculty facilitation in promoting deeper inquiry and enabling sustained dialogue around ethics, access to care, cost barriers, disparities in clinical trials, and social determinants of health shaping immunotherapy outcomes.
- Theme 4: Recommendations for Course Enhancement included requests for clearer expectations for oral presentations, streamlined weekly readings, and expanded case-based discussions, particularly those incorporating literature focused on marginalized populations.

These qualitative insights not only informed targeted course refinements but also aligned with the evaluative dimensions of the RE-AIM and Kirkpatrick models described in the "Methods" section. Students also expressed strong appreciation for the structured discussion prompts, faculty engagement, and the primer's effectiveness in scaffolding learning.

Discussion

Principal Findings

This mixed methods evaluation demonstrated that a fourth-year online cancer immunotherapy elective improved students' foundational understanding of immunotherapy principles, strengthened their ability to apply mechanistic knowledge to clinical scenarios, and enhanced their confidence discussing emerging therapies. Students showed measurable gains in core immunology and immunotherapy concepts on the postelective knowledge assessment, and their final oral presentations provided evidence of higher-order reasoning, integration of basic and clinical science, and accurate articulation of complex mechanisms such as PD-1/PD-L1 and CTLA-4 pathways, CAR T-cell engineering, and immune-related toxicities. Qualitative analysis of postelective survey responses further revealed that students valued the elective's relevance, clarity, and emphasis on clinical translation, noting that it addressed significant gaps in their prior training and supported their ability to think critically about current and emerging cancer immunotherapies. Collectively, these findings indicate that the elective achieved its intended goal of enhancing immunotherapy competency among senior medical students.

Implications of Findings

The findings from this evaluation demonstrate that a structured, 2-week online cancer immunotherapy elective can enhance medical students' self-reported understanding of fundamental immunotherapy concepts, mechanisms of action, and FDA-approved treatments. Students also indicated that the elective strengthened their ability to critique scientific literature and articulate clinical applications, skills reinforced through

structured discussion prompts and a scored oral presentation. These results suggest that even a short, focused elective can support competency development in emerging therapeutic domains.

The mixed methods data further indicate that pairing asynchronous foundational content (eg, the immunology primer) with guided synchronous discussions and a final presentation provides an effective instructional sequence for fourth-year learners. Students consistently rated the course as useful, relevant, and appropriately rigorous, reflecting the value of integrating preexisting knowledge, analytic tasks, and application-based assessments. This supports the feasibility of using similar instructional designs to introduce other rapidly evolving biomedical topics.

In addition, the elective's structure, which included evidence-based discussion prompts and opportunities for students to explain mechanisms to peers, offered a model that other institutions may adapt when integrating novel scientific content into the curriculum. Implementing a peer-informed approach to resource development (eg, the student-written primer) may also enhance perceived relevance and learning efficiency for medical students.

These implications extend primarily to curriculum design, instructional sequencing, and feasibility of implementation in undergraduate medical education. While the course addressed selected ethical and access-related issues within immunotherapy during discussions, the study did not evaluate growth in students' understanding of social determinants of health or culturally competent care. Future research using targeted assessments could explore these domains more explicitly.

Comparison With Literature

The integration of immunotherapy into undergraduate medical education remains limited, yet our findings mirror a growing body of evidence supporting earlier and more structured exposure to advanced cancer therapeutics. Several studies have reported that medical trainees feel underprepared to discuss immunotherapy, interpret mechanism-of-action data, or counsel patients about associated risks and benefits, a gap that our elective similarly sought to address [38]. Recent educational interventions have been developed during graduate medical education, including short courses, case-based modules, and interdisciplinary workshops.

Other research teams have emphasized the value of combining foundational basic science reviews with clinical case discussions to enhance long-term retention and clinical reasoning, an approach consistent with our spiral design and primer-supported structure [39,40]. These parallels support the relevance and effectiveness of our elective's hybrid instructional format.

There is also a small but emerging literature on the importance of tailoring immunotherapy education to local contexts, including programs serving diverse or medically underserved communities. These studies underscore the need for cultural and demographic relevance when teaching advanced therapeutics, a theme that aligns with, but does not overextend beyond, the reflections shared by our students regarding practice readiness and communication skills [41-45].

Together, these converging findings suggest that structured, evidence-informed immunotherapy curricula can enhance medical students' conceptual understanding and clinical preparedness. They also highlight the broader national momentum toward incorporating cutting-edge oncologic therapeutics earlier in medical training. Future multi-institutional studies will be essential to compare outcomes across diverse curricular models and to determine best practices for sustainable integration.

Strengths and Limitations

A major strength of this elective was its alignment with up-to-date clinical and scientific standards in cancer immunotherapy. Due to the field evolving rapidly, faculty routinely reviewed and updated course content using evidence-based guidelines from the American Society of Clinical Oncology and the Society for Immunotherapy of Cancer. This process ensured that students engaged with current FDA-approved therapies, emerging research directions, and clinically relevant case examples, thereby reinforcing the scientific rigor and clinical applicability of the curriculum.

Another strength was the elective's intentional design for a modular, hybrid structure, blending asynchronous learning with synchronous faculty-led discussion, allowing the course to adapt to varying institutional schedules and curricular models. In addition, its use of structured discussion prompts and a capstone innovation project created flexible instructional components that other medical schools could readily adopt or tailor to their own contexts.

The course also benefited from strong interdisciplinary collaboration. Faculty from immunology, oncology, biochemistry, clinical genetics, and internal medicine jointly developed and implemented the elective, ensuring cohesion across basic science and clinical content. Student cocreation, particularly in the development of the immunology primer, added a peer-informed dimension that enhanced clarity, relevance, and learner engagement.

Despite these strengths, several limitations warrant attention. First, participation in both the elective and the postcourse survey was voluntary, introducing potential self-selection and response bias that may limit the generalizability of the findings. Second, the retrospective nature of some perception-based survey items may have contributed to recall bias. Third, although the pre- and postcourse quizzes assessed knowledge gains, all data were analyzed at the aggregate level to maintain anonymity. The absence of individually paired responses limited our ability to conduct more detailed analyses of individual learning trajectories or matched-pair statistical testing.

The elective also faced implementation challenges, including variable student schedules, faculty workload constraints, and occasional technological barriers inherent to online instruction. Students on clinical rotations sometimes experienced difficulties synchronizing schedules with synchronous components. While some students pursued optional shadowing opportunities with oncologists, the elective did not include a formal clinical rotation, representing a potential area for future enhancement.

To address these challenges, we incorporated contingency measures such as recorded sessions, flexible deadlines, and alternative participation formats for discussion activities. Regular faculty meetings and scheduled check-ins with students allowed the instructional team to respond to emerging needs in real time. Attrition was minimal, as all enrolled students completed the elective; however, not all students completed the postcourse survey.

Lessons Learned

Implementing this online cancer immunotherapy elective provided valuable insights into curriculum development, learner engagement, and program scalability. Early and meaningful collaboration between basic scientists, clinicians, and students was essential to ensure content relevance and clarity. Notably, the cocreation of the immunotherapy primer by medical students offered a peer-informed perspective that enhanced both accessibility and professional development for the authors.

Flexibility in course delivery, through asynchronous modules supplemented by live sessions, proved crucial for accommodating students' clinical schedules and different learning styles. However, we learned that clear alignment between asynchronous and synchronous content was essential to avoid redundancy and promote continuity. Later course iterations were adjusted to better integrate these components.

Embedding structured opportunities for innovation, such as the final oral presentation task, fostered critical thinking and interdisciplinary application. Students were encouraged to propose novel therapeutic strategies, and faculty feedback emphasized this capstone experience as one of the most valuable elements of the elective.

Continuous improvement based on student feedback was also key. The anonymous survey asked students to identify strengths and areas for improvement, which led to enhancements in discussion prompts, session timing, and technical delivery. These efforts reinforced a culture of responsiveness and reflective teaching and demonstrated the importance of iterative design in building a sustainable, scalable intervention.

Future Directions

Future research should explore how this elective contributes to long-term competence in cancer immunotherapy, especially during postgraduate training. While short-term outcomes demonstrated knowledge gain and student satisfaction, it remains critical to assess whether these improvements translate to clinical proficiency, leadership in immunotherapy initiatives, and patient-centered decision-making.

A key next step is longitudinal tracking of alumni who completed the elective, evaluating how early exposure to cancer immunotherapy influences clinical performance, advocacy, and continued professional development. This may include follow-up surveys, interviews, or program director assessments to gauge real-world application of course content.

Alternatively, conducting a cross-sectional cohort study comparing students who completed the elective with those who did not would offer insight into differences in long-term knowledge retention, confidence in discussing immunotherapy with patients, and likelihood of pursuing related research or clinical specialties.

To improve generalizability, future iterations of the elective should be piloted across additional medical schools with varying curricular structures and student demographics. Multi-institutional collaborations could help evaluate the adaptability, feasibility, and sustainability of this instructional model and identify institution-specific facilitators and barriers. Incorporating paired pre-post knowledge assessments or longitudinal tracking of clinical application skills may also strengthen future evaluation designs.

Expanding the elective to include interprofessional learners, such as pharmacy, nursing, and physician assistant students, could enhance team-based education and cancer care delivery. Finally, multisite implementation at diverse institutions will be critical for testing the generalizability and adaptability of the curriculum, and future iterations must remain agile to incorporate emerging therapies and updated guidelines in this rapidly evolving field.

Acknowledgments

We thank the students for participating in the course and the research study.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Data Availability

Data can be accessed upon request to the corresponding author.

Authors' Contributions

Conceptualization: MR, RH, BG, SM, HA, CP, JC

Data curation: RH, HA, JC

Formal analysis: RH, HA, JC

Investigation: MR, RH, SM, HA, CP, JC

Methodology: MR, RH, SM, HA, CP, JC

Project administration: SM, HA, CP, JC

Resources: MR, RH, BG, SM, HA, CP, JC

Software: MR, RH, HA, JC

Supervision: SM, HA, CP, JC

Validation: SM, HA, CP, JC

Visualization: MR, RH, HA, JC

Writing – original draft: MR, RH, BG, SM, HA, CP, JC

Writing – review & editing: MR, RH, BG, SM, HA, CP, JC

Conflicts of Interest

None declared.

Multimedia Appendix 1

Pre- and postelective quiz instrument. This quiz assessed medical students' baseline and postcourse knowledge of cancer immunotherapy, including mechanisms of action, clinical applications, and immune-related adverse events. The same set of questions was administered before and after the elective to evaluate knowledge gains and support course effectiveness analysis.

[[PDF File, 143 KB - mededu_v12i1e71628_app1.pdf](#)]

Multimedia Appendix 2

Immunology primer overview. This primer provided foundational content on the immune system, including innate and adaptive immunity, key immune cells, cytokine signaling, and tumor immunology concepts. It was designed to ensure that students entered the elective with a baseline understanding necessary to engage in advanced cancer immunotherapy discussions.

[[PDF File, 1298 KB - mededu_v12i1e71628_app2.pdf](#)]

Multimedia Appendix 3

Comprehensive impact of the immunotherapy course on medical students' knowledge and critical thinking. (A) Students strongly agreed that the course was instrumental in learning current immunotherapy techniques, providing up-to-date content and practical applications that prepared them well for real-world settings. (B) The new immunotherapy course significantly enhanced students' critical thinking in developing novel cancer treatments, with 19 out of 20 students affirming its impact. (C) Feedback from 20 medical students unanimously confirming that the content covered in the cancer immunotherapy course significantly filled gaps and deficiencies in their knowledge of immunotherapy. Each student reported agreement with the comprehensive nature of the course material, highlighting its effectiveness in addressing both fundamental concepts and advanced applications in the field of cancer treatment.

[[JPG File, 127 KB - mededu_v12i1e71628_app3.jpg](#)]

References

1. Abbott M, Ustoyev Y. Cancer and the immune system: the history and background of immunotherapy. *Semin Oncol Nurs* 2019 Oct;35(5):150923. [doi: [10.1016/j.soncn.2019.08.002](https://doi.org/10.1016/j.soncn.2019.08.002)]
2. Ginex PK, Brassil K, Ely B. Immunotherapy: exploring the state of the science. *Clin J Oncol Nurs* 2017 Apr 1;21(2 Suppl):9-12. [doi: [10.1188/17.CJON.S2.9-12](https://doi.org/10.1188/17.CJON.S2.9-12)] [Medline: [28315554](#)]
3. Le I, Dhandayuthapani S, Chacon J, Eiring AM, Gadad SS. Harnessing the immune system with cancer vaccines: from prevention to therapeutics. *Vaccines (Basel)* 2022 May 21;10(5):816. [doi: [10.3390/vaccines10050816](https://doi.org/10.3390/vaccines10050816)]
4. Kye Y, Nagineni L, Gadad S, et al. The identification and clinical applications of mutated antigens in the era of immunotherapy. *Cancers (Basel)* 2022 Aug 31;14(17):4255. [doi: [10.3390/cancers14174255](https://doi.org/10.3390/cancers14174255)]
5. Ramirez F, Zambrano A, Hennis R, Holland N, Lakshmanaswamy R, Chacon J. Sending a message: use of mRNA vaccines to target the tumor immune microenvironment. *Vaccines (Basel)* 2023 Sep 7;11(9):1465. [doi: [10.3390/vaccines11091465](https://doi.org/10.3390/vaccines11091465)]
6. Ha MT, Siddiqui ZS. Understanding medical students' transition to clinical training: a qualitative study of transformative learning and professional identity formation. *BMJ Open* 2025 Jun 17;15(6):e098675. [doi: [10.1136/bmjopen-2024-098675](https://doi.org/10.1136/bmjopen-2024-098675)] [Medline: [40527562](#)]
7. Mirzadeh P, Pituskin E, Au I, Sneath S, Buick CJ. Cancer immunotherapy: the role of nursing in patient education, assessment, monitoring, and support. *Curr Oncol* 2025 Jul 9;32(7):7. [doi: [10.3390/curroncol32070392](https://doi.org/10.3390/curroncol32070392)] [Medline: [40710202](#)]
8. Holtrup JS, Estabrooks PA, Gaglio B, et al. Understanding and applying the RE-AIM framework: clarifications and resources. *J Clin Transl Sci* 2021;5(1):e126. [doi: [10.1017/cts.2021.789](https://doi.org/10.1017/cts.2021.789)] [Medline: [34367671](#)]
9. Glasgow RE, Battaglia C, McCreight M, et al. Use of the reach, effectiveness, adoption, implementation, and maintenance (RE-AIM) framework to guide iterative adaptations: applications, lessons learned, and future directions. *Front Health Serv* 2022;2:959565. [doi: [10.3389/frhs.2022.959565](https://doi.org/10.3389/frhs.2022.959565)] [Medline: [36925843](#)]

10. Rouse DN. Employing Kirkpatrick's evaluation framework to determine the effectiveness of health information management courses and programs. *Perspect Health Inf Manag* 2011 Apr 1;8(Spring):1c. [Medline: [21464860](#)]
11. Chacon JA, Cervantes JL, Perry CN, Pfarr CM, Ayoubieh H. Personalized medicine in undergraduate medical education: a spiral learning model. *Med Sci Educ* 2020 Dec;30(4):1741-1744. [doi: [10.1007/s40670-020-01066-0](#)] [Medline: [34457840](#)]
12. DeSilets LD. An update on Kirkpatrick's model of evaluation: part two. *J Contin Educ Nurs* 2018 Jul 1;49(7):292-293. [doi: [10.3928/00220124-20180613-02](#)] [Medline: [29939374](#)]
13. Johnston S, Coyer FM, Nash R. Kirkpatrick's evaluation of simulation and debriefing in health care education: a systematic review. *J Nurs Educ* 2018 Jul 1;57(7):393-398. [doi: [10.3928/01484834-20180618-03](#)] [Medline: [29958308](#)]
14. Whitcomb ME. Medical education research. *Acad Med* 2004 Oct;79(10):907. [doi: [10.1097/00001888-200410000-00001](#)] [Medline: [15383344](#)]
15. Hankir AK, Northall A, Zaman R. Stigma and mental health challenges in medical students. *BMJ Case Rep* 2014 Sep 2;2014:bcr2014205226. [doi: [10.1136/bcr-2014-205226](#)] [Medline: [25183806](#)]
16. Jupina M, Sidle MW, Rehmeyer Caudill CJ. Medical student mental health during the COVID-19 pandemic. *Clin Teach* 2022 Oct;19(5):e13518. [doi: [10.1111/tct.13518](#)] [Medline: [35909320](#)]
17. Pacheco JP, Giacomini HT, Tam WW, et al. Mental health problems among medical students in Brazil: a systematic review and meta-analysis. *Braz J Psychiatry* 2017;39(4):369-378. [doi: [10.1590/1516-4446-2017-2223](#)] [Medline: [28876408](#)]
18. Ayoubieh H, Perry C, Cervantes J, Pfarr C, Chacon J. Development of an online SARS-CoV-2/COVID-19 elective course for fourth-year medical students. *Med Sci Educ* 2023 Feb;33(1):21-25. [doi: [10.1007/s40670-023-01726-x](#)] [Medline: [36713276](#)]
19. Guo MZ, Allen J, Sakamoto M, Pahwa A, Santhosh L. Reimagining Undergraduate Medical Education in a Post-COVID-19 Landscape. *J Gen Intern Med* 2022 Jul;37(9):2297-2301. [doi: [10.1007/s11606-022-07503-7](#)] [Medline: [35710661](#)]
20. Kim KJ, Kim G, Kang Y. Faculty perceptions and use of e-learning resources for medical education and future predictions. *Korean J Med Educ* 2023 Dec;35(4):325-334. [doi: [10.3946/kjme.2023.270](#)] [Medline: [38062680](#)]
21. Tosto SA, Alyahya J, Espinoza V, McCarthy K, Tcherni-Buzzeo M. Online learning in the wake of the COVID-19 pandemic: mixed methods analysis of student views by demographic group. *Soc Sci Humanit Open* 2023;8(1):100598. [doi: [10.1016/j.ssaho.2023.100598](#)] [Medline: [37366390](#)]
22. Harden RM. The integration ladder: a tool for curriculum planning and evaluation. *Med Educ* 2000 Jul;34(7):551-557. [doi: [10.1046/j.1365-2923.2000.00697.x](#)] [Medline: [10886638](#)]
23. 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](#)]
24. Modo M, Kinchin I. A conceptual framework for interdisciplinary curriculum design: a case study in neuroscience. *J Undergrad Neurosci Educ* 2011;10(1):A71-A79. [Medline: [23626496](#)]
25. Chaffey L, de Leeuw EJ, Finnigan G. Facilitating students' reflective practice in a medical course: literature review. *Educ Health* 2012;25(3):198. [doi: [10.4103/1357-6283.109787](#)]
26. Rogers SL, Priddis LE, Michels N, Tieman M, Van Winkle LJ. Applications of the reflective practice questionnaire in medical education. *BMC Med Educ* 2019 Feb 7;19(1):47. [doi: [10.1186/s12909-019-1481-6](#)] [Medline: [30732611](#)]
27. Sawyer T, White M, Zaveri P, et al. Learn, see, practice, prove, do, maintain: an evidence-based pedagogical framework for procedural skill training in medicine. *Acad Med* 2015 Aug;90(8):1025-1033. [doi: [10.1097/ACM.0000000000000734](#)] [Medline: [25881645](#)]
28. Engqvist Boman L, Sundberg K, Petersson LM, Backman M, Silén C. A pedagogical model to enhance nurses' ability to support patient learning: an educational design research study. *Int J Med Educ* 2022 Jul 29;13:176-186. [doi: [10.5116/ijme.62c2.b9c4](#)]
29. Manninen K, Henriksson EW, Scheja M, Silén C. Patients' approaches to students' learning at a clinical education ward—an ethnographic study. *BMC Med Educ* 2014 Jul 2;14:131. [doi: [10.1186/1472-6920-14-131](#)] [Medline: [24989155](#)]
30. Haidet P, Jarecke J, Adams NE, et al. A guiding framework to maximise the power of the arts in medical education: a systematic review and metasynthesis. *Med Educ* 2016 Mar;50(3):320-331. [doi: [10.1111/medu.12925](#)] [Medline: [26896017](#)]
31. Alharbi NS. Evaluating competency-based medical education: a systematized review of current practices. *BMC Med Educ* 2024 Jun 3;24(1):612. [doi: [10.1186/s12909-024-05609-6](#)] [Medline: [38831271](#)]
32. Lee GB, Chiu AM. Assessment and feedback methods in competency-based medical education. *Ann Allergy Asthma Immunol* 2022 Mar;128(3):256-262. [doi: [10.1016/j.anai.2021.12.010](#)] [Medline: [34929390](#)]
33. Levine MF, Shorten G. Competency-based medical education: its time has arrived. *Can J Anaesth* 2016 Jul;63(7):802-806. [doi: [10.1007/s12630-016-0638-6](#)] [Medline: [27030130](#)]
34. Postow MA, Callahan MK, Wolchok JD. Immune checkpoint blockade in cancer therapy. *J Clin Oncol* 2015 Jun 10;33(17):1974-1982. [doi: [10.1200/JCO.2014.59.4358](#)] [Medline: [25605845](#)]
35. Wiley K, LeFebvre KB, Wall L, et al. Immunotherapy administration: oncology nursing society recommendations. *Clin J Oncol Nurs* 2017 Apr 1;21(2 Suppl):5-7. [doi: [10.1188/17.CJON.S2.5-7](#)] [Medline: [28315551](#)]
36. Galioto M, Mucenski J. Immunotherapy summit: proceedings and identified priorities for safe administration and care. *Clin J Oncol Nurs* 2019 Jun 1;23(3):E60-E65. [doi: [10.1188/19.CJON.E60-E65](#)] [Medline: [31099794](#)]

37. Lim AJS, Hong DZ, Pisupati A, et al. Portfolio use in postgraduate medical education: a systematic scoping review. *Postgrad Med J* 2023 Jul 21;99(1174):913-927. [doi: [10.1093/postmj/qgac007](https://doi.org/10.1093/postmj/qgac007)] [Medline: [36961214](https://pubmed.ncbi.nlm.nih.gov/36961214/)]
38. Fulton-Ward T, Bain R, Khoury EG, et al. Benefits of mentoring in oncology education for mentors and mentees: pre-post interventional study of the British Oncology Network for Undergraduate Societies' national oncology mentorship scheme. *JMIR Med Educ* 2023 Sep 11;9:e48263. [doi: [10.2196/48263](https://doi.org/10.2196/48263)] [Medline: [37695662](https://pubmed.ncbi.nlm.nih.gov/37695662/)]
39. Alkurran O, Al-Mehaisen L, Abu Mahfouz I, et al. Distance electronic learning strategy in medical teaching during the COVID-19 pandemic: cross-sectional survey study. *JMIR Med Educ* 2023 Dec 5;9:e42354. [doi: [10.2196/42354](https://doi.org/10.2196/42354)] [Medline: [38051556](https://pubmed.ncbi.nlm.nih.gov/38051556/)]
40. Kenjrawi Y, Dashash M. The first asynchronous online evidence-based medicine course for Syrian health workforce: effectiveness and feasibility pilot study. *JMIR Form Res* 2022 Oct 25;6(10):e36782. [doi: [10.2196/36782](https://doi.org/10.2196/36782)] [Medline: [36282556](https://pubmed.ncbi.nlm.nih.gov/36282556/)]
41. Emens LA, Romero PJ, Anderson AC, et al. Challenges and opportunities in cancer immunotherapy: a Society for Immunotherapy of Cancer (SITC) strategic vision. *J Immunother Cancer* 2024 Jun 19;12(6):6. [doi: [10.1136/jitc-2024-009063](https://doi.org/10.1136/jitc-2024-009063)] [Medline: [38901879](https://pubmed.ncbi.nlm.nih.gov/38901879/)]
42. Gupta A, Omeogu C, Islam JY, et al. Socioeconomic disparities in immunotherapy use among advanced-stage non-small cell lung cancer patients: analysis of the National Cancer Database. *Sci Rep* 2023 May 20;13(1):8190. [doi: [10.1038/s41598-023-35216-2](https://doi.org/10.1038/s41598-023-35216-2)]
43. Vasquez VM Jr, McCabe M, McKee JC, et al. Transforming cancer care: a narrative review on leveraging artificial intelligence to advance immunotherapy in underserved communities. *J Clin Med* 2025 Jul 29;14(15):15. [doi: [10.3390/jcm14155346](https://doi.org/10.3390/jcm14155346)] [Medline: [40806968](https://pubmed.ncbi.nlm.nih.gov/40806968/)]
44. Ramirez F, Alvarado L, Dwivedi A, Chacon J. Immunotherapy treatment disparities: a Texas Cancer Registry analysis of patients with cutaneous melanoma. *J Immigr Minor Health* 2024 Feb;26(1):23-33. [doi: [10.1007/s10903-023-01536-8](https://doi.org/10.1007/s10903-023-01536-8)] [Medline: [37615821](https://pubmed.ncbi.nlm.nih.gov/37615821/)]
45. Ramirez F, Vasquez R, Gill J, et al. Understanding the association between social determinants of health and receipt of immunotherapy for melanoma. *J Drugs Dermatol* 2024 May 1;23(5):311-315. [doi: [10.36849/JDD.7803](https://doi.org/10.36849/JDD.7803)] [Medline: [38709695](https://pubmed.ncbi.nlm.nih.gov/38709695/)]

Abbreviations

CAR: chimeric antigen receptor

CTLA-4: cytotoxic T-lymphocyte antigen 4

FDA: Food and Drug Administration

IRB: institutional review board

PD-1: programmed death-1

PD-L1: programmed death ligand-1

RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance

TTUHSC: Texas Tech University Health Sciences Center

Edited by B Lesselroth; submitted 22.Jan.2025; peer-reviewed by AN Abbasi, RR Kameny, S Manivasakan; revised version received 01.Dec.2025; accepted 03.Dec.2025; published 21.Jan.2026.

Please cite as:

Raynor M, Hock R, Godinich B, Maharaj S, Ayoubieh H, Perry C, Chacon J

Implementation and Evaluation of a Cancer Immunotherapy Elective for Medical Students: Mixed Methods Descriptive Study

JMIR Med Educ 2026;12:e71628

URL: <https://mededu.jmir.org/2026/1/e71628>

doi: [10.2196/71628](https://doi.org/10.2196/71628)

© Mark Raynor, Rivers Hock, Brandon Godinich, Satish Maharaj, Houriya Ayoubieh, Cynthia Perry, Jessica Chacon. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 21.Jan.2026. 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.

Original Paper

Effect of an Online Continuing Professional Development Course on Physicians' Intention to Approach a Colleague in Difficulty: Mixed Methods Convergent Study

Florence Lizotte¹, MD, MSc; Martin Tremblay², PhD; Caroline Biron^{3,4}, PhD; Éloi Lachance¹, BA; Souleymane Gadio³, MSc; Roberta de Carvalho Corôa³, MSc, PhD; Claude Bernard Uwizeye³, MSc; Sam J Daniel^{2,5}, MD, MS; France Légaré^{1,3,6}, BArch, MD, MSc, PhD

¹Faculty of Medicine, Université Laval, Québec, QC, Canada

²Fédération des médecins spécialistes du Québec, Montréal, QC, Canada

³VITAM – Centre de recherche en santé durable, Université Laval, Québec, QC, Canada

⁴Department of Management, Université Laval, Québec, QC, Canada

⁵Faculty of Medicine, McGill University, Montreal, QC, Canada

⁶Centre de recherche du CHU de Québec, Université Laval, Québec, QC, Canada

Corresponding Author:

France Légaré, BArch, MD, MSc, PhD

VITAM – Centre de recherche en santé durable

Université Laval

2480 Chemin de la Canardière

Québec, QC, G1J 2G1

Canada

Phone: 1 (418) 663 5313 ext 12286

Email: France.Legare@fmed.ulaval.ca

Abstract

Background: Burnout and psychological distress are prevalent among physicians. Peer support appears to play a protective role, yet little is known about training interventions that motivate physicians to approach peers in difficulty, as such effects are often overlooked or assessed using nonvalidated tools.

Objective: We evaluated the effects of an online continuing professional development (CPD) course designed to increase physicians' intention to approach a colleague in difficulty.

Methods: Physicians who completed a 1-hour asynchronous online CPD course between March 2022 and May 2024 were invited to participate in this mixed methods convergent study. The e-learning course aimed to increase physicians' confidence in approaching colleagues in difficulty by recognizing signs of psychological distress, offering support, and referring them to appropriate resources. Participant characteristics were collected, and behavioral intention to approach a colleague in difficulty along with its determinants were measured pre- and postcourse using the validated CPD-REACTION tool. Differences in mean pre-post intention scores were assessed using 2-tailed paired *t* tests (*n*=466) and generalized estimating equations. Factors associated with postcourse intention were examined using multivariate analysis (*n*=466). Four months later, the proportion of physicians reporting adoption of the behavior was calculated (*n*=61). Qualitative responses to open-ended questions were analyzed thematically using behavior change models, and behavior change techniques used in the course were identified. Quantitative and qualitative results were triangulated. We reported results following STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) and SRQR (Standards for Reporting Qualitative Research) guidelines for quantitative and qualitative analyses, respectively.

Results: Among 792 participating physicians, 466 (58.8%) completed online questionnaires pre- and postcourse. The average participant age was 48 (SD 12.4) years; 43.5% (332/762) were women, and 86% (655/762) were specialists. The average precourse intention score was 3.88 (SD 1.73) and average postcourse intention score was 4.92 (SD 1.40), for an adjusted mean difference of 1.06 (95% CI 0.93-1.20; *P*<.001). Factors associated with postcourse intention were beliefs about capabilities ($\beta=0.52$; *P*<.001), social influences ($\beta=0.27$; *P*<.001), and moral norm ($\beta=0.26$; *P*=.03; $R^2=0.22$). Four months later, 41% (25/61; 95% CI 28.6%-54.3%) of participants reported having approached a colleague in difficulty. Frequently reported reasons for intention to

adopt behavior were beliefs about capabilities, beliefs about consequences, and knowledge. Quantitative and qualitative results converged on beliefs about capabilities but diverged regarding beliefs about consequences. A total of 7 behavioral change techniques were identified in the CPD course: goal setting, increasing competence, planning, persuasive communication, behavior-related information, modeling, and behavioral experiments.

Conclusions: This online CPD course increased physicians' intention to approach a colleague in difficulty. The results highlight beliefs about capabilities as a key determinant of this behavioral intention. The study suggests that online learning has strong potential to raise awareness about peer support and ultimately build a culture of care among health care workers.

(*JMIR Med Educ* 2026;12:e80199) doi:10.2196/80199

KEYWORDS

burnout; continuing medical education; continuing professional development; distance education; helping behavior; occupational stress; peer support; physicians; psychological well-being; social support

Introduction

Burnout is prevalent among physicians [1-4]. In 2021, the Canadian Medical Association reported that half of Canadian physicians experienced at least 1 symptom of burnout on a weekly basis or more frequently [1]. Physicians experience greater social isolation compared to workers in other sectors [5]. Social isolation is linked to higher levels of burnout in physicians, but the direction of this relationship—whether it is a cause, a consequence, or bidirectional—remains unclear [5]. Furthermore, the “second victim” phenomenon, or health care providers’ distress after a patient adverse event, can lead to burnout [6,7]. A patient adverse event is an unexpected but often preventable event which results in consequences for the patient of varying seriousness, including prolonging treatment, causing discomfort, disability, or death [7]. Moral distress was also associated with burnout in health care workers during the COVID-19 pandemic [8]. Moral distress is experienced when a health care worker “feels unable to act in accordance with core values and obligations, or attempted actions fail to achieve the desired outcome” [9]. Burnout impacts physicians’ health and can have consequences on patient safety, satisfaction, and physician retention [10-12]. Improving physician well-being is important and the well-being of health care and social service workers is one element of the Quintuple Aim framework, whose goals are to improve population health, enhance care experiences, reduce costs, improve clinician well-being, and improve health equity [13].

A 2022 scoping review showed that peer support, informally or formally provided by health care providers with similar work conditions and experiences, can help health care providers after a distressing patient adverse event [14]. Furthermore, physicians themselves appear to find it the best form of support [15,16]. A 2024 trial found a peer support program protected health care providers aged 30 years or younger from psychological distress, although the effect was not significant for all age categories combined [17]. In another study, a peer support program reduced psychological distress among anesthesia professionals [18].

Components of a peer support conversation include initial outreach, listening, reflecting, reframing, eliciting the peer’s personal coping mechanisms, closing the discussion, and referring the peer to available resources [19]. Experts recommend that peer supporters should receive training on listening skills, psychological first aid, and information about

referral options to fulfill their role adequately [20]. Many formal peer-support programs have been implemented in health care organizations and include dedicated training courses for peer supporters [15,17,21-49]. Some of these training programs are inspired by (1) the Stress First Aid model, initially developed for high-risk occupations such as the military service [45,47-49]; (2) the psychological first aid training developed by George Everly [22,44]; (3) the peer-support approach developed by Shapiro and Galowitz [27,32,35,37,39] and (4) the Scott Three-Tier Intervention Model of peer support [15,21,23,28,30,41,42]. However, evaluation of the effect of courses about peer support is often omitted in program evaluations or assessed using cross-sectional surveys with nonvalidated measures [15,22,26,27,30,34,38-40,43]. Furthermore, most courses studied are in-person workshops rather than online continuing professional development (CPD) courses [15,22,26,27,30,34,38-40,43]. Therefore, we sought to assess the effect of an online CPD course on the intention of physicians to approach a colleague in difficulty using a mixed methods convergent research design and validated measures. We also analyzed why participants considered approaching a colleague in difficulty by conducting thematic analyses of qualitative data and triangulating these findings with quantitative results.

Methods

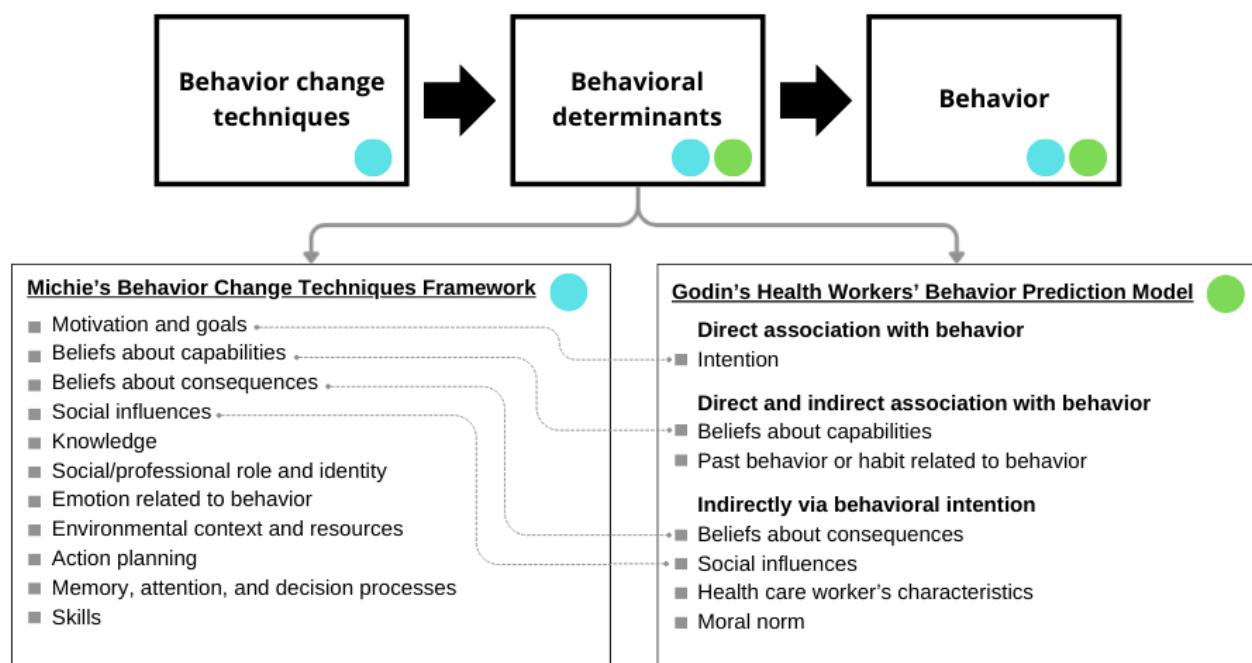
Study Design

We used a mixed methods convergent research design to triangulate results obtained from quantitative and qualitative methods [50]. Quantitative analysis involved a quasi-experimental before-and-after study design to evaluate the effect of an online CPD course that is part of the CPD program of the Fédération des médecins spécialistes du Québec (FMSQ). In this design, measurements before the CPD course served as comparators for the same measurements taken after the CPD course for the same participants [50]. We analyzed responses to open questions with a qualitative thematic deductive approach to understand why participants considered integrating the targeted behavior into their professional practice [50]. For both quantitative and qualitative analyses, we used Godin’s integrated model for predicting health professionals’ behavior as a conceptual framework [51] and we used Michie’s conceptual framework on behavior change techniques (BCTs) to explore our qualitative data in detail [52]. As shown in Figure

1 [52,53], these frameworks share the concept that behavioral determinants lead to behavior adoption [51-53]. Both originate from Ajzen's Theory of Planned Behavior and have many behavioral determinants in common [51-53]. Michie et al [52] propose that BCTs can influence and target behavioral determinants, which in turn translate into behavior change. Possible behavioral change techniques include planning, persuasive communication and others [52]. Details on these 2

conceptual frameworks are available in [Multimedia Appendix 1](#) [51-54]. We reported results following the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guideline for quantitative results ([Multimedia Appendix 2](#)) [55] and used the SRQR (Standards for Reporting Qualitative Research) for qualitative results ([Multimedia Appendix 3](#)) [56].

Figure 1. Framework combining Godin's integrated model for predicting health professionals' behavior and Michie's framework on behavior change techniques [52,53].



Participants and Recruitment

The study took place in Quebec, Canada, from March 1, 2022, to June 20, 2024. Targeted participants were mainly specialists from the FMSQ membership. However, general practitioners who subscribe to the FMSQ's online learning platform MÉDUSE could also participate in the study. The study population was physicians who registered for the course and completed an online consent form. After registering, participants were informed about the study by a system-generated message displayed at the beginning of the online course. The CPD course was not mandatory but was recognized by the Royal College of Physicians and Surgeons of Canada as a 1-hour group learning credit (Section 1) [57]. This convenience sampling was based on time constraints and access to MÉDUSE. Inclusion criteria were being a licensed physician in Quebec and participating in the online CPD course. There were no exclusion criteria.

Intervention

We describe the intervention according to the Template for Intervention Description and Replication (TIDieR) guidelines [58]. The intervention was a 1-hour asynchronous online CPD course developed by the FMSQ to help physicians approach colleagues in difficulty (eg, burnout, psychological distress, second victim symptoms). Thus, this course focused on a form of peer support. The term "colleague" included anyone in the

health and social services sector, including administrative staff. The scientific committee comprised 3 physicians, 2 of whom had peer support experience through their work with the Québec Physicians' Health Program (QPHP). The QPHP organization also helped with course design. A scenario-based video, filmed with professional actors, illustrates the QPHP's 5-Step Caring Approach: (1) becoming aware of context and stressors, (2) recognizing the warning signs, (3) initiating a conversation with the colleague, (4) providing compassionate support, and (5) respecting your own limits and practicing self-care [59]. Approaching a colleague in difficulty corresponds to Tier 1 of the Scott Three-Tier Model of Support, which includes immediate team members of a health care worker experiencing distress [21].

A CPD expert with more than 10 years of experience (MT) used Godin's integrated model for predicting health professionals' behavior [51] for instructional design to increase participants' willingness to approach a colleague in difficulty. For instance, the model was used to increase participants' beliefs in the consequences of approaching a colleague in difficulty. In the first module of the course, viewers are shown how a compassionate approach, which enables a colleague to share emotions related to stressful situations, can help mitigate the impact of these sources of stress. In the second module, physicians are shown the negative consequences that may occur if a colleague in difficulty is not helped or if help is delayed.

Participants completed an intervention plan during the CPD course to use when approaching a colleague at their workplace. Other educational strategies include narrated vignettes, knowledge-check questions, additional readings, and reflective writing, which often refers back to the scenario shown in the video. During the online course, physicians are also provided with a self-evaluation tool to identify when they or a colleague may need help. The tool provides individual strategies to maintain or improve mental health at each step along the continuum from being healthy to having important mental health problems [60].

During the course, participants have the option of downloading 10 summaries (ranging from 1 to 3 pages) of course content or supplementary materials [59-67]. Bloom's taxonomy defined the course's learning objectives [68]. The CanMEDS role targeted by this course was the professional role, specifically with the objective of emphasizing the commitment to physicians' health and well-being, which ultimately promotes optimal patient care [69]. The CanMEDS framework highlights skills that are essential throughout medical practice, providing guidance for medical education and ongoing professional development for practicing physicians, all within a competency-based approach [69]. The course was made available on MÉDUSE, the FMSQ's online CPD platform [70]. Physicians completing the course were sent an automated reminder after 2 weeks of inactivity, and the unfinished course remained available on the physicians' dashboard on MÉDUSE as a reminder to complete it.

Outcomes

The primary outcome was the intention of physicians to approach a colleague in difficulty. Godin defined intention, from the work of Ajzen and Fishbein on the Theory of Reasoned Action, as the motivation to adopt a specific behavior [51]. In

this study, this targeted behavior was approaching a colleague in difficulty.

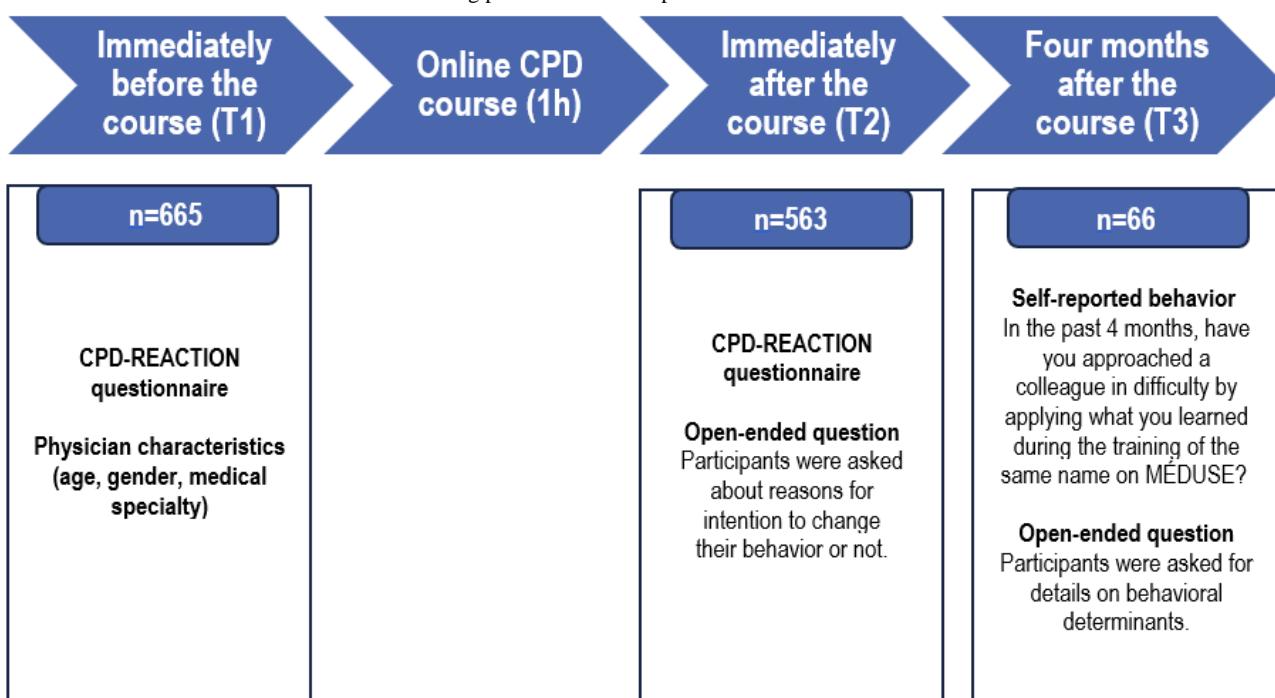
Intention to adopt a behavior is a precursor of behavior adoption in the integrated model for predicting health professionals' behavior [51]. Although intention is not a perfect proxy for future behavior, directly observing peer support behavior among health care workers is neither feasible nor ethical, as these conversations are sensitive and confidential [71]. Using self-reported behavioral intention as a proxy allowed for a practical and more timely evaluation of the CPD course [71]. Self-reported behavior 4 months after course completion was also measured but not used as a primary outcome because there is often a loss to follow-up when measures are not taken soon after course completion [72].

Data Sources and Measurements

Overview

The FMSQ collected physicians' responses on the MÉDUSE online learning platform from March 1, 2022, to June 20, 2024. Data were obtained from three sources: (1) quantitative data, (2) qualitative data, and (3) course content from the MÉDUSE platform (Figure 2). Quantitative data and qualitative data were collected using a questionnaire. The original version administered in French is available in [Multimedia Appendix 4](#) [73,74] and the English translation is available in [Multimedia Appendix 5](#) [73,74]. Physicians completed these online self-administered questionnaires at three timepoints: immediately before the course (T1), immediately after the course (T2), and 4 months after the course (T3) (Figure 2). Questionnaires at T1 and at T2 were administered by a system-generated message displayed at the beginning of the online course, and the questionnaire at T3 was sent automatically by email by the MÉDUSE online platform.

Figure 2. Data sources and collection. CPD: continuing professional development.



Quantitative Data Sources and Measurements

Before the CPD course, we collected data on the age, gender, and medical specialty of participants using an online questionnaire developed by the FMSQ. Before and immediately after the course (T1 and T2), we used the French version of CPD-REACTION to measure physicians' intention to approach a colleague in difficulty (dependent variable) and the determinants of this intention. Finally, we measured their self-reported behavior of approaching a colleague in difficulty 4 months after course completion (T3).

The CPD-REACTION is a measuring tool whose constructs were developed through a systematic review and whose items were validated using factor analysis [73]. Based on Godin's integrated model for predicting health professionals' behavior, the instrument measures intention to adopt a behavior and 4 determinants of intention [53,54,73]. It comprises 12 items grouped into 5 constructs: intention to adopt a behavior (2 items) and four determinants of intention: (1) moral norm (2 items), (2) beliefs about capabilities (3 items), (3) beliefs about consequences (2 items), and (4) social influence (3 items) [73]. Definitions of these determinants are available in [Multimedia Appendix 1](#). Each construct is determined by averaging 2 or 3 items from the instrument [73]. The Cronbach α coefficients for each construct range from 0.79 to 0.89 [73], indicating good reliability [75].

Qualitative Data Sources

Overall, participants were asked 8 open-ended questions. Through 2 open-ended questions, physicians were asked about their reasons for intending to adopt the new behavior as well as factors influencing nonadoption of the behavior 4 months later. The remaining 6 questions, which assessed participants' appreciation of the CPD course, fall outside the scope of this work.

Course Content in the MÉDUSE platform

Two reviewers (F Lizotte and EL) were given access to all course content described in the intervention section above. The methodology for the analysis of this data is detailed in the section title "Identifying Behavioral Change Techniques Present in the Course."

Data Analysis

For both quantitative and qualitative analyses, we used Godin's integrated model for predicting health professionals' behavior as a conceptual framework [51]. In addition, we used Michie's conceptual framework on BCTs to analyze qualitative responses in detail as well as course content [52]. We analyzed all quantitative data with SAS software (version 9.4; SAS Institute Inc). We performed verification with nonparametric tests using R software (R Foundation for Statistical Computing). The NVivo 15 software (Lumivero) was used for qualitative analyses.

Participants' Intention and Its Determinants to Approach a Colleague in Difficulty, Pre-, and Postcourse

First, we performed descriptive statistics for all variables, including participant characteristics. Then, we calculated the difference between the pre- and postintention mean for each participant using 2-tailed paired *t* tests, assuming normality

given our large sample size [76]. To confirm the robustness of the results, we also calculated the statistical significance of intention differences using a nonparametric Wilcoxon signed-rank test [77]. We did the same for each of the 4 determinants of intention to adopt the new behavior (social influences, beliefs about capabilities, beliefs about consequences, and moral norm). Finally, we verified that the results did not differ according to participants' survey response pattern or participant characteristics (gender, age, and domain of practice). We adopted the age categories used by the Canadian Medical Association in its survey on the wellness of Canadian physicians, as these categories reflect different stages of a medical career [78]. To describe effect size, we calculated Cohen d_z for within-subjects designs by dividing the mean of the within-person differences with the SD of those differences [79].

Estimation of the Effect of the CPD Course on the Intention to Approach a Colleague in Difficulty

We used generalized estimating equations (GEE) with a Gaussian distribution and identity link to estimate the marginal effect of the training program on intention scores over time (pre- vs postcourse [80]). The model included time (pre or post) as the main predictor, with an exchangeable working correlation structure to account for within-individual correlation, and was adjusted for age, sex, domain of practice, and baseline determinants of intention. We included age as a categorical variable since that variable did not respect the postulate of linearity.

Predicting Intention Postcourse

To predict intention to approach a colleague in difficulty after the course, we built a multiple linear regression model [81]. Behavioral intention was the dependent variable, while the 4 determinants of this intention and 3 sociodemographic characteristics (age, gender, and domain of practice) were independent variables. A full model including all candidate variables from Godin's framework was first fitted [51]. Using the backward elimination method, variables with a coefficient *P* value greater than .05 were successively eliminated for a more parsimonious model [82]. Sensitivity analyses were performed at critical α values of .10 or .20 to verify whether different choices changed the predictive performance of our model or the variables included in the final model. We presented the full and the fitted models.

Intention and its determinants were measured as aggregate multiple ordered responses of Likert items from 1 to 7 [73]. Thus, these composite scores approximate continuous latent variables and are commonly analyzed with linear models [83,84]. Simulation and methodological work have shown that treating such multi-item Likert scales (with ≥ 5 response options) as continuous yields valid estimates and tests in reasonably sized samples [83,84]. We checked for collinearity, normality of residuals, homoscedasticity, and linearity of residuals.

Power Analysis

Most participants were already recruited, so the sample size could not be changed. Therefore, we performed a post hoc power analysis. Based on 2 studies examining the effect of CPD courses on physicians' intentions to adopt new behaviors, we

should have been able to detect a mean difference of at least 0.5 in intention before vs after the course [72,85]. The statistical power of our paired *t* test for the primary outcome (mean difference in intention) was 99.9% at a significance level of $\alpha=.05$ to detect this mean difference of at least 0.5. This indicates that the sample size used in this study was sufficiently large to detect a meaningful effect. We analyzed all quantitative data with SAS software (version 9.4; SAS Institute Inc). We performed verification with nonparametric tests using R software (R Foundation for Statistical Computing).

Physicians' Reasons for Intention to Adopt the New Behavior From Qualitative Data

Through 2 open-ended questions, physicians were asked about reasons for intention to adopt the new behavior as well as factors influencing nonadoption of the behavior 4 months later. The answers were analyzed using a thematic deductive approach [50]. We created an initial codebook by integrating 2 conceptual frameworks: Godin's integrated model for predicting health professionals' behavior [51] and Michie's BCTs with their respective behavioral domains [52]. Two researchers (F Lizotte and EL) organized 10% of the data into themes. Emergent themes were then discussed and added inductively to the framework. We documented these choices in a journal held by F Lizotte. The 2 reviewers (F Lizotte and EL) then coded data independently for greater credibility of the results and compared their analyses [50,86]. During qualitative analysis, F Lizotte, a medical resident, reflected critically about certain themes with EL, who has a background in psychology, and RCC with training in sociology, to improve the trustworthiness of the results. EL was completely blinded to the quantitative analysis process. To obtain consensus on results, any coding differences between the 2 reviewers (F Lizotte and EL) were resolved through discussion with RCC, whose main expertise was qualitative analysis. We described intercoder reliability of thematic analysis of open-ended questions using percentage agreement [87].

Triangulating Quantitative and Qualitative Data

We triangulated quantitative and qualitative data to enhance the credibility and validity of the research findings and offer practical theory-based recommendations for CPD course development on peer support [50,88]. We compared 4 determinants of intention, measured with CPD-REACTION and obtained using quantitative methods, with the reasons to adopt the new behavior obtained using qualitative methods [50,88]. We observed where results converged, diverged, or provided additional information [88].

Self-Reported Behavior Adoption 4 Months After CPD Course

We calculated the proportion of physicians who self-reported approaching a colleague in difficulty 4 months after the online CPD course. We compared the mean intention of those who adopted the behavior with those who did not using Wilcoxon rank tests because of the small sample sizes of the data at T3 and verified the results using paired *t* tests.

Identifying Behavioral Change Techniques Present in the Course

We developed an observation grid ([Multimedia Appendix 6](#) [51,89,90]) from the conceptual framework by Michie et al [52]. Using this grid, 2 independent reviewers (F Lizotte and EL) identified BCTs present in the online CPD course. We analyzed whether the behavioral domains targeted by the BCTs identified in the course were the same as the determinants of intention that had significantly changed after the CPD course. This analysis explored which components of the course might be the active ingredients in changing physicians' behavioral intentions. Using the entirety of the interrater data, we calculated Cohen kappa to describe interrater reliability for BCT identification [91]. We chose Cohen kappa as the interrater reliability statistic, as it has the advantage of taking into account agreement by chance and can be applied to categorial nominal variables of the observation grid [87].

Handling of Missing Data

Item-level missing data were handled by following the CPD-REACTION Questionnaire User Manuel [74]. A nonresponse to 1 of the 2 items measuring a construct of the CPD-REACTION resulted in missing data for the observation of the measured variable [74]. A missing response to an item in the 3-item measure of the determinants of intention was handled with mean imputation [74]. Only 0.25% of responses to items on the CPD-REACTION questionnaire administered at T1 were missing, with item-level missingness ranging from 0.00% to 0.90% ([Multimedia Appendix 7](#)). Similarly, approximately 0.27% of responses were missing for items on the CPD-REACTION questionnaire administered at T2, with a range of 0.00% to 1.07% depending on the item ([Multimedia Appendix 7](#)). There were no missing sociodemographic characteristics of physicians who completed the CPD-REACTION questionnaire before and after the CPD course.

For wave-level missing data, the 466 participants with data at T1 and T2 were included in the analysis (ie, perfect sets). Sensitivity analyses were conducted to verify if results differed if all data, regardless of data pairing, were analyzed. A complete case analysis was performed in all models.

Ethical Considerations

We obtained ethical exemption from a hospital research ethics committee (Comité d'éthique de la recherche sectorielle en santé des populations et première ligne—Centre intégré universitaire de santé et de services sociaux de la Capitale-Nationale [CIUSSCN]; project #2025-3127). The research project participants had given their free and informed consent to take part in the study. Before each questionnaire, a consent form detailed the objective and intended use of the collected data for research ([Multimedia Appendix 4](#) contains the French version and [Multimedia Appendix 5](#) contains the English version). The database was deidentified to protect the confidentiality of participants. A password was required to access the data saved in an Excel file, which has been stored on a secure server. Data analysis was carried out on a CIUSSCN computer that was also password-protected. Participants experiencing emotional

difficulties could speak with peer supporters from the QPHP. Physicians volunteered to participate and received no financial compensation.

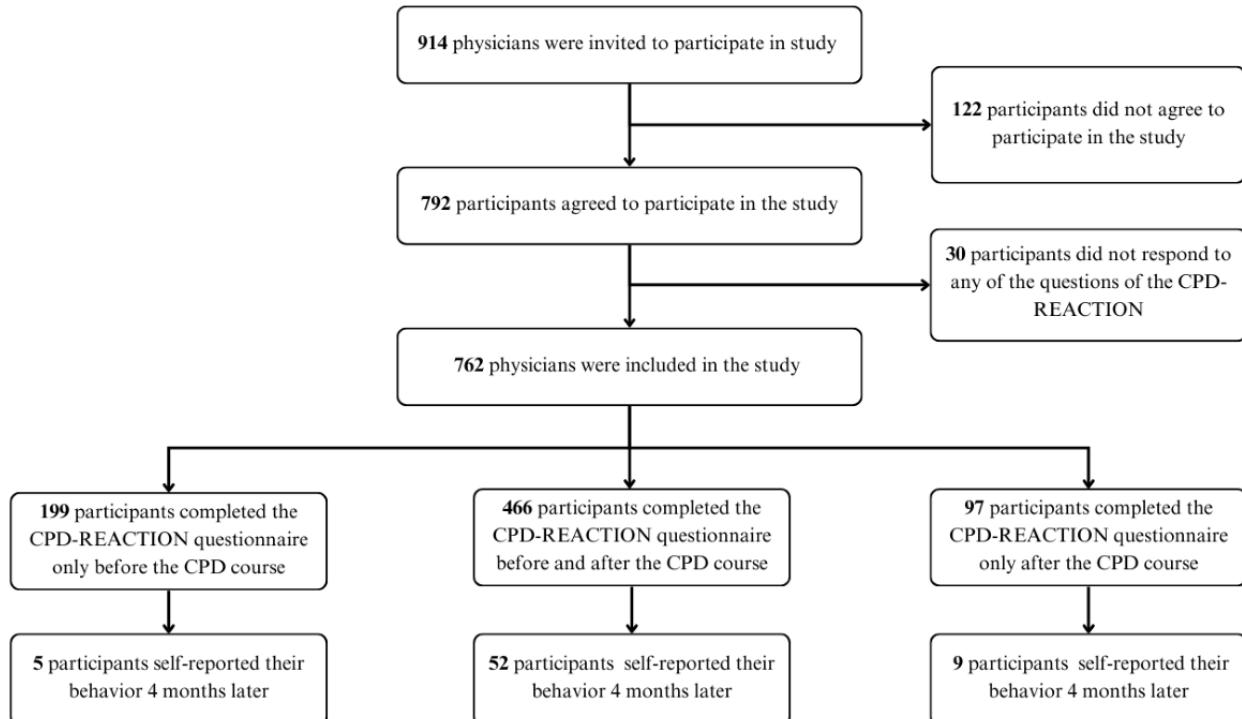
Results

Participant Flow and Characteristics

During the 28-month study period, 914 physicians were invited, of whom 792 (87%) agreed to participate (Figure 3). A total of 30 participants were excluded because they did not answer a

single CPD-REACTION questionnaire item. We performed analyses on 61% (466/762) of participants who completed the CPD-REACTION questionnaires before and after the course (T1 and T2). Finally, the frequency of self-reported behavior was calculated on 8% (61/762) of participants who completed both the self-reported behavior change questionnaire 4 months after the course (T3) and the CPD-REACTION questionnaire immediately after the CPD course (T2). We excluded 5 participants who did not complete the postcourse survey (T2) from the analysis of data at T3, as it was only at T2 that we could verify their completion of the course.

Figure 3. Study flowchart. CPD: continuing professional development.



Among the participants, the woman-to-man ratio was approximately 1:1, and the majority of physicians (655/762, or 86%) were specialists (Table 1). The average age of participants was 48 (SD 12.4) years, 43.5% (332/762) were women, and 86% (655/762) practiced in a specialty. Physicians who

responded both before and after the course (n=466) had similar sociodemographic characteristics. We did not have sociodemographic characteristics for the 97 physicians who responded only to the T2 questionnaire without responding to the T1 questionnaire.

Table 1. Characteristics of participants.

Sociodemographic characteristics of physicians	Value (n=762)	Participants who completed the CPD-REACTION questionnaire before and after the CPD ^a course (n=466)
Gender, n (%)		
Woman	332 (43.5)	241 (51.7)
Man	325 (42.7)	222 (47.6)
Other, nonbinary, or prefer not to disclose	8 (1.1)	3 (0.7)
Missing	97 (12.7)	0 (0)
Domain of practice, n (%)		
Surgical specialty	293 (38.5)	208 (44.7)
Laboratory specialty	102 (13.4)	71 (15.2)
Medical specialty	246 (32.3)	169 (36.3)
General practitioner	10 (1.3)	9 (1.9)
Nonspecified specialty	14 (1.8)	9 (1.9)
Missing	97 (12.7)	0 (0)
Age (in years), n; mean (SD)	665; 48.1 (12.4)	466; 48.6 (12.3)

^aCPD: continuing professional development.

Participants' Intention (and its Determinants) to Approach a Colleague in Difficulty, Pre- and Postcourse

Intention means increased from 3.88 (SD 1.73) before to 4.92 (SD 1.40) after the course, as measured on a scale of 1 to 7. This mean difference increase of 1.04 ($P<.001$) is statistically significant (Table 2). This 1-point increase was also clinically significant, with a Cohen d_z value of 0.72 representing a moderate effect size. Bivariate analyses of changes in the 4 determinants of intention (social influences, beliefs about capabilities, beliefs about consequences, and moral norm) were all statistically significant at an α level of .05. Determinants of

intention (beliefs about capabilities, social influences, and beliefs about consequences) increased by statistically significant mean differences ranging from 0.14 to 0.85. Moral norm decreased by a statistically significant mean difference of 0.30 ($P<.001$). Lower extreme values after the CPD course did not explain the decrease in moral norm. The effect size on the increase in beliefs in capacities and on the increase in social influences was moderate. The effect size on the decrease in moral norm was small, and the effect size on the increase in beliefs about consequences was not clinically significant (Table 2). Mean intention and determinants of intention were similar whether participants had responded to both questionnaires (T1 and T2) or to only 1 questionnaire (Table S2 in [Multimedia Appendix 8](#) [73,74]).

Table 2. Comparison of behavioral intention before and after the continuing professional development (CPD) course.

CPD-REACTION constructs ^a	Frequency, n	Before CPD course, mean (SD)	After CPD course, mean (SD)	Before vs after CPD course			
				Mean difference (CSD ^b)	95% CI	P value ^c	Cohen d_z ^d
Intention	455	3.88 (1.72)	4.92 (1.40)	1.04 (1.45)	0.91 to 1.17	<.001	.72
Beliefs about capabilities	459	4.65 (1.17)	5.49 (0.74)	0.85 (1.11)	0.74 to 0.95	<.001	.76
Moral norm	462	6.09 (1.04)	5.79 (0.77)	-0.30 (1.06)	-0.40 to -0.21	<.001	.29
Social influences	464	3.36 (0.99)	3.81 (0.99)	0.45 (0.74)	0.38 to 0.52	<.001	.61
Beliefs about consequences	465	5.53 (1.04)	5.67 (0.65)	0.14 (0.90)	0.06 to 0.22	<.001	.16

^aResponse scale of the CPD-REACTION questionnaire ranged from 1 to 7 [73,74].

^bCSD: combined SD.

^cP values were calculated using 2-tailed paired t test.

^dCohen d_z estimates effect size (<0.2=not significant, 0.2-0.5=small, 0.5 to 0.8 medium, >0.8=large) [91].

Subgroup Analyses of Principal Outcome: Difference of Intention to Approach a Colleague in Difficulty

The increase in the mean difference of intention was approximately 1 point and was statistically significant for all categories of gender, age, and domain of practice (laboratory, surgical, and medical domains) that included at least 10 participants ([Multimedia Appendix 9](#)). We did not interpret subcategories containing fewer than 10 participants (eg, general practitioners). Results of the multiple linear regression model confirmed that the intention difference was not modified by gender, age, or domain of practice when all variables were included in the same model ([Multimedia Appendix 9](#)).

Effect of the CPD Course on the Intention to Approach a Colleague in Difficulty

In the GEE model, participation in the CPD course was associated with a higher intention to approach a colleague in difficulty. On average, intention scores were 1.06 points higher in participants after the CPD course compared to before the CPD course ($\beta=1.06$; $P<.001$), after adjustment for determinants and for sociodemographic characteristics (Table S1 in

[Multimedia Appendix 8](#)). Beliefs about capabilities ($\beta=0.37$; $P<.001$), social influence ($\beta=0.22$; $P=.002$), and beliefs about consequences ($\beta=0.15$; $P=.03$) before the CPD course showed significant positive associations with mean intention, whereas moral norm was not significant. No sociodemographic characteristics were independently associated with intention to approach a colleague in difficulty (Table S1 in [Multimedia Appendix 8](#)).

Postcourse Determinants of Intention

In the final multivariate regression analysis model, beliefs about capabilities ($\beta=0.52$; $P<.001$), social influences ($\beta=0.27$; $P<.001$), and moral norm ($\beta=0.26$; $P<.03$) predicted physicians' intention to approach a colleague in difficulty ([Table 3](#)). Our model accounted for 22% of the variance of intention after the course. Sensitivity analyses verified the stability of the other variables and showed consistent results if we also included physicians who had not responded to the precourse questionnaire ([Multimedia Appendix 10](#)), except for the variable belief in consequences, whose statistical significance changed when we modeled physicians who had not responded to the precourse questionnaire.

Table 3. Predictive factors of physicians' intention to approach a colleague experiencing difficulties.

Variables ^a	Full multivariate model ^b (n=457)		Fitted multivariate model ^c (n=457)	
	β (95% CI)	P value	β (95% CI)	P value
Beliefs about capabilities	0.39 (0.12 to 0.67)	.005	0.52 (0.28 to 0.75)	<.001
Social influence	0.25 (0.12 to 0.37)	<.001	0.27 (0.15 to 0.39)	<.001
Moral norm	0.26 (0.03 to 0.49)	.02	0.26 (0.03 to 0.48)	.03
Beliefs about consequences	0.22 (-0.01 to 0.45)	.06	N/A ^d	N/A
Age (in years; reference: 35-54)		.39 ^e		
<35	-0.22 (-0.56 to 0.13)	.21	N/A	N/A
>54	-0.11 (-0.37 to 0.14)	.38	N/A	N/A
Domain (reference: surgical)		.25 ^e		
Laboratory	0.23 (-0.12 to 0.57)	.20	N/A	N/A
Medical	0.29 (0.03 to 0.55)	.03	N/A	N/A
Family medicine	0.39 (-0.45 to 1.23)	.36	N/A	N/A
Other	0.30 (-0.54 to 1.13)	.48	N/A	N/A
Gender (reference: women)		.54 ^e		
Men	-0.12 (-0.37 to 0.12)	.33	N/A	N/A
Other, nonbinary, or prefer not to disclose	-0.44 (-1.86 to 0.97)	.54	N/A	N/A

^aWe included psychosocial variables measured after the course.

^b $R^2=0.24$.

^c $R^2=0.22$.

^dN/A: not applicable, as variables with a coefficient P value greater than .05 were successively removed using the backward elimination method.

^eGlobal P value.

Postulates of independence of observations and homoscedasticity were respected. Postulates of normality and linearity of residuals were not respected. We verified the correlations between the independent variables and found they

did not explain our model's low predictive capacity ([Multimedia Appendix 10](#)).

Physicians' Reasons for Intention to Adopt the New Behavior From Qualitative Data

Of the 251 open-ended responses related to behavioral intentions (T2), 153 verbatims expressed clear behavioral intentions (eg, “I will help my peer”), 10 showed hesitation (eg, “I will try to”), 89 described reasons for intention to approach a colleague in difficulty (eg, “I realize more fully my ability to act”), and 8 referred to other related behaviors. The above categories are not mutually exclusive. Physicians most frequently intended to

(1) be more attentive to their colleagues' signs of distress, (2) approach a colleague with kindness or help or support a colleague, and (3) be more available or listen more to colleagues ([Multimedia Appendix 11](#)). The reasons for intention to change behavior physicians most frequently cited were beliefs about capabilities, beliefs about consequences, and knowledge of the behavior ([Table 4](#)). For each of the 10 categories of reasons for wanting to adopt the new behavior, intercoder percentage agreement on thematic analysis varied between 98.1% and 99.5%.

Table 4. Reasons for intention to adopt new behavior of approaching a colleague in difficulty.

Reasons for intention to adopt new behavior	Representative excerpts of verbatim	Agreement, n (%)
Beliefs about capabilities	<ul style="list-style-type: none"> “I realize more fully my ability to act.” “I have more tools to approach a colleague in difficulty.” “I feel more comfortable doing it.” 	41 (98.4)
Beliefs about consequences	<ul style="list-style-type: none"> “This limits the consequences for the colleague and the team.” “To improve the collegiality of my department.” “I think that the current context of healthcare depends on the support we provide to our colleagues, so it is essential to do it.” 	16 (99.2)
Knowledge	<ul style="list-style-type: none"> “Better knowledge of how to detect a colleague's distress and how to intervene in such cases.” “Better understanding of the approach.” “I will better assess where my colleague stands on the distress scale.” 	11 (96)
Social/professional role and identity	<ul style="list-style-type: none"> “Necessary with my hospital duties.” “Feeling more legitimized to approach a colleague in difficulty.” “Better understanding and responsibility of collegiality.” 	6 (98.5)
Moral norm	<ul style="list-style-type: none"> “I like the idea that doing nothing is not an option.” “Not acting is not an option.” “...understanding that some approaches may be better than others, but anything is better than doing nothing.” 	5 (97.9)
Past behavior or habit related to behavior	<ul style="list-style-type: none"> “It has provided me with new tools to continue supporting colleagues in difficulty.” “I already apply many similar principles.” “Continue to approach my colleague when they seem to be having difficulties.” 	4 (99)
Emotion related to behavior	<ul style="list-style-type: none"> “Being more willing to get uncomfortable.” “It reassures me.” “I feel more equipped and reassured about my reflections.” 	3 (99.5)
Environmental context and resources	<ul style="list-style-type: none"> “Colleague to help.” “Frequent problem.” 	2 (98.9)
Action planning	<ul style="list-style-type: none"> “Do not procrastinate.” “Do not hesitate to create opportunities outside of work.” 	2 (98.1)
Memory, attention, and decision processes	<ul style="list-style-type: none"> “Because I realize that often, I get too absorbed in my professional or personal concerns to look around me. Being a witness allows for perspective and a new outlook.” 	1 (99)

The reasons physicians gave for not wanting to change their practice were mostly related to their past behavior, for example, that they were already helping a peer and thus had already adopted the behavior ([Multimedia Appendix 11](#)). Other reasons for not intending to change behavior were their environmental context (eg, “I'm now in solo practice”) and that helping a peer was not relevant to their particular medical practice. Four months after the CPD course, the behavioral determinants most

frequently given for not approaching a colleague in difficulty were having had no occasion to adopt the behavior, having no colleagues in difficulty, and not being actively in practice or not being in full-time practice ([Multimedia Appendix 11](#)).

Triangulated Quantitative and Qualitative Data

Qualitative and quantitative results showed most physicians were motivated to approach a colleague in difficulty following

the online CPD course. Both qualitative and quantitative results pointed to beliefs about capabilities as being an important determinant of intention. Beliefs about consequences was an important theme in the qualitative data, but it was not an important determinant of postcourse intention to approach a colleague in difficulty. Sensitivity analysis of the 251 physicians who provided open-ended responses related to behavioral intentions (T2) also showed a divergence regarding beliefs in consequences between qualitative and quantitative results.

Self-Reported Behavior Adoption 4 Months After CPD Course

Among the 66 physicians who completed the self-administered questionnaire at T3, 61 had completed CPD-REACTION immediately after the course. Of these, 41% (25/61; 95% CI 28.6%-54.3%) reported approaching a colleague in difficulty in the 4 months following the CPD course. Those who adopted the behavior had a mean intention after the CPD course of 5.40

(SD 1.01), and those who did not adopt the behavior had a mean intention of 5.00 (SD 1.40). This 0.4 (SD 1.26; $P=.24$) mean behavioral intention difference is not statistically significant.

Behavioral Change Techniques Identified as Present in the Course

A total of 7 BCTs were found in the online CPD course: goal setting, increasing competence, planning, persuasive communication, behavior-related information, modeling, and behavioral experiments (Table 5). Cohen kappa measuring interrater reliability was 0.53 (95% CI 0.14-0.91) for evaluation of BCTs identified the course, which represents a moderate strength of agreement [92]. Of these 7 BCTs, 4 were related to behavioral intentions, 3 involved competence, 2 involved behavioral regulation, and 2 were related to beliefs about consequences (Multimedia Appendix 11). Beliefs about capabilities, knowledge, social influences, and memory, attention, and decision processes were each targeted by 1 BCT.

Table 5. Behavioral change techniques (BCTs) identified in the continuing professional development (CPD) course.

BCT	Definition	Domains targeted
Goal setting	Clearly indicate the targeted behavior.	Competence, behavioral intention, behavioral regulation
Increasing competence	Enhance skills through problem-solving, decision-making, and goal setting.	Competence, beliefs about capabilities, behavioral intention
Action planning	Identify the components of the behavior and establish a plan to execute each component, including when and where the behavior will be performed (ie, scheduling behaviors).	Memory, attention, and decision processes, behavioral regulation
Persuasive communication	A credible source presents arguments in favor of the behavior. There must be a presentation of arguments; general pro-behavior communication does not qualify.	Beliefs about consequences, behavioral intention
Behavior-related information	Provide information about the antecedents or consequences of the behavior, the connections between them, or relevant BCTs.	Knowledge, beliefs about consequences, behavioral intention
Modeling	Observe the behavior of others as a model for action.	Competence, social influences
Behavioral experiments	Ask individuals to test hypotheses about the behavior, its causes, and its consequences by collecting and interpreting data, including personal experiences.	No consensus on targeted domains

Discussion

Principal Findings

We assessed the effect of an online CPD course on physicians' intention to approach a colleague in difficulty and analyzed why participants considered integrating this peer support approach or not into their practice. Mean behavioral intention increased significantly after the course. Postcourse, significant determinants of intention were social influences, beliefs about capabilities, and moral norm. Most frequently stated reasons for intention to change behavior were beliefs about capabilities, beliefs about consequences, and knowledge of behavior. Triangulation showed an alignment of quantitative and qualitative results on beliefs about capabilities as a key factor but diverged on beliefs about consequences. In the 4 months following the CPD course, 41% (25/61) of physicians reported approaching a colleague in difficulty. There was no significant difference in the intention to approach a colleague in difficulty between physicians who adopted this behavior at the 4-month

follow-up and those who did not. Those who did adopt the behavior had a nonstatistically significant higher intention after the CPD course than those who did not. Finally, we identified 7 BCTs in the online CPD course: goal setting, increasing competence, planning, persuasive communication, behavior-related information, modeling, and behavioral experiments. These results lead us to make the following observations.

Comparison of Participants' Behavioral Intention Pre- and Postcourse With Prior Work

First, we found that, on average, physicians reported a stronger intention to approach a colleague in difficulty after completing the CPD course compared to before. This 1-point intention increase was statistically significant and also clinically significant, having a moderate effect size. This increase was observed in bivariate analysis and confirmed with multivariate analysis. This suggests that the CPD course motivated physicians to approach colleagues in difficulty. Enhancing peer support among physicians through CPD courses is particularly important

as it can play a protective role regarding psychological distress in physicians [17,18].

Also, to the best of our knowledge, this 1-point intention increase represents a greater increase in intention than has been observed in most other studies using the CPD-REACTION to evaluate changes in health professionals' intentions following an intervention [72,85,93-102]. The before-after difference of intention was more than double the difference observed in a study by Ayivi-Vinz et al [85] who evaluated intention before and after 9 online CPD courses taken by Quebec physicians, all on topics other than this one. The 95% CIs of difference in intention between their study and ours do not overlap [85]. Thus, the observed difference between the 2 studies is statistically significant. An important difference is that Godin's integrated model for predicting health professionals' behavior was used by the senior instructional designer (MT) to design the course, while the 9 courses in the study by Ayivi-Vinz et al [85] were developed by physicians who may not have used behavior change models in course design. This may give support to the importance of using theory to produce CPD courses. Another explanation is that preintervention intention among the Quebec physicians in the study by Ayivi-Vinz et al [85] was already high at 5.5 out of 7, while in ours it was 3.9 out of 7, and thus there was more room for improvement. Overall, this result suggests that there are benefits of using behavior change theories, especially for CPD courses targeting behaviors that physicians are initially less inclined to adopt. It also suggests that it is possible to increase behavioral intention in physicians using online training formats.

Change in Determinants of the Intention to Approach a Colleague in Difficulty

Social influences, beliefs about capabilities, and beliefs about consequences increased. An unexpected postcourse decrease in moral norm occurred. Moral norm had the highest mean among the 4 determinants of intention both before and after the intervention, indicating that participants already felt a moral obligation to adopt the targeted behavior and thus there was little room for improvement. Also, the apparent decrease had a small effect size, which may not translate into clinical significance. One possible explanation for this unexpected decrease in moral norm is that participants may have given more nuanced answers after the CPD course as they better understood the ethical and deontological implications of approaching a colleague in difficulty. In the precourse questionnaire, 38.3% (177/462) of the physicians answered with the maximum score of 7/7 for moral norm, compared to none after the intervention. Nevertheless, further qualitative data analysis is needed to explore this result. Analyzing which BCTs target moral norm could also be important, as this is not studied by Michie et al [52].

No Modifying Effect of Age, Gender, or Domain of Practice on Primary Outcome

Second, we also examined whether participant characteristics modified the impact of the course on their behavioral intention. The before-after intention difference was the same regardless of age, gender, and domain of practice (laboratory, surgical, and medical domains). Thus, the course successfully reached

physicians regardless of their sociodemographic characteristics and does not need to be adapted for a specific population. For example, a previous study found that surgeons were the least likely physicians to use peer support [34], hence the importance of our finding that surgeons also responded favorably to the online CPD course.

Comparison of Determinants of Intention Pre- and Postcourse With Prior Work

Third, we found that determinants of intention to approach a colleague in difficulty after the CPD course included social influences, beliefs about capabilities, and moral norm. Beliefs about capabilities was the strongest predictive variable of the intention to approach a colleague in difficulty. This means improving physicians' perception of their self-efficacy in adopting a behavior, as well as their perceived control over performing that behavior, is important to motivate them to approach a colleague in difficulty. Thus, CPD course developers should target beliefs about capabilities through BCTs for peer support training. Beliefs about capabilities was also the most significant determinant of intention in Godin's systematic review on health care workers' behaviors [53]. BCTs that are known to be effective for this include: (1) increasing skills, (2) coping strategies, (3) repetition or practice of the behavior, (4) feedback, (5) planned self-affirmations, (6) motivational interviewing, and (7) graded tasks (starting with tasks that are easy to accomplish and gradually making them more difficult). In previous studies, beliefs about capabilities, moral norm, and beliefs about consequences were all predictive of intention to adopt behaviors related to patient care (eg, "Prescribe opioids in a safe manner when managing perioperative pain in my patients") postcourse [72,85]. Our results differ slightly from these, with social influence also being predictive of this intention. Approaching a colleague in difficulty is a social behavior that can be adopted by a whole departmental team [38]. Our results suggest that the social context may be a more important influence on team-based behaviors than on clinical behaviors, which are associated more with individual clinicians.

Comparing Predictive Capacity Postcourse With Prior Work

Our model had a low predictive capacity ($R^2=0.22$), explaining only 22% of the variance in intention. In comparison, previous literature showed that the predictive capacity of models analyzing 1 single behavior varied between 19% and 62%, depending on the type of behavior [53]. Bakwa et al's [72] study of multiple behaviors related to patient care among Quebec physicians explained 82% of the variance in physicians' intention to adopt the clinical behavior. However, Bakwa et al [72] included multiple behaviors in the same model, while we studied only 1 behavior, and moreover one that is specific to the well-being of physicians rather than geared to a clinical task. Indeed, existing studies mainly predict behaviors related to patient care [53,72]. While variables of Godin's integrated model for predicting health professionals' behavior are sufficient to predict such behaviors [51], approaching a colleague is not a clinical behavior and may be influenced by factors not measured by Godin's model such as lack of time, lack of privacy, and organizational culture [36,103]. This is significant

because the Quintuple Aim—adopted by many health care systems—now emphasizes 5 key outcomes, and not all of them clinical: patient outcomes, patient experience, efficiency, clinician well-being, and equity [13]. Therefore, further development is needed regarding fostering behaviors related to clinician well-being.

Triangulated Quantitative and Qualitative Data

Fourth, triangulating quantitative with qualitative results confirmed that beliefs about capabilities was a key determinant of intention to approach a colleague in difficulty. Interestingly, results regarding beliefs about consequences diverged slightly: while it was the second most frequent reason for intention to change behavior stated in physicians' open-ended responses, it was not a statistically significant determinant of intention in our quantitative results. One explanation is that statistical significance varied according to the other variables included in the model, perhaps because the P value for beliefs about consequences was so close to the critical α value of .05. Another possible explanation is that beliefs about consequences, as described by Godin [51], seems to refer to the consequences for oneself. Items for measuring beliefs about consequences quantitatively were worded as follows "Overall, I think that *for me* approaching a colleague in difficulty would be...." [73]. In contrast, in our qualitative analysis, beliefs about the consequences of adopting the behavior could also refer to its consequences for the colleague being approached, the team, the department, or for society (Table 5). Indeed, many physicians taking part in this study perceived that the consequences of approaching a colleague in difficulty would be beneficial for the team and the department, a theme often mentioned in the open-ended responses. Thus the perception of the consequences of team-focused care may include organizational benefits, team benefits, patient benefits, and benefits to individual team members [104]. Also, approaching a colleague in difficulty is a team-focused behavior that takes place between 2 health care workers, while behaviors related to patient care occur between a clinician and a patient (eg, prescribing medication). We hypothesize that such team-focused behaviors may depend on a wider variety of determinants, such as organizational culture, than do individual behaviors related to patient care [36,104,105]. Whether peer support strategies can be applied without fear of judgment or retaliation within the organization may depend on its climate of psychosocial safety, or the perception among employees that their psychological health is a priority for the management [106,107]. A study by Tolin et al [38] reported that physicians' perception of institutional support had a positive impact on implementation of a physician-focused peer support program. Broadening the item statements on beliefs about consequences to include effects on others and adding questions with regards to role and identity may be a good starting point for improving measuring tools regarding team-based behaviors.

Other determinants to measure in future studies should include social and professional role or identity and past behavior or habits related to the behavior, both mentioned by participants as determinants of intention and present in Godin's integrated model for predicting health professionals' behavior (Table 4) [51]. Our qualitative analysis identified the following further potential determinants: (1) knowledge; (2) action planning; (3)

memory, attention, and decision processes; (4) environmental context and resources; and (5) emotion related to behavior. The first 3 of these determinants are also behavioral domains targeted by the BCTs we identified in the CPD course.

Self-Reported Behavior Adoption 4 Months After CPD Course

Fifth, 4 of 10 physicians reported having approached a colleague in difficulty in the 4 months following the CPD course, and those who did so had a higher intention to adopt the behavior after the CPD course. No statistically significant difference in intention to approach a colleague in difficulty was found between those who did and those who did not adopt the behavior (5.4 vs 5.0; $P=.24$). This unexpected finding is possibly due to the limited statistical power of the analysis, estimated at 23%. Also, environmental factors, such as practicing solo, were the barriers most mentioned by physicians. Some of these physicians who might have been willing to approach a colleague in difficulty were simply not in the right environment to be able to do so. Other studies report on peer support courses that are part of comprehensive peer support programs that include an organization-wide referral system [15,22,26,27,30,34,38-40,43]. Thus, a physician could be the only physician in a department and still participate in a peer support program thanks to the referral system. It would be valuable to further evaluate whether formalizing peer support through these programs helps overcome certain environmental barriers or, conversely, introduces new barriers, compared to informal peer support outside of a program, as occurred in our study.

Behavioral Change Techniques Present in the Course

Lastly, we identified 7 BCTs in the online CPD course. This is the first time that BCTs have been identified in an FMSQ course and triangulated with before-and-after measures of CPD-REACTION. The results suggest that the BCTs effectively targeted determinants of intention, such as beliefs about capabilities and social influence, which had changed after the course. However, this study, which analyzed only 1 course, cannot make recommendations on which BCTs to use for which behaviors. Nevertheless, our results offer significant methodological progress in evaluating CPD course content, which could lead to innovations in CPD course development. Further studies on the use of theory-informed BCTs in CPD course design to enhance clinician adoption of targeted behaviors could further our understanding of effective course development strategies.

Limitations

First, the quasi-experimental before-and-after study design is vulnerable to confounding concomitant events that could influence study outcomes because the design comprises only a single group [50]. However, the short duration of the online CPD course (1 hour) made it less likely for other events to introduce bias. We chose this before-and-after study design as it is better suited to the constraints of practice settings than randomized control trials and better reflects the real-world CPD context [50]. Furthermore, we confirmed bivariate analysis with the GEE model to better measure the effect of the CPD course.

Triangulation of quantitative with qualitative methods improved the internal validity of the study [50,108].

Second, our results mainly apply to specialist physicians who chose to follow a nonmandatory CPD course and thus had some interest in the subject. The effect of the CPD course could have differed if it were mandatory for all physicians, as seen in other education programs [109]. Furthermore, general practitioners were underrepresented in this study (n=9). Thus, we do not know how general practitioners would respond to the CPD course but have no reason to think that they would behave differently than specialists. In addition, the well-being of all health care professionals is paramount, and it is not clear if the online CPD course would have any impact on other types of health professionals.

Third, there was a possible social desirability bias whereby physicians overrated their intention to approach a colleague in difficulty [50]. Confidentiality and anonymity were ensured to limit this bias. Of note, we measured intention using the same instrument used by Bakwa et al [72] and by Ayivi-Vinz et al [85] in their studies, so the desirability bias would have applied to their results as well and does not explain the higher before-and-after intention difference we observed in this study.

Fourth, this study does not allow us to confirm the reasons for which results regarding beliefs about consequences diverged, reasons to change behavior that were prominent in qualitative results but unstable or nonsignificant in the multiple linear regression. Although hypotheses are detailed in the Discussion, further studies would be needed to validate them.

Fifth, another limitation is that we used a backward elimination procedure, which can increase the risk of overfitting and biased effect estimates. However, results in the full model and fitted model were similar.

Sixth, we lacked statistical power to measure the difference in intention to approach a colleague in difficulty between those who did and those who did not adopt the behavior.

Lastly, self-reporting of behavior may also overestimate behavior adoption, as participants may not have used the approach taught in the course and may still report having adopted the behavior. Furthermore, only a small proportion of participants completed the self-reported behavior questionnaire 4 months after the course. This may indicate a selection bias in favor of the most motivated physicians, which could overestimate the proportion of those adopting the behavior. The sensitivity analysis in [Multimedia Appendix 8](#) shows that this potential selection bias on number of questionnaires answered

does not affect the primary outcome (mean difference in intention to approach a colleague in difficulty) but only the secondary outcome of self-reported behavior 4 months after the training.

Conclusions

To the best of our knowledge, this is the first study that uses validated before-and-after intervention measures to examine the effect of a CPD course on physicians' intention to approach a colleague in difficulty (experiencing distress, burnout, or second-victim symptoms after a patient adverse event). The CPD course produced an increase in physicians' intention regardless of age, gender, or domain of practice (laboratory, surgical, and medical domains). Beliefs about capabilities was an important determinant of intention to approach a colleague in difficulty. Although the CPD course increased physicians' intention to approach a colleague in difficulty, there is still a gap in understanding what other factors predict this intention. Within 4 months of the CPD course, 41% (25/61) of the physicians reported approaching a colleague in difficulty using strategies taught in the online CPD course. Identifying BCTs in the CPD course design was a methodological innovation for FMSQ course evaluation.

Although the CPD course demonstrated a positive effect on physicians' intention to approach a colleague in difficulty, more research is needed to better understand the mechanisms through which the intervention works. This could be done by translating the course into English, making it available to English-speaking physicians in Quebec and other Canadian provinces or internationally, and measuring the effect of the CPD course on approaching a colleague in difficulty in other contexts. Furthermore, instructional designers should focus on enhancing physicians' beliefs about their capabilities to facilitate their adoption of this peer support behavior. Instructional designers could consider using BCTs, especially when targeted behaviors are less widely accepted by clinicians or perceived as more difficult to adopt. Evaluating the effect of BCTs on behavioral intention could deepen our understanding of interventions aimed at changing clinicians' behaviors.

Our results suggest that peer support courses help build a culture of care among physicians. By enhancing our knowledge of interventions aimed at improving the peer support capabilities of physicians, this project indirectly aimed to improve the well-being of health care professionals. Peer-supported professionals could then be better able to contribute to resilient health care systems and to providing quality care to patients.

Acknowledgments

We would like to highlight the contributions of the scientific committee to the development of this training program, including Dr. Pierre Laramée, Cardiologist in the Department of Medicine at the Centre hospitalier de l'Université de Montréal; Dr. Anne Magnan, General Practitioner and former Medical Advisor and Executive Director at the Quebec Physicians' Assistance Program; and Dr. Yves Tremblay, Pediatrician specializing in Developmental Pediatrics, Public Health, and Preventive Medicine, and Medical Advisor at the Quebec Physicians' Assistance Program. Furthermore, we underline the contributions of several collaborators at various stages of this research project: Louisa Blair (Editor) for editing this paper; Sabrina Guay-Belanger (PhD) for her contribution to designing the protocol and submitting it to the ethics committee; Georgina Suelene Dofara (MSc) for her assistance

with the ethics committee application and review of the paper; and Janyce Gnanvi, who carried out diagnostic evaluations of the linear regression model and programmed the generalized estimating equations (GEE) approach to corroborate findings from bivariate analyses.

The authors declare the use of generative artificial intelligence (GAI) in the research and writing process. According to the Generative Artificial Intelligence–Driven Education Taxonomy (GAIDeT, 2025), optimization of less than 10% of code, the proofreading and editing of 25% of manuscript, translation of questionnaires included in the Multimedia Appendices, and reformatting of certain sentences were delegated to GAI tools under full human supervision. The GAI tools used were Copilot (Microsoft) and ChatGPT (OpenAI). The suggestions made by the GAI tools were verified, modified, and approved by the authors, who bear responsibility for results presented in this study.

Funding

The work was supported by the Fédération des médecins spécialistes du Québec (FMSQ). France Légaré is supported by a Canada Research Chair in Shared Decision Making and Knowledge Mobilization.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors contributed to the conceptualization of the study. MT and SJD carried out the investigation. RCC, EL, and F Lizotte developed the methodology for qualitative data analysis and RCC supervised qualitative analysis. F Lizotte and EL analyzed the qualitative data. F Lizotte performed data curation and conducted quantitative data analysis. SG, as biostatistician, validated all quantitative data analyses except for the diagnostic evaluations of the linear regression model and the programming of the GEE approach, which were carried out entirely by JG. CBU contributed to project administration and participated in drafting the original manuscript. F Légaré and CB supervised the overall research activities. All authors were involved in reviewing and editing the manuscript and have approved the final submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Conceptual framework.

[[DOCX File , 216 KB - mededu_v12i1e80199_app1.docx](#)]

Multimedia Appendix 2

Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Checklist.

[[DOCX File , 117 KB - mededu_v12i1e80199_app2.docx](#)]

Multimedia Appendix 3

Standards for Reporting Qualitative Research (SRQR) Checklist.

[[DOCX File , 113 KB - mededu_v12i1e80199_app3.docx](#)]

Multimedia Appendix 4

Complete Questionnaires and Consent Forms (French version).

[[DOCX File , 122 KB - mededu_v12i1e80199_app4.docx](#)]

Multimedia Appendix 5

Complete Questionnaires and Consent Forms (English version).

[[DOCX File , 120 KB - mededu_v12i1e80199_app5.docx](#)]

Multimedia Appendix 6

Grid for evaluating the CPD course "Approaching a Colleague in Difficulty" offered by the Fédération des médecins spécialistes du Québec (FMSQ) according to Michie's Behavior Change Techniques.

[[DOCX File , 115 KB - mededu_v12i1e80199_app6.docx](#)]

Multimedia Appendix 7

Number and percentage of participants with missing data for each variable.

[[DOCX File , 124 KB - mededu_v12i1e80199_app7.docx](#)]

Multimedia Appendix 8

Assessing Robustness in Bivariate Comparisons.

[[DOCX File , 126 KB - mededu_v12i1e80199_app8.docx](#)]

Multimedia Appendix 9

Analysis of mean intention difference by age, gender and by domain of medical specialty.

[[DOCX File , 114 KB - mededu_v12i1e80199_app9.docx](#)]

Multimedia Appendix 10

Sensitivity analysis.

[[DOCX File , 115 KB - mededu_v12i1e80199_app10.docx](#)]

Multimedia Appendix 11

Supplementary qualitative results.

[[DOCX File , 120 KB - mededu_v12i1e80199_app11.docx](#)]

References

1. Physician wellness: new 2021 National Physician Health Survey findings — burnout, short-staffing and an overburdened system take their toll. Canadian Medical Association. URL: <https://www.cma.ca/physician-wellness-hub/content/physician-wellness-new-2021-national-physician-health-survey> [accessed 2023-01-26]
2. Shanafelt TD, West CP, Sinsky C, Trockel M, Tutty M, Satele DV. Changes in burnout and satisfaction with work-life integration in physicians and the general US working population between 2011 and 2017. Mayo Clin Proc 2019;94(9):1681-1694 [[FREE Full text](#)] [doi: [10.1016/j.mayocp.2018.10.023](https://doi.org/10.1016/j.mayocp.2018.10.023)] [Medline: [30803733](https://pubmed.ncbi.nlm.nih.gov/30803733/)]
3. Shanafelt TD, West CP, Dyrbye LN, Trockel M, Tutty M, Wang H. Changes in burnout and satisfaction with work-life integration in physicians during the first 2 years of the COVID-19 pandemic. Mayo Clin Proc 2022;97(12):2248-2258 [[FREE Full text](#)] [doi: [10.1016/j.mayocp.2022.09.002](https://doi.org/10.1016/j.mayocp.2022.09.002)] [Medline: [36229269](https://pubmed.ncbi.nlm.nih.gov/36229269/)]
4. Rotenstein LS, Torre M, Ramos MA, Rosales RC, Guille C, Sen S. Prevalence of burnout among physicians: a systematic review. JAMA 2018;320(11):1131-1150 [[FREE Full text](#)] [doi: [10.1001/jama.2018.12777](https://doi.org/10.1001/jama.2018.12777)] [Medline: [30326495](https://pubmed.ncbi.nlm.nih.gov/30326495/)]
5. West CP, Dyrbye LN, Sinsky CA, Trockel M, Tutty M, Satele D, et al. Social isolation and burnout, professional fulfillment, and suicidal ideation among US physicians. Mayo Clin Proc 2025;100(12):2115-2127 [[FREE Full text](#)] [doi: [10.1016/j.mayocp.2025.07.025](https://doi.org/10.1016/j.mayocp.2025.07.025)] [Medline: [41175133](https://pubmed.ncbi.nlm.nih.gov/41175133/)]
6. Miller SC, Scott SD, Beck M. Second victims and mindfulness: A systematic review. J Patient Saf Risk Manag 2019;24(3):108-117. [doi: [10.1177/2516043519838176](https://doi.org/10.1177/2516043519838176)]
7. Seys D, Wu AW, Van Gerven E, Vleugels A, Euwema M, Panella M. Health care professionals as second victims after adverse events: a systematic review. Eval Health Prof 2013;36(2):135-162. [doi: [10.1177/0163278712458918](https://doi.org/10.1177/0163278712458918)] [Medline: [22976126](https://pubmed.ncbi.nlm.nih.gov/22976126/)]
8. Whitehead PB, Haisch CE, Hankey MS, Mutcheson RB, Dewitt SA, Stewart CA. Studying moral distress and moral injury among inpatient and outpatient healthcare professionals during the COVID-19 pandemic. Int J Psychiatry Med 2024;59(4):469-486. [doi: [10.1177/00912174231205660](https://doi.org/10.1177/00912174231205660)] [Medline: [37807925](https://pubmed.ncbi.nlm.nih.gov/37807925/)]
9. Hamric AB. A case study of moral distress. J Hosp Palliat Nurs 2014;16(8):457-463. [doi: [10.1097/njh.00000000000000104](https://doi.org/10.1097/njh.00000000000000104)]
10. Dewa CS, Loong D, Bonato S, Trojanowski L. The relationship between physician burnout and quality of healthcare in terms of safety and acceptability: a systematic review. BMJ Open 2017;7(6):e015141 [[FREE Full text](#)] [doi: [10.1136/bmjopen-2016-015141](https://doi.org/10.1136/bmjopen-2016-015141)] [Medline: [28637730](https://pubmed.ncbi.nlm.nih.gov/28637730/)]
11. Hodkinson A, Zhou A, Johnson J, Geraghty K, Riley R, Zhou A. Associations of physician burnout with career engagement and quality of patient care: systematic review and meta-analysis. BMJ 2022;378:e070442 [[FREE Full text](#)] [doi: [10.1136/bmj-2022-070442](https://doi.org/10.1136/bmj-2022-070442)] [Medline: [36104064](https://pubmed.ncbi.nlm.nih.gov/36104064/)]
12. Dewa CS, Jacobs P, Thanh NX, Loong D. An estimate of the cost of burnout on early retirement and reduction in clinical hours of practicing physicians in Canada. BMC Health Serv Res 2014;14:254 [[FREE Full text](#)] [doi: [10.1186/1472-6963-14-254](https://doi.org/10.1186/1472-6963-14-254)] [Medline: [24927847](https://pubmed.ncbi.nlm.nih.gov/24927847/)]
13. 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](https://doi.org/10.1001/jama.2021.25181)] [Medline: [35061006](https://pubmed.ncbi.nlm.nih.gov/35061006/)]
14. Carbone R, Ferrari S, Callegarin S, Casotti F, Turina L, Artioli G. Peer support between healthcare workers in hospital and out-of-hospital settings: a scoping review. Acta Biomed 2022;93(5):e2022308 [[FREE Full text](#)] [doi: [10.23750/abm.v93i5.13729](https://doi.org/10.23750/abm.v93i5.13729)] [Medline: [36300208](https://pubmed.ncbi.nlm.nih.gov/36300208/)]

15. Finney RE, Czinski S, Fjerstad K, Arteaga GM, Weaver AL, Riggan KA. Evaluation of a second victim peer support program on perceptions of second victim experiences and supportive resources in pediatric clinical specialties using the Second Victim Experience and Support Tool (SVEST). *J Pediatr Nurs* 2021;61:312-317. [doi: [10.1016/j.pedn.2021.08.023](https://doi.org/10.1016/j.pedn.2021.08.023)] [Medline: [34500175](#)]
16. Burlison JD, Scott SD, Browne EK, Thompson SG, Hoffman JM. The second victim experience and support tool: validation of an organizational resource for assessing second victim effects and the quality of support resources. *J Patient Saf* 2017;13(2):93-102 [FREE Full text] [doi: [10.1097/PTS.0000000000000129](https://doi.org/10.1097/PTS.0000000000000129)] [Medline: [25162208](#)]
17. Meredith LS, Ahluwalia S, Chen PG, Dong L, Farmer CM, Bouskill KE. Testing an intervention to improve health care worker well-being during the COVID-19 pandemic: a cluster randomized clinical trial. *JAMA Netw Open* 2024;7(4):e244192 [FREE Full text] [doi: [10.1001/jamanetworkopen.2024.4192](https://doi.org/10.1001/jamanetworkopen.2024.4192)] [Medline: [38687482](#)]
18. Pelikan M, Finney RE, Jacob A. Use of the Second Victim Experience and Support Tool (SVEST) to assess the impact of a departmental peer support program on anesthesia professionals' second victim experiences and perceptions of support two years after implementation. *AANA J* 2023;91(5):371-379. [Medline: [37788179](#)]
19. Shapiro J, Galowitz P. Peer support for clinicians: a programmatic approach. *Acad Med* 2016;91(9):1200-1204. [doi: [10.1097/ACM.0000000000001297](https://doi.org/10.1097/ACM.0000000000001297)] [Medline: [27355784](#)]
20. Creamer MC, Varker T, Bisson J, Darte K, Greenberg N, Lau W. Guidelines for peer support in high-risk organizations: an international consensus study using the Delphi method. *J Trauma Stress* 2012;25(2):134-141. [doi: [10.1002/jts.21685](https://doi.org/10.1002/jts.21685)] [Medline: [22522726](#)]
21. Scott SD, Hirschinger LE, Cox KR, McCoig M, Hahn-Cover K, Epperly KM. Caring for our own: deploying a systemwide second victim rapid response team. *Jt Comm J Qual Patient Saf* 2010;36(5):233-240. [doi: [10.1016/s1553-7250\(10\)36038-7](https://doi.org/10.1016/s1553-7250(10)36038-7)] [Medline: [20480757](#)]
22. Edrees H, Connors C, Paine L, Norvell M, Taylor H, Wu AW. Implementing the RISE second victim support programme at the Johns Hopkins Hospital: a case study. *BMJ Open* 2016;6(9):e011708 [FREE Full text] [doi: [10.1136/bmjopen-2016-011708](https://doi.org/10.1136/bmjopen-2016-011708)] [Medline: [27694486](#)]
23. Davidson JE, Graham P, Montross-Thomas L, Norcross W, Zerbi G. Code Lavender: cultivating intentional acts of kindness in response to stressful work situations. *Explore (NY)* 2017;13(3):181-185. [doi: [10.1016/j.explore.2017.02.005](https://doi.org/10.1016/j.explore.2017.02.005)] [Medline: [28668136](#)]
24. Merandi J, Winning A, Liao N, Rogers E, Lewe D, Gerhardt C. Implementation of a second victim program in the neonatal intensive care unit: an interim analysis of employee satisfaction. *J Patient Saf Risk Manag* 2018;23(6):231-238 [FREE Full text] [doi: [10.1177/2516043518809457](https://doi.org/10.1177/2516043518809457)]
25. Lane MA, Newman BM, Taylor MZ, O'Neill M, Ghetti C, Wolzman RM. Supporting clinicians after adverse events: development of a clinician peer support program. *J Patient Saf* 2018;14(3):e56-e60 [FREE Full text] [doi: [10.1097/PTS.0000000000000508](https://doi.org/10.1097/PTS.0000000000000508)] [Medline: [29878948](#)]
26. El Hechi MW, Bohnen JD, Westfal M, Han K, Cauley C, Wright C. Design and impact of a novel surgery-specific second victim peer support program. *J Am Coll Surg* 2020;230(6):926-933. [doi: [10.1016/j.jamcollsurg.2019.10.015](https://doi.org/10.1016/j.jamcollsurg.2019.10.015)] [Medline: [31857209](#)]
27. Keyser EA, Weir LF, Valdez MM, Aden JK, Matos RI. Extending peer support across the Military Health System to decrease clinician burnout. *Mil Med* 2021;186(Suppl 1):153-159. [doi: [10.1093/milmed/usaa225](https://doi.org/10.1093/milmed/usaa225)] [Medline: [32830273](#)]
28. Klatt TE, Sachs JF, Huang CC, Pilarski AM. Building a program of expanded peer support for the entire health care team: no one left behind. *Jt Comm J Qual Patient Saf* 2021;47(12):759-767. [doi: [10.1016/j.jcq.2021.08.010](https://doi.org/10.1016/j.jcq.2021.08.010)] [Medline: [34580016](#)]
29. Schuster MA. Creating the Hematology/Oncology/Stem Cell Transplant Advancing Resiliency Team: a nurse-led support program for hematology/oncology/stem cell transplant staff. *J Pediatr Oncol Nurs* 2021;38(5):331-341. [doi: [10.1177/10434542211011046](https://doi.org/10.1177/10434542211011046)] [Medline: [33988484](#)]
30. Wolf M, Smith K, Basu M, Heiss K. The prevalence of second victim syndrome and emotional distress in pediatric intensive care providers. *J Pediatr Intensive Care* 2023;12(2):125-130 [FREE Full text] [doi: [10.1055/s-0041-1731666](https://doi.org/10.1055/s-0041-1731666)] [Medline: [37082466](#)]
31. Hinzmann D, Forster A, Koll-Krüsmann M, Schießl A, Schneider F, Sigl-Erkel T. Calling for help—peer-based psychosocial support for medical staff by telephone: a best practice example from Germany. *Int J Environ Res Public Health* 2022;19(23): [FREE Full text] [doi: [10.3390/ijerph192315453](https://doi.org/10.3390/ijerph192315453)] [Medline: [36497526](#)]
32. Wilkie T, Tajirian T, Stergiopoulos V. Advancing physician wellness, engagement and excellence in a mental health setting: a Canadian perspective. *Health Promot Int* 2022;37(1). [doi: [10.1093/heapro/daab061](https://doi.org/10.1093/heapro/daab061)] [Medline: [33942064](#)]
33. Barghouth U, Gianakos A, Verran D, Klein SE, Butler KL, Miller P. Peer support helps physicians navigate workplace conflict. *J Gen Intern Med* 2023;38(7):1768-1770 [FREE Full text] [doi: [10.1007/s11606-023-08092-9](https://doi.org/10.1007/s11606-023-08092-9)] [Medline: [36917432](#)]
34. Doyle K, Murray T, Fong IC, Chavez A, Rounds G, Linenberger M. Building a culture of support at a pediatric surgery center through multidisciplinary peer support. *J Surg Res* 2023;291:90-96 [FREE Full text] [doi: [10.1016/j.jss.2023.04.029](https://doi.org/10.1016/j.jss.2023.04.029)] [Medline: [37354705](#)]
35. Liu JJW, Lum PA, Foxcroft L, Lim R, Richardson JD. Faculty-wide peer-support program during the COVID-19 pandemic: design and preliminary results. *JMIR Form Res* 2023;7:e37527 [FREE Full text] [doi: [10.2196/37527](https://doi.org/10.2196/37527)] [Medline: [36862464](#)]

36. Simms L, Ottman KE, Griffith JL, Knight MG, Norris L, Karakcheyeva V. Psychosocial peer support to address mental health and burnout of health care workers affected by COVID-19: a qualitative evaluation. *Int J Environ Res Public Health* 2023;20(5) [FREE Full text] [doi: [10.3390/ijerph20054536](https://doi.org/10.3390/ijerph20054536)] [Medline: [36901545](#)]
37. Simpson SL, Khan S, Schiferl LM, Boehl L, Horewitz D, Hausfeld J. Implementation of a peer-to-peer support program in a quaternary pediatric medical center. *Acad Pediatr* 2023;23(8):1481-1488 [FREE Full text] [doi: [10.1016/j.acap.2023.07.004](https://doi.org/10.1016/j.acap.2023.07.004)] [Medline: [37482296](#)]
38. Tolins ML, Rana JS, Lippert S, LeMaster C, Kimura YF, Sax DR. Implementation and effectiveness of a physician-focused peer support program. *PLoS One* 2023;18(11):e0292917 [FREE Full text] [doi: [10.1371/journal.pone.0292917](https://doi.org/10.1371/journal.pone.0292917)] [Medline: [37910457](#)]
39. Fall F, Hu YY, Walker S, Baertschiger R, Gaffar I, Saltzman D. Peer support to promote surgeon well-being: the APSA program experience. *J Pediatr Surg* 2024;59(9):1665-1671. [doi: [10.1016/j.jpedsurg.2023.12.022](https://doi.org/10.1016/j.jpedsurg.2023.12.022)] [Medline: [38272766](#)]
40. Carne B, Furyk J. Supporting clinicians post exposure to potentially traumatic events: emergency department peer support program evaluation. *Emerg Med Australas* 2025;37(1):e14518. [doi: [10.1111/1742-6723.14518](https://doi.org/10.1111/1742-6723.14518)] [Medline: [39397272](#)]
41. High AE, Forest S. Peer support targeting the second victim phenomenon: implementation and outcomes. *Cureus* 2025;17(2):e78854. [doi: [10.7759/cureus.78854](https://doi.org/10.7759/cureus.78854)] [Medline: [39949643](#)]
42. Graham P, Zerbi G, Norcross W, Montross-Thomas L, Lobbstaedt L, Davidson J. Testing of a caregiver support team. *Explore (NY)* 2019;15(1):19-26. [doi: [10.1016/j.explore.2018.07.004](https://doi.org/10.1016/j.explore.2018.07.004)] [Medline: [30424993](#)]
43. Finney RE, Jacob A, Johnson J, Messner H, Pulos B, Siggum H. Implementation of a second victim peer support program in a large anesthesia department. *AANA J* 2021;89(3):235-244. [Medline: [34042575](#)]
44. Katzman JW, Tomedi LE, McCoy-Hayes S, Richardson K, Romero E, Rosenbaum N. The Project ECHO first responder resiliency program: curriculum development, listening groups and lessons learned to support providers virtually during a pandemic. *Int Rev Psychiatry* 2021;33(8):682-690 [FREE Full text] [doi: [10.1080/09540261.2021.2021867](https://doi.org/10.1080/09540261.2021.2021867)] [Medline: [35412429](#)]
45. Bellehsen MH, Cook HM, Shaam P, Burns D, D'Amico P, Goldberg A. Adapting the Stress First Aid model for frontline healthcare workers during COVID-19. *Int J Environ Res Public Health* 2024;21(2):497-506 [FREE Full text] [doi: [10.3390/ijerph21020171](https://doi.org/10.3390/ijerph21020171)] [Medline: [38397662](#)]
46. Knudsen T, Abrahamsen C, Jørgensen JS, Schrøder K. Validation of the Danish version of the Second Victim Experience and Support Tool. *Scand J Public Health* 2022;50(4):497-506 [FREE Full text] [doi: [10.1177/14034948211004801](https://doi.org/10.1177/14034948211004801)] [Medline: [33860696](#)]
47. McLean CP, Betsworth D, Bihday C, Daman MC, Davis CA, Kayser D. Helping the helpers: adaptation and evaluation of Stress First Aid for healthcare workers in the Veterans Health Administration during the COVID-19 pandemic. *Workplace Health Saf* 2023;71(4):162-171 [FREE Full text] [doi: [10.1177/21650799221148650](https://doi.org/10.1177/21650799221148650)] [Medline: [36726298](#)]
48. Gidengil C, Meredith LS, Ahluwalia S, Chen PG, Dong L, Farmer C. Protecting the well-being of healthcare workers during the COVID-19 pandemic—the COVER-HCW project. Patient-Centered Outcomes Research Institute (PCORI) 2024 [FREE Full text]
49. Timmins G, Williamson S, Cassells A, Davis K, Dong L, Tobin JN. Health care worker experiences with a brief peer support and well-being intervention during the COVID-19 pandemic. *BMC Health Serv Res* 2025;25(1):1253 [FREE Full text] [doi: [10.1186/s12913-025-13268-6](https://doi.org/10.1186/s12913-025-13268-6)] [Medline: [41029307](#)]
50. Fortin MF, Gagnon J. Fondements et étapes du processus de recherche: Méthodes quantitatives et qualitatives. 4th Edition. Montréal: Chenelière Éducation; 2022:496.
51. Godin G. Les comportements dans le domaine de la santé: comprendre pour mieux intervenir. Montréal: Presses de l'Université de Montréal; 2012.
52. Michie S, Johnston M, Francis J, Hardeman W, Eccles M. From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. *Appl Psychol* 2008;57(4):660-680 [FREE Full text] [doi: [10.1111/j.1464-0597.2008.00341.x](https://doi.org/10.1111/j.1464-0597.2008.00341.x)] [Medline: [18631386](#)]
53. Godin G, Bélanger-Gravel A, Eccles M, Grimshaw J. Healthcare professionals' intentions and behaviours: a systematic review of studies based on social cognitive theories. *Implement Sci* 2008;3(6):36 [FREE Full text] [doi: [10.1186/1748-5908-3-36](https://doi.org/10.1186/1748-5908-3-36)] [Medline: [18631386](#)]
54. Légaré F, Freitas A, Turcotte S, Borduas F, Jacques A, Luconi F. Responsiveness of a simple tool for assessing change in behavioral intention after continuing professional development activities. *PLoS One* 2017;12(5):e0176678 [FREE Full text] [doi: [10.1371/journal.pone.0176678](https://doi.org/10.1371/journal.pone.0176678)] [Medline: [28459836](#)]
55. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *Epidemiology* 2007;18(6):805-835. [doi: [10.1097/EDE.0b013e3181577511](https://doi.org/10.1097/EDE.0b013e3181577511)] [Medline: [18049195](#)]
56. 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](https://doi.org/10.1097/ACM.0000000000000388)] [Medline: [24979285](#)]
57. About CPD accreditation. Royal College of Physicians and Surgeons of Canada. URL: <https://news.royalcollege.ca/en/standards-and-accreditation/cpd-accreditation> [accessed 2025-03-29]

58. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014;348:g1687 [FREE Full text] [doi: [10.1136/bmj.g1687](https://doi.org/10.1136/bmj.g1687)] [Medline: [24609605](https://pubmed.ncbi.nlm.nih.gov/24609605/)]

59. Novick JS. Six conseils pour soutenir un collègue. Fédération des médecins omnipraticiens du Québec (FMOQ). 2021. URL: https://www.pamq.org/wp-content/uploads/2021/10/Ecouter_6-conseils-pour-soutenir-un-collegue_fmoq.pdf [accessed 2025-02-27]

60. Prendre ses signes vitaux psychologiques. Québec Physicians' Health Program. 2021. URL: <https://www.pamq.org/wp-content/uploads/2021/08/Prendre-ses-signes-vitaux-psychologiques.pdf> [accessed 2025-02-27]

61. Événements personnels pouvant avoir un impact psychologique sur les médecins qui les vivent. Québec Physicians' Health Program. 2021. URL: <https://www.pamq.org/wp-content/uploads/2021/10/Evenements-personnels-pouvant-avoir-un-impact.pdf> [accessed 2025-02-27]

62. Situations pouvant amener un médecin à intervenir auprès d'un collègue. Québec Physicians' Health Program. 2021. URL: <https://www.pamq.org/wp-content/uploads/2021/10/Situations-pouvant-amener-un-medecin-a-intervenir.pdf> [accessed 2026-02-27]

63. Practice transformation series: preventing physician distress and suicide: sample scripting for approaching distressed physicians. American Medical Association. URL: <https://tinyurl.com/572v9ubw> [accessed 2026-01-20]

64. L'approche bienveillante axée sur le rétablissement. Québec Physicians' Health Program. 2021. URL: <https://www.pamq.org/wp-content/uploads/2022/12/LApproche-bienveillante-axee-sur-le-retablissement.pdf> [accessed 2024-05-05]

65. Événements professionnels stressants en médecine. Québec Physicians' Health Program. 2021. URL: <https://www.pamq.org/wp-content/uploads/2021/10/Evenements-professionnels-stressants.pdf> [accessed 2025-02-27]

66. L'identité rétrécie. Québec Physicians' Health Program. 2021. URL: <https://www.pamq.org/wp-content/uploads/2021/10/Identite-retrecie.pdf> [accessed 2025-02-27]

67. Stratégies individuelles pour traverser une période difficile. Québec Physicians' Health Program. 2021. URL: <https://www.pamq.org/wp-content/uploads/2021/10/Strategies-individuelles-pour-traverser-une-période-difficile.pdf> [accessed 2025-02-27]

68. Bloom BS. Taxonomie des objectifs pédagogiques. 2nd Edition. Montréal: Presses de l'Université du Québec; 1977.

69. Frank JR, Snell L, Sherbino J, Boucher A. Référentiel de compétences CanMEDS 2015 pour les médecins. Ottawa: Collège royal des médecins et chirurgiens du Canada; 2015.

70. Specialists for you: FMSQ. Federation of Medical Specialists of Quebec. 2025. URL: <https://www.fmsq.org/en> [accessed 2026-01-20]

71. Hrisos S, Eccles MP, Francis JJ, Dickinson HO, Kaner EFS, Beyer F. Are there valid proxy measures of clinical behaviour? A systematic review. *Implement Sci* 2009;4(17):37-41 [FREE Full text] [doi: [10.1186/1748-5908-4-37](https://doi.org/10.1186/1748-5908-4-37)] [Medline: [19575790](https://pubmed.ncbi.nlm.nih.gov/19575790/)]

72. Bakwa Kanyinga F, Gogovor A, Dofara SG, Gadio S, Tremblay M, Daniel SJ. Evaluating the impact of continuing professional development courses on physician behavioral intention: a pre-post study with follow-up at six months. *BMC Med Educ* 2023;23(1):629 [FREE Full text] [doi: [10.1186/s12909-023-04597-3](https://doi.org/10.1186/s12909-023-04597-3)] [Medline: [37661265](https://pubmed.ncbi.nlm.nih.gov/37661265/)]

73. Légaré F, Borduas F, Freitas A, Jacques A, Godin G, Luconi F. Development of a simple 12-item theory-based instrument to assess the impact of continuing professional development on clinical behavioral intentions. *PLoS One* 2014;9(3):e91013 [FREE Full text] [doi: [10.1371/journal.pone.0091013](https://doi.org/10.1371/journal.pone.0091013)] [Medline: [24643173](https://pubmed.ncbi.nlm.nih.gov/24643173/)]

74. Légaré F, Borduas F, Freitas A, Turcotte S. User Manual – The Continuing Professional Development (CPD) Reaction Questionnaire. 2015. URL: https://www.dropbox.com/scl/fi/npjz7cbdyjoc5f3ixvzm/CPD-Reaction_User-Manual.pdf?rlkey=5gw4tjicr3su0mof5aun4w8pj&e=1&dl=0 [accessed 2026-01-20]

75. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ* 2011;2:53-55 [FREE Full text] [doi: [10.5116/ijme.4dfb.8dfd](https://doi.org/10.5116/ijme.4dfb.8dfd)] [Medline: [28029643](https://pubmed.ncbi.nlm.nih.gov/28029643/)]

76. Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. *Annu Rev Public Health* 2002;23:151-169. [doi: [10.1146/annurev.publhealth.23.100901.140546](https://doi.org/10.1146/annurev.publhealth.23.100901.140546)] [Medline: [11910059](https://pubmed.ncbi.nlm.nih.gov/11910059/)]

77. Ancelle T. Statistique épidémiologie. 5th Edition. Paris: Maloine; 2023.

78. From data to action: understanding the drivers of physician wellness. Regression analyses – CMA 2021 National Physician Health Survey results. Canadian Medical Association. 2022. URL: <https://www.cma.ca/sites/default/files/2022-11/NPHS-regression-analysis-FR.pdf> [accessed 2024-03-18]

79. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychol* 2013;4:863 [FREE Full text] [doi: [10.3389/fpsyg.2013.00863](https://doi.org/10.3389/fpsyg.2013.00863)] [Medline: [24324449](https://pubmed.ncbi.nlm.nih.gov/24324449/)]

80. Ballinger GA. Using generalized estimating equations for longitudinal data analysis. *Organ Res Methods* 2004;7(2):127-150. [doi: [10.1177/1094428104263672](https://doi.org/10.1177/1094428104263672)] [Medline: [15227000](https://pubmed.ncbi.nlm.nih.gov/15227000/)]

81. Vittinghoff E. Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models. New York: Springer; 2012.

82. Hastie T, Tibshirani R, Friedman JH. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer; 2009.

83. Huh I, Gim J. Exploration of Likert scale in terms of continuous variable with parametric statistical methods. *BMC Med Res Methodol* 2025;25(1):218 [FREE Full text] [doi: [10.1186/s12874-025-02668-1](https://doi.org/10.1186/s12874-025-02668-1)] [Medline: [41023822](https://pubmed.ncbi.nlm.nih.gov/41023822/)]

84. Rhemtulla M, Brosseau-Liard P, Savalei V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol Methods* 2012;17(3):354-373. [doi: [10.1037/a0029315](https://doi.org/10.1037/a0029315)] [Medline: [22799625](https://pubmed.ncbi.nlm.nih.gov/22799625/)]

85. Ayivi-Vinz G, Tremblay M, Gadio S, Dofara SG, Daniel SJ, Talbot D. Impact of educational activity formats, online or in-person, on the intention of medical specialists to adopt a clinical behaviour: a comparative study. *J CME* 2024;13(1):2363550 [FREE Full text] [doi: [10.1080/28338073.2024.2363550](https://doi.org/10.1080/28338073.2024.2363550)] [Medline: [38873619](https://pubmed.ncbi.nlm.nih.gov/38873619/)]

86. Kivits JL, Balard FDR, Fournier CC, Winance M, Kivits J, Balard F. Les recherches qualitatives en santé. Malakoff: Armand Colin; 2016.

87. Cheung K, Tai K. The use of intercoder reliability in qualitative interview data analysis in science education. *Res Sci Technol Educ* 2021;41(3):1155-1175. [doi: [10.1080/02635143.2021.1993179](https://doi.org/10.1080/02635143.2021.1993179)]

88. Östlund U, Kidd L, Wengström Y, Rowa-Dewar N. Combining qualitative and quantitative research within mixed method research designs: a methodological review. *Int J Nurs Stud* 2011;48(3):369-383 [FREE Full text] [doi: [10.1016/j.ijnurstu.2010.10.005](https://doi.org/10.1016/j.ijnurstu.2010.10.005)] [Medline: [21084086](https://pubmed.ncbi.nlm.nih.gov/21084086/)]

89. Sutton S. Interventions très brèves sur le changement de comportement dans des environnements de soins de santé. *Practical Health Psychology* 2017 [FREE Full text]

90. Reed SK. In: Chipman SEF, editor. *The Oxford Handbook of Cognitive Science*. London: Oxford University Press; 2017:231-248.

91. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educ Psychol Meas* 1960;20(1):37-46 [FREE Full text] [doi: [10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104)]

92. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159 [FREE Full text] [doi: [10.2307/2529310](https://doi.org/10.2307/2529310)] [Medline: [38589863](https://pubmed.ncbi.nlm.nih.gov/38589863/)]

93. Bos-van den Hoek DW, van Laarhoven HWM, Ali R, Bakker SD, Goosens A, Hendriks MP. Blended online learning for oncologists to improve skills in shared decision making about palliative chemotherapy: a pre-posttest evaluation. *Support Care Cancer* 2023;31(3):184 [FREE Full text] [doi: [10.1007/s00520-023-07625-6](https://doi.org/10.1007/s00520-023-07625-6)] [Medline: [36820944](https://pubmed.ncbi.nlm.nih.gov/36820944/)]

94. Taqif H, Adisso L, Gomes Souza L, Dofara SG, Ghio SC, Rivest LP. Changes in intention to use an interprofessional approach to decision-making following training: a cluster before-and-after study. *BMC Health Serv Res* 2024;24(1):437-450 [FREE Full text] [doi: [10.1186/s12913-024-10899-z](https://doi.org/10.1186/s12913-024-10899-z)] [Medline: [38589863](https://pubmed.ncbi.nlm.nih.gov/38589863/)]

95. van de Water LF, Kuijper SC, Henselmans I, van Alphen EN, Kooij ES, Calff MM. Effect of a prediction tool and communication skills training on communication of treatment outcomes: a multicenter stepped wedge clinical trial (the SOURCE trial). *EClinicalMedicine* 2023;64:102244 [FREE Full text] [doi: [10.1016/j.eclim.2023.102244](https://doi.org/10.1016/j.eclim.2023.102244)] [Medline: [37781156](https://pubmed.ncbi.nlm.nih.gov/37781156/)]

96. Terrill BN, Pearce A, Chau A, Young MA. Navigating genomic testing: evaluation of an e-learning module with general practitioners. *Focus Health Prof Educ* 2024;25(1):37-50. [doi: [10.11157/fohpe.v25i1.630](https://doi.org/10.11157/fohpe.v25i1.630)]

97. Latif A, Waring J, Chen LC, Pollock K, Solomon J, Gulzar N. Supporting the provision of pharmacy medication reviews to marginalised (medically underserved) groups: a before/after questionnaire study investigating the impact of a patient-professional co-produced digital educational intervention. *BMJ Open* 2019;9(9):e031548 [FREE Full text] [doi: [10.1136/bmjopen-2019-031548](https://doi.org/10.1136/bmjopen-2019-031548)] [Medline: [31530620](https://pubmed.ncbi.nlm.nih.gov/31530620/)]

98. Yu C, Choi D, Bruno BA, Thorpe KE, Straus SE, Cantarutti P. Impact of MyDiabetesPlan, a web-based patient decision aid on decisional conflict, diabetes distress, quality of life, and chronic illness care in patients with diabetes: cluster randomized controlled trial. *J Med Internet Res* 2020;22(9):e16984 [FREE Full text] [doi: [10.2196/16984](https://doi.org/10.2196/16984)] [Medline: [32996893](https://pubmed.ncbi.nlm.nih.gov/32996893/)]

99. Lawani MA, Côté L, Morin M, Witteman HO, Caron D, Kroger E. Professional training on shared decision making with older adults living with neurocognitive disorders: a mixed-methods implementation study. *BMC Med Inform Decis Mak* 2020;20(1):189. [doi: [10.1186/s12911-020-01197-9](https://doi.org/10.1186/s12911-020-01197-9)]

100. Urban K, Wright PB, Hester AL, Curran G, Rojo M, Tsai PF. Evaluation of an education strategy versus usual care to implement the STEADI algorithm in primary care clinics in an academic medical center. *Clin Interv Aging* 2020;15(6):1059-1066 [FREE Full text] [doi: [10.2147/CIA.S256416](https://doi.org/10.2147/CIA.S256416)] [Medline: [32753856](https://pubmed.ncbi.nlm.nih.gov/32753856/)]

101. McWilliams S. Addressing compassion fatigue of a family caregiver of adolescents in a mental health setting. *Int J Nurs Health Care Res* 2020;3(11):1206 [FREE Full text] [doi: [10.29011/2688-9501.101206](https://doi.org/10.29011/2688-9501.101206)]

102. Cheung IW, Bland VR, Castro-Noriega E, Silva DS, Miller KJ, Michalak EE. Pilot of a training format for clinicians' continuing professional development: open house with rotating short segments. *J Contin Educ Health Prof* 2022;42(1):e92-e95 [FREE Full text] [doi: [10.1097/CEH.0000000000000366](https://doi.org/10.1097/CEH.0000000000000366)] [Medline: [34108385](https://pubmed.ncbi.nlm.nih.gov/34108385/)]

103. Rivera-Chiauzzi EY, Smith HA, Moore-Murray T, Lee C, Goffman D, Bernstein PS. Healing our own: a randomized trial to assess benefits of peer support. *J Patient Saf* 2022;18(1):e308-e314. [doi: [10.1097/PTS.0000000000000771](https://doi.org/10.1097/PTS.0000000000000771)] [Medline: [32925571](https://pubmed.ncbi.nlm.nih.gov/32925571/)]

104. Babiker A, El Husseini M, Al Nemri A, Al Frayh A, Al Juryyan N, Faki MO. Health care professional development: working as a team to improve patient care. *Sudan J Paediatr* 2014;14(2):9-16 [FREE Full text] [Medline: [27493399](https://pubmed.ncbi.nlm.nih.gov/27493399/)]

105. Kruper A, Domeyer-Klenske A, Treat R, Pilarski A, Kaljo K. Secondary traumatic stress in Ob-Gyn: a mixed methods analysis assessing physician impact and needs. *J Surg Educ* 2021;78(3):1024-1034. [doi: [10.1016/j.jsurg.2020.08.038](https://doi.org/10.1016/j.jsurg.2020.08.038)] [Medline: [32948508](https://pubmed.ncbi.nlm.nih.gov/32948508/)]
106. Dollard MF, Bakker AB. Psychosocial safety climate as a precursor to conducive work environments, psychological health problems, and employee engagement. *J Occup Organ Psychol* 2010;83(3):579-599 [FREE Full text] [doi: [10.1348/096317909X470690](https://doi.org/10.1348/096317909X470690)] [Medline: [17570838](https://pubmed.ncbi.nlm.nih.gov/17570838/)]
107. Loh MY, Idris MA, Dollard MF, Isahak M. Psychosocial safety climate as a moderator of the moderators: Contextualizing JDR models and emotional demands effects. *J Occup Organ Psychol* 2018;91(3):620-644. [doi: [10.1111/joop.12211](https://doi.org/10.1111/joop.12211)]
108. O'Cathain A, Murphy E, Nicholl J. Why, and how, mixed methods research is undertaken in health services research in England: a mixed methods study. *BMC Health Serv Res* 2007;7:85. [doi: [10.1186/1472-6963-7-85](https://doi.org/10.1186/1472-6963-7-85)]
109. Pawson R, Tilley N. *Realistic Evaluation*. Thousand Oaks: Sage Publications; 1997:231-247.

Abbreviations

BCT: behavior change technique

CIUSSCN: Centre intégré universitaire de santé et de services sociaux de la Capitale-Nationale

CPD: continuing professional development

FMSQ: Fédération des médecins spécialistes du Québec

GEE: generalized estimating equations

QPHP: Québec Physicians' Health Program

SRQR: Standards for Reporting Qualitative Research

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

TIDieR: Template for Intervention Description and Replication

Edited by A Stone, T Leung; submitted 29.Aug.2025; peer-reviewed by P Whitehead, H Kwon; comments to author 05.Nov.2025; revised version received 10.Dec.2025; accepted 31.Dec.2025; published 05.Feb.2026.

Please cite as:

Lizotte F, Tremblay M, Biron C, Lachance É, Gadio S, de Carvalho Corôa R, Uwizeye CB, Daniel SJ, Légaré F

Effect of an Online Continuing Professional Development Course on Physicians' Intention to Approach a Colleague in Difficulty: Mixed Methods Convergent Study

JMIR Med Educ 2026;12:e80199

URL: <https://mededu.jmir.org/2026/1/e80199>

doi: [10.2196/80199](https://doi.org/10.2196/80199)

PMID: [41474517](https://pubmed.ncbi.nlm.nih.gov/41474517/)

©Florence Lizotte, Martin Tremblay, Caroline Biron, Éloi Lachance, Souleymane Gadio, Roberta de Carvalho Corôa, Claude Bernard Uwizeye, Sam J Daniel, France Légaré. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 05.Feb.2026. 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.

Combining Problem-Based Learning Methods With the WeChat Platform in Teaching Ophthalmology: Randomized Controlled Trial

Fang Fang, MD, PhD; Bing Bu, BM; Wenmin Jiang, MD, PhD

Department of Ophthalmology in the Second Xiangya Hospital, Central South University, 139 Renmin Road, Changsha, China

Corresponding Author:

Wenmin Jiang, MD, PhD

Department of Ophthalmology in the Second Xiangya Hospital, Central South University, 139 Renmin Road, Changsha, China

Abstract

Background: Ophthalmology poses distinct learning challenges for medical students due to the complex anatomy of the eye and the requirement of essential hands-on skills. Problem-based learning (PBL), a student-centered approach, fosters clinical reasoning and self-directed learning. To address the time and logistical constraints of traditional teaching methods, this study implemented a WeChat-based PBL model that leveraged the platform's efficiency and interactivity to enhance student engagement and skill acquisition in ophthalmology.

Objective: This study aimed to evaluate the effectiveness of WeChat-based PBL in ophthalmology education, focusing on its impact on students' self-perception of learning and clinical skills compared to traditional teaching methods.

Methods: This study involved 108 undergraduate students who successfully passed the Chinese National Entrance Examination. Among them, 54 (50%) were randomly selected to participate in the WeChat-based PBL, while the other 54 (50%) received traditional teaching. Students were placed into 6 groups (18 students for each group) using a random number table, and the new teaching methods were tested outside their regular class time. Three groups were randomly selected to receive PBL using WeChat as the platform, while the remaining 3 groups received conventional teaching.

Results: Our analysis indicated that although students in the WeChat-based PBL group scored marginally lower in memorization compared to their peers in the traditional teaching group (traditional group: mean 37.6, SD 2.8; WeChat group: mean 32.0, SD 4.1; $P=.006$; $n=54$), they exhibited markedly superior levels of understanding (traditional group: mean 24.1, SD 1.8; WeChat group: mean 28.0, SD 1.3; $P=.007$; $n=54$) and knowledge application (traditional group: mean 24.3, SD 1.9; WeChat group: mean 27.6, SD 1.3; $P=.008$; $n=54$). This suggests that the WeChat-based PBL method promotes deeper engagement, enabling students to better comprehend essential concepts, even with a diminished emphasis on rote learning. Additionally, students in the WeChat group reported increased collaboration (traditional group: mean 3.8889, SD 0.8393; WeChat group: mean 1.7222, SD 0.5961; $P<.001$); motivation (traditional group: mean 3.5471, SD 0.7915; WeChat group: mean 1.8333, SD 0.5746; $P=.004$); knowledge acquisition (traditional group: mean 3.6667, SD 0.7770; WeChat group: mean 1.8704, SD 0.7017; $P<.001$); self-learning ability (traditional group: mean 3.5741, SD 0.7673; WeChat group: mean 1.8519, SD 0.4917; $P<.001$); clinical reasoning (traditional group: mean 2.9444, SD 0.8777; WeChat group: mean 1.9630, SD 0.6132; $P=.01$); and problem-solving skills (traditional group: mean 3.2037, SD 0.6553; WeChat group: mean 1.8519, SD 0.5287; $P=.001$) than the students in the traditional group.

Conclusions: Integrating PBL methods with WeChat has been shown to improve ophthalmic education outcomes compared to traditional teaching, suggesting that this method may offer a superior alternative to conventional teaching.

(*JMIR Med Educ* 2026;12:e65279) doi:[10.2196/65279](https://doi.org/10.2196/65279)

KEYWORDS

problem-based learning; WeChat; China; ophthalmology; undergraduate; medical students

Introduction

Medical education increasingly incorporates innovative strategies in undergraduate training to better prepare students for the challenges of clinical practice. Key strategies such as problem-based learning (PBL) [1,2], case-based learning [3], team-based learning [4], flipped classroom-based learning [5,6], and competency-based medical education [7] have been

emphasized to bridge the gap between theoretical knowledge and clinical skills. These strategies are particularly relevant in fields such as ophthalmology, which represents a distinct and highly specialized discipline within clinical medicine [8]. The intricate anatomy of the eye and the nonintuitive nature of its structure necessitate the use of specialized equipment and tests for accurate diagnosis and management [9]. These factors contribute to the considerable hurdles medical students

encounter while attempting to master this subject. Therefore, effective teaching methodologies are vital for adequately preparing students for clinical practice and developing their competencies as future clinicians [10].

The conventional teacher-centered educational model often relies on the passive transmission of knowledge from instructors to students. While efficient for delivering core knowledge, especially in large classes common in China, this method can limit students' engagement and active learning [11]. As a result, students may struggle to operate ophthalmic instruments or perform fundus examinations, thereby diminishing their confidence in applying ophthalmology skills. Moreover, although textbooks provide detailed descriptions of the clinical manifestations and signs of ophthalmic diseases, students frequently encounter challenges when attempting to apply this theoretical knowledge to real-world clinical scenarios [12].

PBL, which inherently starts with a clinically based problem as its name, is an active learning strategy that fosters student-centered engagement [13]. Evidence shows that PBL enhances learning motivation, promotes self-guided learning, and improves the integration of theoretical knowledge with clinical practice [14,15]. Students exposed to PBL report greater satisfaction and participation compared with those in traditional learning settings [16,17]. In today's digital era, college students are adept at using the internet and mobile networks to access information, helping to overcome the temporal and spatial limitations of traditional PBL methods. With the advancement of mobile networks, WeChat has become a popular social networking platform in China [18]. It supports multiple modes of interaction—including texting, voice messages, images, and multimedia—similar to WhatsApp in Western countries [19,20]. Moreover, WeChat is widely used for disseminating knowledge to medical students due to its convenience and the extensive availability of information [21,22].

This study addresses the challenges faced in conventional ophthalmology education, where passive learning hinders the development of practical skills and confidence among undergraduate medical students. It aims to evaluate the effectiveness and acceptability of a WeChat-based PBL approach in enhancing students' self-perceptions of learning, initiative, organization, clinical skills, and knowledge mastery. The central research question guiding this inquiry is: How does the integration of WeChat in PBL influence the learning outcomes of medical students in ophthalmology? This randomized controlled trial will involve the comparison of student outcomes between those engaged in traditional teaching methods and those using the WeChat platform for online PBL, allowing for a robust assessment of the intervention's

effectiveness. We hypothesize that the use of WeChat for online PBL will significantly improve student engagement and competency in ophthalmology. The findings of this study will provide valuable insights for medical educators, curriculum developers, and policymakers seeking to improve medical education methodologies in China and beyond.

Methods

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki and received approval from the institutional review board and ethics committee of the Second Xiangya Hospital (Z0005-01). In accordance with institutional guidelines for educational research, the intervention was classified as minimal risk and considered part of the routine evaluation of teaching quality. All participants were informed of the study's purpose, procedures, and the planned use of the collected data, and all provided written informed consent before participation. Participant anonymity was strictly protected, with all data deidentified and analyzed only in aggregated form. No study procedures posed risks beyond those ordinarily encountered in standard educational activities.

Study Participants

A total of 108 fifth-year undergraduate students from Xiangya Medical College of Central South University, all of whom had passed the Chinese National Entrance Examination, participated in the study. The cohort consisted of 42 (38.9%) male and 66 (61.1%) female participants, aged 20 to 22 years. Participants were randomly assigned into 6 groups of 18 students each using Microsoft Excel to generate random numbers to ensure unbiased distribution. In this study, students were informed of their group assignments due to the nature of the teaching intervention, but evaluators responsible for assessing the examination results were blinded to group allocation to minimize assessment bias. The new teaching methods were implemented outside their regular class schedules, primarily during evenings and weekends. This approach presented unique challenges, including an increased workload for both students and instructors, as it required students to allocate additional time beyond their standard academic hours to participate in the study. Additionally, coordinating schedules among all faculty members and students involved posed challenges, which required careful planning and communication to ensure participation without conflicting with existing institutional educational policies. Both intervention and control groups received instruction in dedicated classrooms at the hospital, which facilitated a conducive learning environment (Table 1).

Table . Timeline of the WeChat-based problem-based learning activities.

Week	Activity	Duration (h)	Learning objectives
1	The instructor posts guiding questions and students form study groups	2	Students are randomly assigned to study groups of 10 and establish WeChat groups, through which group members communicate about the study content
2	Explore contributing factors	4	Identify and explain major factors and search for medical literature and books on your own to find the answers to the guiding questions provided by the instructor
3	Apply concepts to scenarios	4	Apply learned concepts to hypothetical scenarios and communicate with group members and the instructor through the WeChat group
4	Solve the guiding problems and present the key points	5	Students create relevant course Microsoft PowerPoint presentations according to the assigned groups, which cover all the answers to the instructor's guiding questions and the knowledge points obtained from self-study of the teaching content of this course
5	Evaluate proposed solutions	4	Evaluate the effectiveness of various solutions by presenting them to each other during the formal class

Research Methods

We selected primary angle-closure glaucoma (PACG) as the topic for applying the WeChat-based PBL approach in this study, as the diagnosis and treatment of PACG represent crucial skills that students must acquire. The textbook used for this study was *Ophthalmology*, 9th edition, published by the People's Medical Publishing House [23].

WeChat-Based PBL

Establishment of Learning Groups

Three groups (18 students for each group) were randomly selected to take part in the WeChat-based PBL program. In each group, 1 volunteer self-selected by students was responsible for recording each student's speech and the outcomes of the panel discussions.

Problem Introduction Phase

In alignment with the national syllabus and the ophthalmology textbook, 3 representative clinical cases were selected from the 2022-2023 glaucoma service database of the Second Xiangya Hospital. Each vignette depicted a distinct stage or mechanism of PACG (including asymptomatic primary angle closure, acute angle closure attack, and plateau-iris configuration) covering the essential diagnostic and therapeutic spectrum required for undergraduate training. To guide case selection and design, chapter-level learning objectives were first extracted, encompassing (1) anterior-chamber anatomy and angle structures, (2) the pathophysiology of pupillary-block versus nonpupillary-block mechanisms, (3) biometric risk factors, (4) gonioscopic and imaging diagnosis, (5) laser and surgical indications, and (6) emergency management of acute attacks.

Each objective was then translated into a practical key decision point (eg, interpreting a gonioscopy record, determining when laser peripheral iridotomy is indicated, or choosing between phacoemulsification and trabeculectomy). All selected cases, provided by the instructors, were distributed to students 1 week prior to the lecture to facilitate preparatory discussion and independent learning.

Preclass Guidance Phase

Instructors provided curated readings including key sections drawn from the textbook, supplemented with selected review articles, clinical guidelines (eg, *Asia-Pacific Glaucoma Guidelines, 3rd Edition*), and representative case reports from the hospital's glaucoma database. Additionally, each vignette was accompanied by inquiry-based questions that directed students to identify risk factors, compare pathogenic mechanisms, and determine appropriate management strategies, encouraging active and evidence-based learning. These questions could direct students' literature searches toward key concepts and clinical decision-making points related to PACG. Each group participating in the WeChat-based PBL established its own discussion team on the WeChat platform, fostering an environment of active participation. To encourage full student engagement, instructors emphasized the importance of collaboration and implemented a system that recognized contributions, motivating students to prepare thoroughly before classes. Through this platform, instructors and students engaged in interactive discussions on the provided cases, sharing their diagnostic reasoning, proposing treatment plans, and exploring the latest global developments in disease management. For example, instructors could prompt a discussion by asking specific diagnostic questions, thereby guiding students to reflect

critically. During class sessions, instructors monitored and reviewed the discussions and the online outputs, often projecting key discussion points or summaries from WeChat onto a shared screen for collective evaluation and real-time feedback.

Discussion and Summarization Phase

Each WeChat-based PBL session lasted 2 hours and was supported by the platform's features for sharing text, images, voice recordings, videos, and documents, which enabled immediate and efficient communication. All instructors followed a structured facilitator's guide, identical across groups, and distributed 1 week before the session. The guide provided a minute-by-minute timetable covering the entire 2-hour process (10 min recap, 45 min group presentation, 30 min cross-groups challenge, 25 min instructor synthesis, and 10 min reflection), ensuring balanced pacing and consistent coverage across groups.

At the beginning of each session, a representative student from each group presented their understanding of the assigned disease and summarized unresolved problems identified during preclass discussions. This was followed by an open, interactive discussion in which students from other groups exchanged perspectives and clinical reasoning. To maintain alignment with curricular standards, the instructor's guide specified 5 “must-cover” learning objectives mapped to the national ophthalmology syllabus (eg, explain the mechanism of pupillary block, justify the timing of laser iridotomy, and interpret

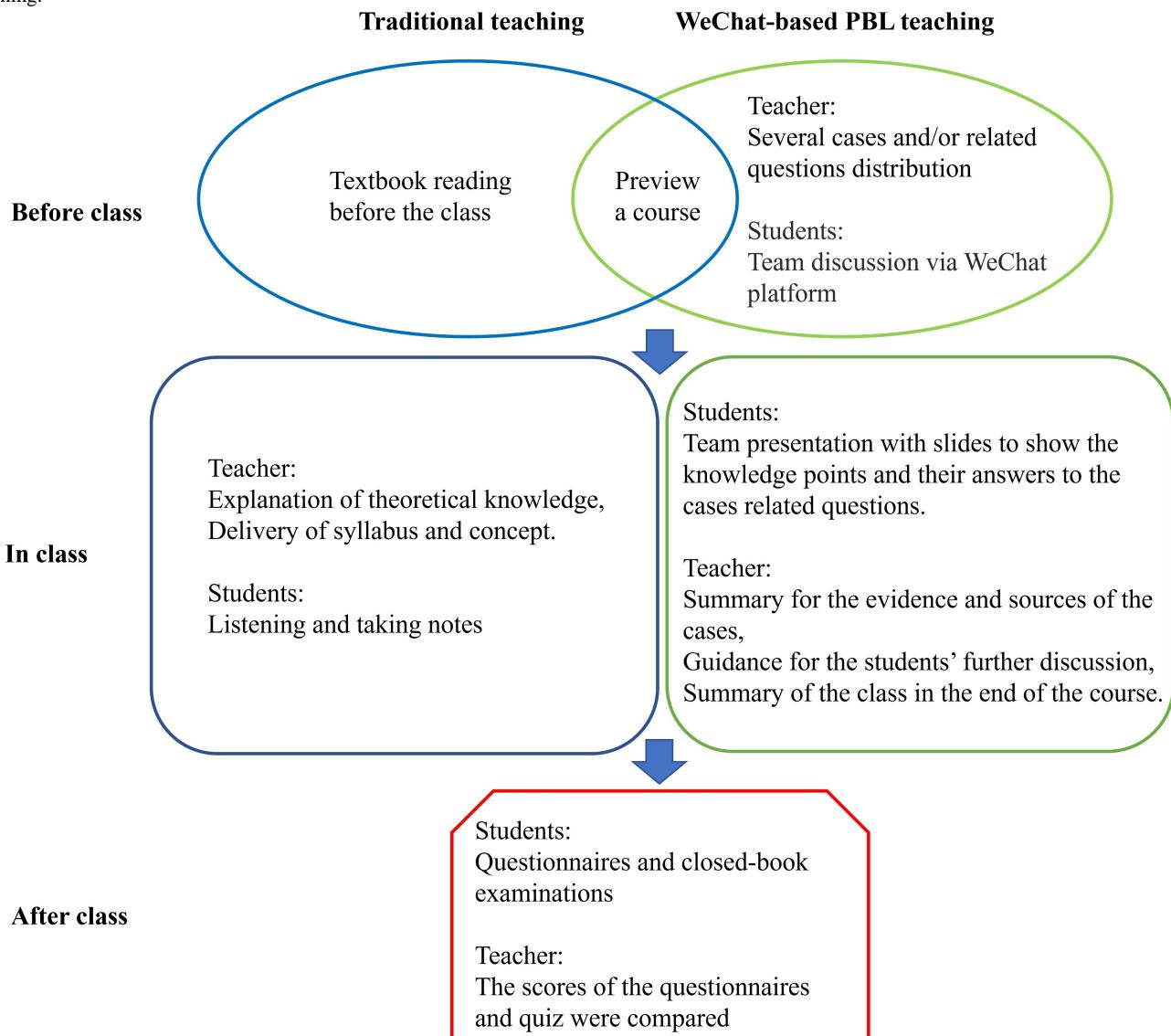
ultrasound biomicroscopy images). It also included prompt questions addressing common misconceptions identified during the pilot phase, and a 1-slide “take-home” summary that instructors were required to display at the end of the session.

After the session, instructors completed a structured checklist to document which objectives had been achieved. They were instructed to intervene only when discussion drifted more than 2 minutes off task or when critical misconceptions arose; otherwise, they guided students by inquiry-based questioning to ensure that all checklist objectives were discussed. The structured facilitation process, combined with WeChat's interactive features, fostered active participation, peer collaboration, and efficient knowledge integration, greatly enhanced the PBL learning experience.

Traditional Teaching Group

The traditional teaching model involved the instructor predominantly delivering comprehensive explanations of foundational knowledge, adhering to a prescribed multimedia teaching outline. In this model, students were not divided into smaller groups and were not required to give formal presentations during the course. Both groups completed the same quizzes and questionnaires, and the results did not affect course grades or performance. The study design is summarized in [Figure 1](#).

Figure 1. The flowchart shows the 2 different models of teaching and learning in ophthalmology for Chinese medical students. PBL: problem-based learning.



Teaching Efficacy Evaluation and Statistical Analysis

The efficacy and satisfaction of both teaching methods were evaluated through a combination of questionnaire surveys for subjective assessment and examinations for objective evaluation.

Subjective Evaluation

The questionnaire (Table 2) was designed and modified based on the Course Experience Questionnaire by Ramsden [24] and the Study Process Questionnaire by Biggs [25]. All the students completed questionnaires to assess their perceptions and self-evaluated competence at the end of the teaching course.

Table . Specific questions used in the questionnaire

Topic area	Questions
The course improves my learning motivation	<ul style="list-style-type: none"> “Can you analyze the impact of the study findings on the field?”
The course helps me to acquire knowledge	<ul style="list-style-type: none"> “Can you define the main terms used in the study?”
The course is helpful for passing the final examination	<ul style="list-style-type: none"> “What are your thoughts on the effectiveness of the methods used in the study?”
The course encourages me to express my opinions	<ul style="list-style-type: none"> “In this course, do you feel that your opinions are respected?” “In this course, do you often take the initiative to express your opinions?”
This course occupies too much of my spare time	<ul style="list-style-type: none"> “How much time do you spend in WeChat-based problem-based learning by using this method to communicate with teachers and group members?”
The course improves my communication skills	<ul style="list-style-type: none"> “During discussions, does the instructor give you sufficient time to express your opinions?”
The course improves my self-learning skills	<ul style="list-style-type: none"> “What are the key concepts discussed in the study material?”
The course improves my clinical thinking ability	<ul style="list-style-type: none"> “How would you apply the concepts learned to a real-world scenario?”
The course improves my ability to analyze and solve problems	<ul style="list-style-type: none"> “How much do you agree that this course has provided you with the tools and techniques to effectively solve problems?” “How often have you used the analytical skills you learned in the course to solve real-world problems outside the classroom?”

Objective Evaluation

Both groups then undertook an identical closed-book examination comprising essay questions, true or false questions, judgment questions, and case analyses to assess students' memory, comprehension, and clinical application abilities. The examination questions were formulated based on the 9th edition of *Ophthalmology*, an ophthalmology textbook published by the People's Medical Publishing House [23]. Notably, both the examination format and evaluation criteria were standardized across both groups. To maintain objectivity in scoring the theoretical examination, 3 instructors performed a blinded evaluation of the students' responses.

Statistical analyses were performed using SPSS (version 11.0; IBM Corp) software. Measurement data were expressed as means (SDs). Questionnaire data were analyzed using the

Mann-Whitney *U* test, and examination scores across the two groups were compared using the independent samples 2-tailed *t* test. A *P* value $<.05$ was considered statistically significant.

Results

Participants' Demographic Data

Figure 2 shows a flowchart of participation for the 108 medical students in this study. Among them, 54 (50%) students were randomly selected to participate in the WeChat-based PBL, while the other 54 (50%) received traditional teaching. Among the participants, there were 42 (38.9%) male and 66 (61.1%) female individuals who were aged 20 to 22 years. The mean age of the WeChat-based group was 20.426 (SD 0.602) years, and the mean age of the traditional teaching group was 20.296 (SD 0.571) years. There was no significant difference in gender and average age between the two groups (Table 3).

Figure 2. Consolidated Standards of Reporting Trials (CONSORT) flowchart.

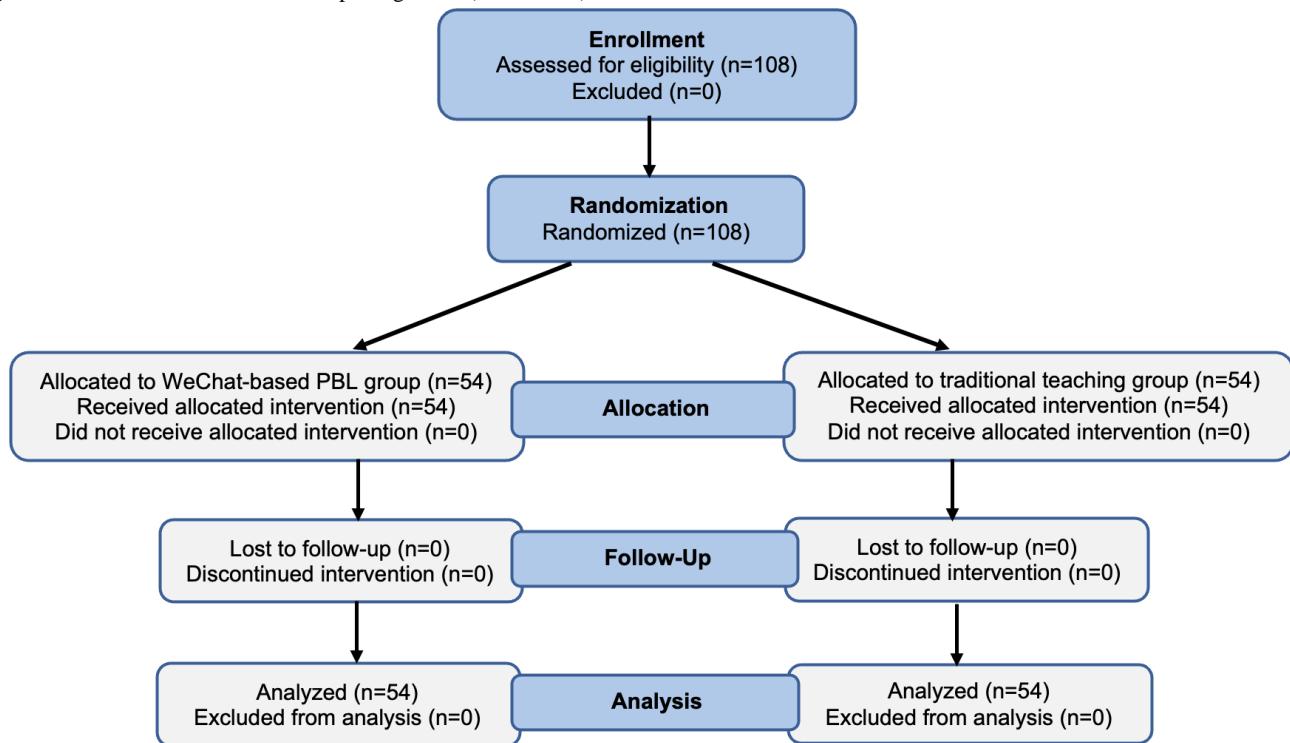


Table . Participants' demographic data of the Second Xiangya Medical College of WeChat-based problem-based learning (PBL) and traditional teaching groups.

Characteristics	WeChat-based PBL (n=54)	Traditional teaching (n=54)	P value
Age (y), mean (SD)	20.426 (0.602)	20.296 (0.571)	.12 ^a
Sex, n (%)			.48 ^b
Male	23 (42.6)	22 (40.7)	
Female	31 (57.4)	32 (59.3)	

^aFor age, independent sample $t_{106}=1.188$.

^bFor sex, $\chi^2_1=1.0$.

Comparison of the WeChat-Based PBL Method and Traditional Teaching Method

The closed-book examination was administered 1 week after course completion to allow students sufficient time for review while minimizing long-term memory effects. This clarification has been included in the revised manuscript. Examination questions were designed by 3 instructors and included 6 types of questions: single-choice questions, multiple-choice questions, noun definitions, short-answer questions, essay questions, and comprehensive questions. The exam assessed 3 cognitive levels—memorization, understanding, and application, which accounted for 50%, 20%, and 30% of the total score, respectively. The memorization section evaluated students' ability to recall fundamental facts, definitions, and terminology. The understanding section assessed their capacity to interpret information, explain concepts in their own words, and demonstrate deeper comprehension beyond simple recall. The application section examined students' ability to apply their knowledge and understanding to new or practical situations, such as problem-solving, formula application, and case analysis.

Notably, the sections on understanding and application emphasized students' flexible mastery of knowledge and its practical implementation. The goal was to enable students to apply theoretical knowledge in clinical contexts, thereby improving their diagnostic and treatment skills as well as their ability to analyze and solve clinical problems.

Our analysis indicated that although students in the WeChat-based PBL group scored marginally lower in memorization compared to their peers in the traditional teaching group (traditional vs WeChat groups: 37.6, SD 2.8 vs 32.0, SD 4.1; $t_{106}=-8.426$; $P=.006$; $n=54$), they exhibited markedly superior levels of understanding (traditional vs WeChat groups: 24.1, SD 1.8 vs 28.0, SD 1.3; $t_{106}=12.753$; $P=.007$; $n=54$) and knowledge application (traditional vs WeChat groups: 24.3, SD 1.9 vs 27.6, SD 1.3; $t_{106}=10.899$; $P=.008$; $n=54$). However, when considering the total score, there was no significant difference between the two groups (traditional vs WeChat: 85.8, SD 4.9 vs 87.4, SD 5.1; $t_{106}=1.694$; $P=.60$; $n=54$).

All 108 (100%) students responded to the questionnaire for the comparison of students' perceptions between the WeChat-based

PBL and lecture-based class (Table 4). More students believed (rather than did not believe) that the WeChat-based PBL class enhanced their learning motivation (traditional group: mean 3.5471, SD 0.7915; WeChat group: mean 1.8333, SD 0.5746; $P=.004$), helped them acquire knowledge (traditional group: mean 3.6667, SD 0.7770; WeChat group: mean 1.8704, SD 0.7017; $P<.001$), and improved their performance on the final examination ($P=.001$). They also believed that it encouraged them to express opinions (traditional group: mean 3.5556, SD 0.7181; WeChat group: mean 1.9630, SD 0.5482; $P=.001$), enhanced their communication skills (traditional group: mean 3.8889, SD 0.8393; WeChat group: mean 1.7222, SD 0.5961;

$P<.001$), improved their self-learning ability (traditional group: mean 3.5741, SD 0.7673; WeChat group: mean 1.8519, SD 0.4917; $P<.001$), improved their clinical thinking (traditional group: mean 2.9444, SD 0.8777; WeChat group: mean 1.9630, SD 0.6132; $P=.012$), and enhanced their ability to analyze and resolve problems (traditional group: mean 3.2037, SD 0.6553; WeChat group: mean 1.8519, SD 0.5287; $P=.001$). However, a majority of students in the WeChat-based PBL class felt that the course occupied too much of their spare time (traditional group: mean 3.7963, SD 0.7618; WeChat group: mean 1.9630, SD 0.5482; $P<.001$).

Table . Comparison of students' perceptions and self-evaluation between WeChat-based problem-based learning (PBL) method and traditional teaching method. The results were ranked as a Likert scale from 1 to 5 (1=strongly agree, 2=agree, 3=neutral, 4=disagree, and 5=strongly disagree).

Questions	WeChat-based PBL, mean (SD)	Control (traditional teaching), mean (SD)	Mann-Whitney <i>U</i> test	<i>P</i> value	Effect size
The course improves my learning motivation	1.8333 (0.5746)	3.5741 (0.7915)	1012	.004	2.5169
The course helps me to acquire the knowledge	1.8704 (0.7017)	3.6667 (0.7770)	914	<.001	2.4264
The course is helpful for passing the final examination	2.7593 (0.7507)	3.2593 (0.8284)	975.5	<.001	0.6325
The course encourages me to express my opinions	1.9630 (0.5482)	3.5556 (0.7181)	963	.001	2.4930
This course occupies too much of my spare time	3.3148 (0.8201)	3.7963 (0.7618)	714	<.001	0.6083
The course improves my communication skills	1.7222 (0.5961)	3.8889 (0.8393)	725	<.001	2.9765
The course improves my self-learning skills	1.8519 (0.4917)	3.5741 (0.7673)	930.5	<.001	2.6725
The course improves my clinical thinking ability	1.9630 (0.6132)	2.9444 (0.8777)	1072	.012	1.2964
The course improves my ability to analyze and solve problems	1.8519 (0.5287)	3.2037 (0.6553)	932	.001	2.2706

Discussion

Principal Findings

This study aimed to assess the effectiveness of WeChat-based PBL in ophthalmic education compared with traditional teaching methods. Our analysis revealed that while students in the WeChat-based PBL group exhibited slightly lower scores in memorization than those in the traditional teaching group, they demonstrated significantly higher levels of understanding and application of knowledge. This suggests that the WeChat-based PBL approach fosters deeper engagement, allowing students to grasp critical concepts more effectively, despite having less focus on rote memorization. Moreover, students in the WeChat group reported enhanced collaboration, motivation, and willingness to participate actively in their learning process.

WeChat has been used as a platform for PBL because of its time efficiency and convenience [26-28]. Students were able to present questions and engage in discussions with instructors at any time and from any location via WeChat. Moreover, it allows the sharing of images, videos, links, and other related resources. This capability greatly aids discussion and ensures that all members can access case-related or problem-related resources promptly. Everyone can share their opinions and suggestions, receiving feedback from others anytime and anywhere. This timely and efficient communication via WeChat offers a better understanding among team members [26,27,29]. Successful collaboration guarantees the productivity of the team, and instructors can actively participate in students' discussions and debates, responding to questions and offering advice to enhance the efficiency of the learning process [26,27].

Although the primary benefit of WeChat-based PBL is time-saving and convenience, as traditional PBL can be a time-intensive approach [30], the perception that WeChat-based PBL occupied too much spare time underscores the need for careful consideration of time management and workload in educational program design. One of the reasons could be the additional time dedicated to self-learning [31] and discussion among the team members increased the pressure and burden on the students. Increased preparation time can lead to deeper understanding and better retention of the material as students engage with content more thoroughly before class discussions. This proactive approach enhances participation during PBL sessions and ultimately improves learning outcomes. To optimize learning efficiency while preventing overload, several refinements are warranted. Specifically, preclass materials should be streamlined to emphasize key concepts and reduce redundancy, and brief in-class reviews can help reinforce complex topics, minimizing postclass study demands. Periodic guided learning sessions can also be scheduled to help balance students' independent study and supervised learning time. These refinements aim to alleviate cognitive load while maintaining the effectiveness of the teaching model. Moreover, it is essential to refine the new teaching approach to effectively integrate the advantages of the PBL model with the interactive capabilities of the WeChat platform. Furthermore, instructors should also receive appropriate training to embrace this innovative teaching approach [32]. Dolmans and Wolthagen [33] have highlighted that a tutor's actions significantly influence both the productivity and effectiveness of a PBL group's efforts. To enhance the effectiveness of tutors in PBL tutorials, ongoing training that encourages reflective practice on their development as educators is recommended [34,35]. These efforts can effectively strike a balance between optimal learning outcomes and manageable student workload.

At the same time, the increase in students' out-of-class learning hours also implied additional workload for instructors. Although instructor workload in asynchronous chat facilitation was not quantitatively measured, informal feedback indicated a moderate increase. Instructors reported spending an additional 15 to 20 minutes per day monitoring discussions and providing timely feedback. Despite this, they generally considered this extra time acceptable and valuable, as it enabled closer observation of students' reasoning and early identification of misconceptions. The main challenges involved balancing prompt responses with other duties and managing message overload during peak activity. Nonetheless, most instructors agreed that the improved engagement and learning outcomes justified the additional effort. A rotating instructor schedule for online facilitation and setting clearer expectations for response timelines may help balance instructional workload while maintaining effective student-teacher interaction.

Our statistical analysis revealed that the WeChat-based PBL group exhibited slightly lower scores in memorization than the traditional teaching group. However, their levels of understanding and application of knowledge were higher than the traditional teaching group. This discrepancy may be attributed to the engagement of students in the WeChat-based PBL group with a broader array of relevant information,

extending beyond textbook content. Consequently, while the WeChat-based PBL group demonstrated a better performance in understanding and application of knowledge, their memorization of textbook-specific information was slightly weaker. Nevertheless, the WeChat-based PBL group showed advantages in mastering key knowledge points and achieving higher total scores.

However, several factors must be considered while implementing WeChat-based PBL. For instance, it is advisable to keep the number of students in each team small, as smaller groups facilitate easier administration and instruction. Teachers should contact students at least 1 week in advance to provide questions or case reports, ensuring both teachers and students are well prepared to achieve optimal educational outcomes. By adapting to real-time situations, teachers can enhance students' interest in learning, invigorate the classroom dynamics, and correct students' misconceptions. Teachers should act as a resource and facilitator, responding promptly to students' inquiries throughout their research. Overall, the WeChat-based PBL model marks a significant departure from the traditional didactic "cramming education" approach, fostering a more engaging and active learning environment. Therefore, WeChat-based PBL should be increasingly incorporated into future ophthalmology education to enable continuous identification and refinement of relevant issues [36].

Limitations

There are several limitations that should be acknowledged. First, while students covered the same core content on PACG, differences in teaching methodologies could have influenced memorization outcomes and overall comprehension. This research was performed targeting a single textbook chapter on PACG. Therefore, the conclusion requires further verification in other subject fields. Second, questionnaire results indicated that students in the WeChat-based PBL group spent more extracurricular time preparing course materials. Therefore, understanding their perceptions and attitudes toward this additional workload is particularly important. However, as the students involved in this study have already graduated, it was not possible to collect further feedback from them. This limitation prevents a more comprehensive understanding of their long-term perceptions of the learning burden and its potential impact on sustained engagement and clinical competency development. Future research will include longitudinal follow-up surveys or interviews to further evaluate students' attitudes, learning experiences, and workload perceptions. Third, the necessity for student participation in additional discussions and presentations may have affected their independent study time and overall academic performance. Fourth, the lack of blinding may have introduced biases, as awareness of the teaching method could have influenced both students' and instructors' expectations and engagement levels. Fifth, the generalizability of our findings is also limited by the single-institution context and the sample size of 108 fifth-year undergraduate medical students, which may not reflect broader educational settings or diverse student populations. Finally, in our study, we did not compare the effects of traditional PBL and WeChat-based PBL methods. Further research should compare the functions between these two approaches in future

studies. Moreover, future research should explore the long-term impacts of WeChat-based PBL on learning outcomes across different medical disciplines and institutions. Further empirical testing is necessary to assess the effectiveness of this approach in various contexts and to identify best practices for implementation. Optimizing digital communication tools such as WeChat holds great promise for enhancing educational experience and fostering collaborative learning in medical education.

Conclusions

In summary, compared to traditional teaching, integrating PBL methods combined with WeChat as a communication platform improved the effectiveness of ophthalmic education. Although

students invested more time in preclass, they did not perceive it as a waste of their spare time. Instead, they recognized that the WeChat-based PBL mode facilitated more effective communication among students and between students and teachers. It enhanced their motivation to learn; promoted knowledge acquisition; encouraged expression of opinions; and improved self-learning abilities, clinical reasoning, and problem-solving. This innovative teaching mode may represent a superior alternative to conventional teaching methods. However, further exploration to optimize the students' spare time for effective online communication is warranted. This new method merits continued refinement and evaluation based on the findings of this study.

Funding

This work was supported by the 2024 Education and Teaching Reform Research Project of Central South University (grant 2024jy141), the Changsha Natural Science Foundation (grant kq2403090), the Hunan Province Health Commission (grant 202307027224), and the Natural Science Foundation of Hunan (grants 2023JJ10087 and 2025JJ50538).

Data Availability

The datasets used or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

WJ and FF designed the experiments and drafted the manuscript. FF and BB collected the data and performed the statistical analysis. WJ and BB reviewed the manuscript. All authors read and approved of the final manuscript.

Conflicts of Interest

None declared.

References

1. Jin J, Bridges SM. Educational technologies in problem-based learning in health sciences education: a systematic review. *J Med Internet Res* 2014 Dec 10;16(12):e251. [doi: [10.2196/jmir.3240](https://doi.org/10.2196/jmir.3240)] [Medline: [25498126](https://pubmed.ncbi.nlm.nih.gov/25498126/)]
2. Wyer PC. Evidence-based medicine and problem based learning a critical re-evaluation. *Adv Health Sci Educ Theory Pract* 2019 Dec;24(5):865-878. [doi: [10.1007/s10459-019-09921-5](https://doi.org/10.1007/s10459-019-09921-5)] [Medline: [31617018](https://pubmed.ncbi.nlm.nih.gov/31617018/)]
3. Thistlethwaite JE, Davies D, Ekeocha S, et al. The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME Guide No. 23. *Med Teach* 2012;34(6):e421-e444. [doi: [10.3109/0142159X.2012.680939](https://doi.org/10.3109/0142159X.2012.680939)] [Medline: [22578051](https://pubmed.ncbi.nlm.nih.gov/22578051/)]
4. Burgess A, Ayton T, Mellis C. Implementation of team-based learning in year 1 of a PBL based medical program: a pilot study. *BMC Med Educ* 2016 Feb 4;16:49. [doi: [10.1186/s12909-016-0550-3](https://doi.org/10.1186/s12909-016-0550-3)] [Medline: [26846425](https://pubmed.ncbi.nlm.nih.gov/26846425/)]
5. Han E, Klein KC. Pre-class learning methods for flipped classrooms. *Am J Pharm Educ* 2019 Feb;83(1):6922. [doi: [10.5688/ajpe6922](https://doi.org/10.5688/ajpe6922)] [Medline: [30894772](https://pubmed.ncbi.nlm.nih.gov/30894772/)]
6. Park KH, Park KH, Chae SJ. Experiences of medical teachers in flipped learning for medical students: a phenomenological study. *Korean J Med Educ* 2018 Jun;30(2):91-100. [doi: [10.3946/kjme.2018.84](https://doi.org/10.3946/kjme.2018.84)] [Medline: [29860775](https://pubmed.ncbi.nlm.nih.gov/29860775/)]
7. Succar T, McCluskey P, Grigg J. Enhancing medical student education by implementing a competency-based ophthalmology curriculum. *Asia Pac J Ophthalmol (Phila)* 2017;6(1):59-63. [doi: [10.22608/APO.2016102](https://doi.org/10.22608/APO.2016102)] [Medline: [28161917](https://pubmed.ncbi.nlm.nih.gov/28161917/)]
8. Gostimir M, Sharma RA, Bhatti A. Status of Canadian undergraduate medical education in ophthalmology. *Can J Ophthalmol* 2018 Oct;53(5):474-479. [doi: [10.1016/j.jcjo.2017.11.015](https://doi.org/10.1016/j.jcjo.2017.11.015)] [Medline: [30340714](https://pubmed.ncbi.nlm.nih.gov/30340714/)]
9. Wang A, Xiao R, Zhang C, et al. Effectiveness of a combined problem-based learning and flipped classroom teaching method in ophthalmic clinical skill training. *BMC Med Educ* 2022 Jun 23;22(1):487. [doi: [10.1186/s12909-022-03538-w](https://doi.org/10.1186/s12909-022-03538-w)] [Medline: [35733187](https://pubmed.ncbi.nlm.nih.gov/35733187/)]
10. Schmidt HG. Problem-based learning: does it prepare medical students to become better doctors? *Med J Aust* 1998 May 4;168(9):429-430. [doi: [10.5694/j.1326-5377.1998.tb139020.x](https://doi.org/10.5694/j.1326-5377.1998.tb139020.x)] [Medline: [9612452](https://pubmed.ncbi.nlm.nih.gov/9612452/)]
11. Tu MG, Yu CH, Wu LT, Li TC, Kwan CY. Dental and medical students' perspectives on early exposure to PBL in Taiwan. *J Dent Educ* 2012 Jun;76(6):746-751. [Medline: [22659703](https://pubmed.ncbi.nlm.nih.gov/22659703/)]
12. Al Wadani F, Khan AR. Problem-based learning in ophthalmology: a brief review. *Oman J Ophthalmol* 2014 Jan;7(1):1-2. [doi: [10.4103/0974-620X.127908](https://doi.org/10.4103/0974-620X.127908)] [Medline: [24799792](https://pubmed.ncbi.nlm.nih.gov/24799792/)]

13. Rui Z, Rong-Zheng Y, Hong-Yu Q, Jing Z, Xue-Hong W, Chuan Z. Preliminary investigation into application of problem-based learning in the practical teaching of diagnostics. *Adv Med Educ Pract* 2015;6:223-229. [doi: [10.2147/AMEP.S78893](https://doi.org/10.2147/AMEP.S78893)] [Medline: [25848334](#)]
14. Tudor Car L, Kyaw BM, Dunleavy G, et al. Digital problem-based learning in health professions: systematic review and meta-analysis by the digital health education collaboration. *J Med Internet Res* 2019 Feb 28;21(2):e12945. [doi: [10.2196/12945](https://doi.org/10.2196/12945)] [Medline: [30816846](#)]
15. Khaki AA, Tubbs RS, Zarrintan S, et al. The first year medical students' perception of and satisfaction from problem-based learning compared to traditional teaching in gross anatomy: introducing problem-based anatomy into a traditional curriculum in Iran. *Int J Health Sci (Qassim)* 2007 Jan;1(1):113-118. [Medline: [21475460](#)]
16. Albanese MA, Mitchell S. Problem-based learning: a review of literature on its outcomes and implementation issues. *Acad Med* 1993 Jan;68(1):52-81. [doi: [10.1097/00001888-199301000-00012](https://doi.org/10.1097/00001888-199301000-00012)] [Medline: [8447896](#)]
17. Groves M, Régo P, O'Rourke P. Tutoring in problem-based learning medical curricula: the influence of tutor background and style on effectiveness. *BMC Med Educ* 2005 Jun 7;5(1):20. [doi: [10.1186/1472-6920-5-20](https://doi.org/10.1186/1472-6920-5-20)] [Medline: [15938758](#)]
18. Sun ZJ, Zhu L, Liang M, Xu T, Lang JH. The usability of a WeChat-based electronic questionnaire for collecting participant-reported data in female pelvic floor disorders: a comparison with the traditional paper-administered format. *Menopause* 2016 Aug;23(8):856-862. [doi: [10.1097/GME.0000000000000690](https://doi.org/10.1097/GME.0000000000000690)] [Medline: [27326820](#)]
19. Hussain I, Khan RA, Mahboob U, Latif MZ, Waheed K. Comparison of effectiveness of WhatsApp and Facebook as learning tool for the students of a medical college. *J Pak Med Assoc* 2020 Oct;70(10):1705-1708. [doi: [10.5455/JPMA.35071](https://doi.org/10.5455/JPMA.35071)] [Medline: [33159737](#)]
20. Oyewole BK, Animasahun VJ, Chapman HJ. A survey on the effectiveness of WhatsApp for teaching doctors preparing for a licensing exam. *PLoS One* 2020;15(4):e0231148. [doi: [10.1371/journal.pone.0231148](https://doi.org/10.1371/journal.pone.0231148)] [Medline: [32240259](#)]
21. Zeng F, Deng G, Wang Z, Liu L. WeChat: a new clinical teaching tool for problem-based learning. *Int J Med Educ* 2016 Apr 25;7:119-121. [doi: [10.5116/ijme.5708.e5c4](https://doi.org/10.5116/ijme.5708.e5c4)] [Medline: [27111920](#)]
22. Wang J, Gao F, Li J, et al. The usability of WeChat as a mobile and interactive medium in student-centered medical teaching. *Biochem Mol Biol Educ* 2017 Sep;45(5):421-425. [doi: [10.1002/bmb.21065](https://doi.org/10.1002/bmb.21065)] [Medline: [28581156](#)]
23. Peizeng Y, Xianqun F. Ophthalmology, 9th edition: People's Medical Publishing House; 2018.
24. Ramsden P. A performance indicator of teaching quality in higher education: the course experience questionnaire. *High Educ Stud* 1991 Jan;16(2):129-150. [doi: [10.1080/03075079112331382944](https://doi.org/10.1080/03075079112331382944)]
25. Biggs JB. Study process questionnaire manual: student approaches to learning and studying. : Australian Council for Educational Research Ltd; 1987 URL: <https://eric.ed.gov/?id=ED308200> [accessed 2025-10-05]
26. Zhang W, Li ZR, Li Z. WeChat as a platform for problem-based learning in a dental practical clerkship: feasibility study. *J Med Internet Res* 2019 Mar 19;21(3):e12127. [doi: [10.2196/12127](https://doi.org/10.2196/12127)] [Medline: [30888328](#)]
27. Luo P, Pang W, Wang Y, et al. WeChat as a platform for problem-based learning among hematological postgraduates: feasibility and acceptability study. *J Med Internet Res* 2021 May 25;23(5):e16463. [doi: [10.2196/16463](https://doi.org/10.2196/16463)] [Medline: [34032573](#)]
28. Tu S, Yan X, Jie K, Ying M, Huang C. WeChat: an applicable and flexible social app software for mobile teaching. *Biochem Mol Biol Educ* 2018 Sep;46(5):555-560. [doi: [10.1002/bmb.21170](https://doi.org/10.1002/bmb.21170)] [Medline: [30369037](#)]
29. Zhou Y, Zhang D, Guan X, Pan Q, Deng S, Yu M. Application of WeChat-based flipped classroom on root canal filling teaching in a preclinical endodontic course. *BMC Med Educ* 2022 Mar 2;22(1):138. [doi: [10.1186/s12909-022-03189-x](https://doi.org/10.1186/s12909-022-03189-x)] [Medline: [35236332](#)]
30. Seneviratne RD, Samarasekera DD, Karunathilake IM, Ponnamperuma GG. Students' perception of problem-based learning in the medical curriculum of the Faculty of Medicine, University of Colombo. *Ann Acad Med Singap* 2001 Jul;30(4):379-381. [Medline: [11503544](#)]
31. 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](https://doi.org/10.5688/ajpe6898)] [Medline: [30013248](#)]
32. Gómez-Carrasco CJ, Monteagudo-Fernández J, Moreno-Vera JR, Sainz-Gómez M. Evaluation of a gamification and flipped-classroom program used in teacher training: perception of learning and outcome. *PLoS ONE* 2020;15(7):e0236083. [doi: [10.1371/journal.pone.0236083](https://doi.org/10.1371/journal.pone.0236083)] [Medline: [32673373](#)]
33. Dolmans DH, Wolfhagen IH. Complex interactions between tutor performance, tutorial group productivity and the effectiveness of PBL units as perceived by students. *Adv Health Sci Educ Theory Pract* 2005 Aug;10(3):253-261. [doi: [10.1007/s10459-005-0665-5](https://doi.org/10.1007/s10459-005-0665-5)] [Medline: [16193404](#)]
34. Hendry GD. Problem-based learning tutors' conceptions of their development as tutors. *Med Teach* 2009 Feb;31(2):145-150. [doi: [10.1080/01421590802146026](https://doi.org/10.1080/01421590802146026)] [Medline: [19330673](#)]
35. Bollela VR, Gabarra MHC, da Costa C, Lima RCP. Students and tutors' social representations of assessment in problem-based learning tutorials supporting change. *BMC Med Educ* 2009 Jun 7;9:30. [doi: [10.1186/1472-6920-9-30](https://doi.org/10.1186/1472-6920-9-30)] [Medline: [19500408](#)]
36. Hartling L, Spooner C, Tjosvold L, Oswald A. Problem-based learning in pre-clinical medical education: 22 years of outcome research. *Med Teach* 2010 Jan;32(1):28-35. [doi: [10.3109/01421590903200789](https://doi.org/10.3109/01421590903200789)] [Medline: [20095771](#)]

Abbreviations

PACG: primary angle closure glaucoma
PBL: problem-based learning

Edited by B Lesselroth; submitted 16.Aug.2024; peer-reviewed by E Li, K Zhu; revised version received 14.Nov.2025; accepted 25.Nov.2025; published 05.Jan.2026.

Please cite as:

Fang F, Bu B, Jiang W

Combining Problem-Based Learning Methods With the WeChat Platform in Teaching Ophthalmology: Randomized Controlled Trial

JMIR Med Educ 2026;12:e65279

URL: <https://mededu.jmir.org/2026/1/e65279>

doi:10.2196/65279

© Fang Fang, Bing Bu, Wenmin Jiang. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 5.Jan.2026. 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.

Original Paper

Ultrasound-Guided Regional Anesthesia in a Resource-Limited Hospital: Prospective Pilot Study of a Hybrid Training Program

Jakob E Gamboa¹, MD; Inge Tamm-Daniels¹, MD; Roland Flores¹, MD; Nancy G Sarat Diaz², MD; Mario A Villasenor¹, MD; Mitchell A Gist¹, MD; Aidan B Hoie¹, MD; Christopher Kurinec³, MS; Colby G Simmons¹, DO, MBA

¹Department of Anesthesiology, University of Colorado School of Medicine, Aurora, CO, United States

²Department of Anesthesiology, Hospital Nacional de Coatepeque, Coatepeque, Quetzaltenango, Guatemala

³Department of Biostatistics and Informatics, Colorado School of Public Health, Aurora, CO, United States

Corresponding Author:

Jakob E Gamboa, MD

Department of Anesthesiology

University of Colorado School of Medicine

Mail Stop B113

12401 E. 17th Avenue

Aurora, CO, 80045

United States

Phone: 1 3034826929

Email: jakob.gamboa@cuanschutz.edu

Abstract

Background: Ultrasound-guided regional anesthesia (UGRA) remains underused in low- and middle-income countries due to barriers to training and equipment. Recent advances in portable ultrasound devices and international partnerships have expanded access to UGRA, enhancing patient safety and quality of care.

Objective: This study describes the development and outcomes of a hybrid UGRA training program for anesthesiologists at the Hospital Nacional de Coatepeque (HNC) in Guatemala.

Methods: An educational pilot program for UGRA was developed based on local needs and feedback, comprising 4 weeks of online modules, an in-person educational conference, and 1 month of supervised clinical practice. Evaluation followed the Kirkpatrick framework using preprogram and postprogram surveys adapted from the Global Regional Anesthesia Curricular Engagement model. Outcomes included participants' satisfaction, change in knowledge and skill, and procedural performance. Knowledge and skill assessments were compared before and after the training, and clinical data were recorded for 10 months. Nonparametric tests were used to assess changes and associations with performance outcomes.

Results: All 7 anesthesiologists at HNC completed the training program. Knowledge test scores improved by a median percentage increase of 20.8% (IQR 13.5%-28.1%; $r=0.899$; $P=.02$), and procedural skill rating scores increased by a median percentage of 147.1% (IQR 96.9%-197.3%; $r=0.904$; $P=.03$) at 1 month and 131.4% (IQR 90.5%-172.3%; $r=0.909$; $P=.04$) at 4 months after the program. Participants self-reported high satisfaction and substantial clinical improvement and motivation. A total of 54 peripheral nerve blocks were performed under direct supervision in the first month, with 187 blocks recorded over 10 months. The supraclavicular brachial plexus block was the most frequently used (66/187, 35.3%) and replaced the standard general anesthetic for upper extremity surgery in 70 patients. The procedure success rate was 96.3% (180/187), and there were no observed patient complications.

Conclusions: This hybrid curriculum enabled the successful implementation of UGRA at a public hospital in Guatemala, safely expanding clinical capabilities and reducing reliance on general anesthesia for upper extremity surgery. This practical training model provides a framework for implementing UGRA in similar resource-limited hospitals.

(*JMIR Med Educ* 2026;12:e84181) doi:[10.2196/84181](https://doi.org/10.2196/84181)

KEYWORDS

health education; regional anesthesia; resource-limited settings; capacity building; global health

Introduction

Ultrasound-guided regional anesthesia (UGRA) has become the standard of care for safe and effective perioperative pain management in high-income countries. It is frequently used as the primary anesthetic in patients undergoing extremity surgeries [1]. Beyond the operating room, regional anesthesia is increasingly performed by nonanesthesia clinicians, including emergency physicians, to provide timely pain relief in acute care settings [2,3]. Ultrasound-guided peripheral nerve blocks (PNBs) are now used in emergency departments for conditions such as hip and rib fractures and wound management, demonstrating the growing interdisciplinary role of UGRA worldwide [4-6]. However, limited availability of ultrasound equipment and the lack of formal training programs contribute to the underuse of PNBs in low- and middle-income countries (LMICs) [7-10]. The emergence of portable, reliable, and cost-effective ultrasound devices offers new opportunities to expand the use of UGRA in diverse and resource-limited settings [11]. Advancing global regional anesthesia capacity can improve patient care by providing an alternative to general anesthesia, reducing cost and resource consumption, and enhancing patient comfort and safety [1,12].

Addressing the need for training opportunities is critical to the adoption of regional anesthesia techniques in LMICs. International partnerships help improve global capabilities by promoting educational activities and access to supplies. Although several initiatives have attempted to implement UGRA in diverse settings, few have been formally evaluated, and published evidence on their impact in LMICs remains limited [12]. Brouillet et al [13] described the successful development of an in-person educational program called the Global Regional Anesthesia Curricular Engagement (GRACE). In this program, anesthesia providers in Ghana were trained in UGRA techniques, demonstrating positive outcomes and increased procedure volume. Other collaborations have explored fully remote training models delivered via online communication platforms [14,15]. However, to our knowledge, the use of a hybrid training curriculum for UGRA, combining online didactic instruction with in-person, hands-on training, has not been described. Furthermore, most existing reports focus on programs based in African countries at large teaching hospitals with preexisting ultrasound capabilities and delivered in English. There is a paucity of evidence describing educational initiatives developed specifically for medical professionals in Latin America.

Our program was delivered in Guatemala, a Central American country where approximately 90% of the population relies on the public health sector, and approximately half of the country's 17 million inhabitants live in rural or impoverished communities [16]. Geographic and financial barriers limit access to advanced medical technologies such as ultrasound, which are largely concentrated in metropolitan areas [17]. Additionally, most physician training opportunities are offered in large urban hospitals near the capital of Guatemala City [17]. As a result, smaller public hospitals across the country have limited capacity and access to UGRA. Despite these challenges, there is substantial interest and growing demand for ultrasound-guided procedures nationwide.

<https://mededu.jmir.org/2026/1/e84181>

In the southwestern region of Guatemala, the Hospital Nacional de Coatepeque (HNC) serves as the regional public hospital, facing high surgical demands amid notable resource constraints. In 2023, a needs assessment was performed through a partnership between the department of anesthesiology at the University of Colorado (CU) and HNC and identified a lack of UGRA capabilities in the region. In response, the department of anesthesiology at the HNC requested a formal training program to build local expertise and improve the quality of perioperative care. Our objective was to develop a tailored hybrid training program to establish a sustainable UGRA service at HNC. This evaluation assessed the outcomes and impact of this pilot program. We hypothesized that the implementation of this hybrid training model would increase clinical knowledge and confidence, enhance skill acquisition, and enable the successful adoption of PNBs in this resource-limited setting in Guatemala.

Methods

Overview

We conducted a prospective pilot study of a novel pilot training program for UGRA delivered at HNC in Coatepeque, Guatemala, from April 2024 through June 2024. The evaluation design and materials were adapted from the GRACE model developed by Brouillet et al [13], with the authors' permission.

Setting

HNC is the only public surgical hospital in the town of Coatepeque, Quetzaltenango, Guatemala, and serves as the primary referral center for an estimated 150,000 people in this region. A large proportion of patients come from underserved, agriculture-based communities that experience some of the highest rates of poverty and poorest health outcomes in the country [18]. This hospital has 2 main operating rooms and an additional operating room in the obstetric unit for cesarean deliveries. Based on internal hospital records, approximately 5300 surgeries are performed annually across general surgery, obstetrics and gynecology, and orthopedics and traumatology. Trauma and orthopedic extremity procedures represent a major portion of surgical volume due to the high incidence of motor vehicle accidents in this area. Neuraxial anesthesia is used preferentially when feasible, although general anesthesia is typically required for upper extremity surgery. However, due to limitations in supplies, monitoring equipment, and postoperative nursing capacity at this hospital, the use of long-acting opioid analgesics, deep sedation, and general anesthesia is avoided when possible to optimize patient safety [19].

Needs Assessment

The initial site needs assessment was conducted in February 2023 and identified both strong interest and need for UGRA. A follow-up visit with faculty from CU in August 2023 focused on developing a collaborative action plan, gathering feedback, taking inventory of available supplies, and obtaining approval from hospital leadership. The primary barrier to implementation was the absence of an ultrasound machine due to limited purchasing access. The anesthesia department already

maintained a supply of nerve block needles and local anesthetics for infrequent use of nerve stimulation for distal lower extremity blocks. Standard emergency medications were available, except for lipid emulsions. However, no procedure logs or patient consent forms were in place because a regional anesthesia workflow had not yet been established.

To inform curriculum design, an online survey was distributed to HNC anesthesiologists to assess baseline experience and learning preferences ([Multimedia Appendix 1](#)). Before training, participants were asked to complete the questionnaire to capture demographic information, self-reported practices, educational needs, and motivations for participation.

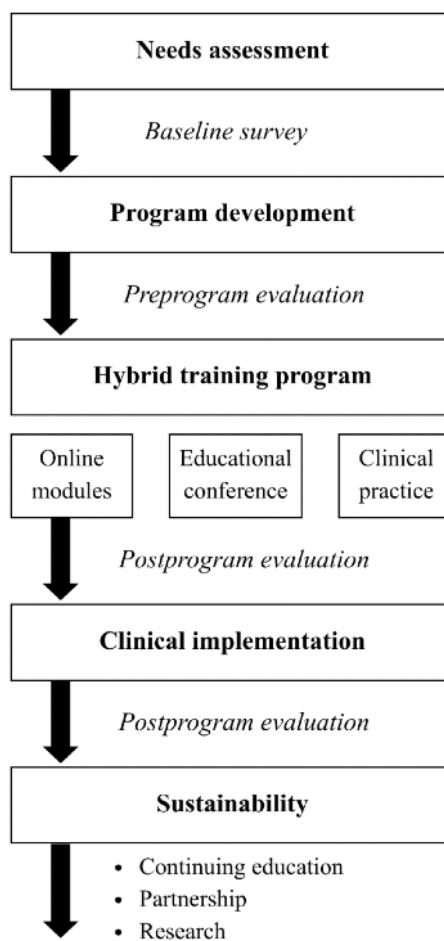
Program Design and Implementation

Curriculum Development

Information gathered from site assessments, planning meetings, and the online survey was used to develop a hybrid training

curriculum. The curriculum was designed collaboratively with local stakeholders and regional anesthesia specialists from CU and subsequently revised based on feedback. All 7 physician anesthesiologists at HNC were invited to participate and provided written consent. The program consisted of three core components: (1) online self-directed learning modules, (2) an in-person educational conference, and (3) clinical practice with direct supervision and bedside teaching ([Figure 1](#)). Training was conducted over a 2-month period from April 2024 to May 2024, with a follow-up site visit 4 months after clinical implementation. To simplify the program and emphasize high-yield applications, a limited number of blocks were selected based on the needs assessment and team discussions. These included interscalene and supraclavicular blocks of the brachial plexus, femoral, saphenous (adductor canal), and popliteal blocks.

Figure 1. Flowchart of the program design and evaluation.



Online Modules

The online learning curriculum was delivered as 4 weekly learning modules on a web-based platform to provide open access to independent learning. As no existing materials met our program objectives for topics or in the Spanish language, we created a dedicated website to address this need [[20](#)]. Content for weekly modules was derived from open access resources, such as publications from the New York School of Regional

Anesthesia ([Multimedia Appendix 2](#)). Additionally, the option for virtual chats or question-and-answer sessions was offered weekly on Fridays at the end of each module.

Educational Conference

The educational conference consisted of a 2-day, 16-hour in-person program that included lectures, clinical demonstrations, and hands-on workshops. Sessions were open to anesthesiologists, nurse anesthetists, nursing staff, and

surgeons (refer to [Multimedia Appendix 3](#) for conference outline).

Clinical Practice

Clinical practice was initiated following the conference, with HNC participants performing nerve blocks during work hours under the supervision of CU preceptors. The program faculty comprised multiple physician anesthesiologists with fellowship training or proficiency in regional anesthesia, ensuring the presence of at least one supervising preceptor during a 4-week period for continuity. Following 1 month of direct supervision, ongoing training and support were offered remotely to address emerging needs. A week-long follow-up visit was completed 4 months after clinical implementation to provide additional support and evaluations.

Equipment Donation

To support program initiation, CU faculty donated essential supplies, including a portable ultrasound machine, tablet, nerve block needles, gel, lidocaine, bupivacaine, and lipid emulsion. In response to site visits and internal requests, hospital leadership committed to increasing local supplies and purchased additional nerve block needles and local anesthetics. Additionally, operating room schedules of the surgical services were adjusted to increase time allotted to orthopedic surgery and increase PNB volume during the clinical practice phase.

Evaluation Framework

Evaluation Approach

The evaluation framework and instruments were adapted from the GRACE model, with approval from the study authors [13]. This model used the method proposed by Kirkpatrick and Kirkpatrick [21], a globally recognized and validated approach for evaluating the effectiveness of educational and training programs across disciplines [22]. This model evaluates outcomes across four levels: (1) reaction, (2) learning, (3) behavior, and (4) outcomes.

Reaction

Participant reactions were assessed through postprogram satisfaction surveys to solicit perceptions and feedback related to the training. This survey was adapted from the original GRACE satisfaction survey, which contained 10 statements where trainees rate their level of agreement on a 5-point Likert scale and 3 open-ended questions for program feedback and suggestions. Two open-ended questions were also added to describe the most and least helpful components of the training ([Multimedia Appendix 4](#)).

Learning

Participant learning was evaluated by analyzing changes in knowledge and clinical skills. Knowledge tests were administered before the program and 1 month after clinical implementation of UGRA. The knowledge test consisted of 24 multiple-choice questions that were adapted from the original GRACE instrument. Modifications were made to remove content not taught in this program, provide clarifications, and include items addressing identified knowledge gaps ([Multimedia Appendix 5](#)). This test was administered through an online

secure platform (Qualtrics XM Platform). Procedural skill acquisition was assessed using a validated global rating scale (GRS) for UGRA developed by Chuan et al [23]. This instrument evaluates 7 domains—preparation, respect for tissue, handling, time and motion, instrument handling, flow, knowledge, and overall performance—graded on a 5-point scale ([Multimedia Appendix 6](#)). Participants were observed during procedures and graded at program initiation, 1 month, and 4 months. To further capture participants' perceptions of the program, additional postsurvey questions were included to self-report changes in confidence, knowledge, and procedural skills using a 5-point Likert scale. All adapted GRACE evaluation instruments were translated into Spanish by 2 language-certified researchers and reviewed for cultural and linguistic accuracy by a local physician coinvestigator.

Behavior

Participants were prompted to report their level of interest, plans to continue performing PNBs, and perceived impact on patient care on the postprogram survey at program completion. The number of procedures performed by the anesthesiologist was recorded to assess adoption into practice.

Clinical Outcomes

Clinical outcomes were analyzed by total number and type of PNB, success and complication rate, and anesthesia method for 10 months after clinical implementation. Block success was defined a priori by the local department of anesthesiology at the HNC as a patient's ability to tolerate surgery without general anesthesia or an analgesic block with moderate to complete pain relief. To address the lack of workflow, a block documentation form and clinical log were developed in Spanish based on the New York School of Regional Anesthesia Universal Documentation Sheet for Peripheral Nerve Blocks to record and track procedure information, and patient consent forms were adapted from CU consent documentation ([Multimedia Appendices 7 and 8](#)) [24]. Additionally, a safety surveillance system was implemented to evaluate for complications and ensure resolution of the nerve block and lack of new neurological symptoms before discharge. Additional training was provided to nursing staff by a specialized CU postanesthesia care unit nurse to manage and identify complications in patients who received PNBs.

Statistical Analysis

Descriptive statistics were calculated and reported as frequencies, percentages, and means with SDs. Preprogram and postprogram knowledge test and GRS scores were summarized using medians and ranges. Changes in scores between preprogram and postprogram tests were assessed using paired Wilcoxon signed-rank tests. Likert scale outcomes across multiple time points were evaluated with the Friedman test, followed by pairwise Wilcoxon signed-rank tests for post hoc comparisons. Effect sizes for the Friedman test (Kendall W) and for Wilcoxon signed-rank tests were quantified with 95% CIs estimated using the percentile bootstrap method from 1000 resamples [25,26]. Standardized effect sizes were interpreted using Cohen general guidelines (0.2=weak, 0.5=medium, and 0.8=strong) while recognizing that these thresholds are context

dependent and should be interpreted cautiously. Interrater reliability (IRR) for preprogram GRS items, assessed across 2 fixed raters, was quantified using a 2-way mixed-effects intracorrelation coefficient and weighted κ , with 95% CIs obtained using the percentile bootstrap method from 1000 resamples. Two-sided P values $<.05$ were considered statistically significant, with post hoc comparisons corrected for false discovery rate. Given the small sample size, these results may not be generalizable and should be interpreted as exploratory. Statistical analyses were performed using R (version 4.5.1; R Foundation for Statistical Computing). The cohort sample size was determined by the number of anesthesiologists available to perform PNBS at HNC.

Ethical Considerations

This evaluation received approval from the local research ethics committee (001-2024) and was considered exempt by the Colorado Multiple Institutional Review Board (24-0212). All participants provided written informed consent before enrollment. To protect participant privacy, all data were

deidentified and managed in accordance with institutional data security standards, with access restricted to authorized study personnel. Results are reported in aggregate to prevent individual identification. Participants did not receive financial or other compensation for participation in the study.

Results

Overview

All anesthesiologists from the HNC completed the training, with a 100% (7/7) response rate for surveys and evaluations. Baseline participant characteristics are provided in [Table 1](#) (complete preprogram survey responses are provided in [Multimedia Appendix 9](#)). While all participants had previously performed PNBS using landmark or nerve stimulator techniques, none had previous experience with ultrasound guidance. Learning style preferences varied, although all participants favored workshops and hands-on clinical practice. Reported barriers included a lack of training opportunities and limited supplies.

Table 1. Baseline participant demographics (N=7).

Characteristic	Values
Participant demographics	
Participants, n (%)	7 (100)
Female, n (%)	4 (57.1)
Experience (y), mean (SD)	9.1 (5.1)
Have you used ultrasound previously? n (%)	
Yes	0 (0)
No	7 (100)
Do you have previous experience with blocks? n (%)	
Yes	7 (100)
No	0 (0)
What blocks have you performed?^a n (%)	
Supraclavicular	2 (28.6)
Axillary	5 (71.4)
Peripheral nerve (radial, medial, and ulnar)	4 (57.1)
Ankle	5 (71.4)
Sciatic	2 (28.6)
Scalp	1 (14.2)
What is your primary motivation for participating in this program?^a n (%)	
Learning new clinical skills	6 (85.7)
Patient outcomes	5 (71.4)
Work satisfaction	2 (28.6)
Patient satisfaction	4 (57.1)
Decreased resource use	3 (42.8)

^aMultiple responses were allowed per participant.

Each participant completed the full online curriculum, attended both in-person workshops, and performed multiple supervised

UGRA procedures during the clinical practice period. The number of blocks performed by each participant varied widely in the 10 months following implementation (range 14-41).

Reaction Outcomes

The overall satisfaction scores for the program were high (Table 2). The most frequently cited strengths were the integration of

Table 2. Participants' reactions to the training program.

Survey question	Score ^a , mean (SD)
Performing my own blocks during work hours was useful	4.71 (0.49)
The content of the program was satisfactory	4.86 (0.38)
The online study program helped prepare me for the in-person sessions	4.86 (0.38)
The conference presentations were useful	4.86 (0.38)
The program was applicable to my practice	5.00 (0)
The program should continue at HNC ^b	5.00 (0)
It was useful to practice with ultrasound during the workshops	5.00 (0)
The clinical teaching and supervision were helpful during procedures	5.00 (0)
The knowledge test was a fair evaluation of the course material	5.00 (0)
This program should be implemented in other hospitals	5.00 (0)

^aResponses rated using a 5-point Likert scale (1=strongly disagree; 5=strongly agree).

^bHNC: Hospital Nacional de Coatepeque.

Learning Outcomes

As shown in Figure 2, all participants demonstrated improved knowledge of UGRA. Knowledge test scores increased from a median of 45.8% (11/24; IQR 33.2%-54.0%) to 66.7% (16/24; IQR 58.3%-87.5%), representing a median absolute improvement of 20.8 (IQR 16.7%-41.7%; $P=.02$) percentage points with a large effect size ($r=0.90$, 95% CI 0.896-0.913). Similarly, all participants reported perceived gains in knowledge and confidence with UGRA (Table 3). Procedural skills improved following training across all domains, with median GRS scores increasing from 1.2 before the program to 3.4 at 1 month and 3.0 at 4 months (Figure 3). This corresponded to a median percentage increase of 147.1% (IQR 96.9%-197.3%;

theory with personalized hands-on practice. No component was identified as unhelpful; however, participants suggested increasing collaboration with surgical specialties and extending the duration of trainers on-site.

$P=.03$) at 1 month and 131.4% (IQR 90.5%-172.3%; $P=.04$) at 4 months, with large effect sizes at both time points ($r=0.90$, 95% CI 0.90-0.92 at 1 month; $r=0.91$, 95% CI 0.90-0.95 at 4 months). IRR for average preprogram scores showed moderate-to-high agreement with wide confidence intervals (intracorrelation coefficient=0.69, 95% CI 0.31-0.88; $\kappa=0.63$, 95% CI 0.47-0.93). The IRR for individual skills is provided in Multimedia Appendix 10, although the IRR could not be meaningfully calculated for some skills due to minimal variability. Only preprogram items had multiple raters due to the limited on-site availability of research personnel, providing preliminary evidence of measurement reliability. IRR was not assessed for postprogram items.

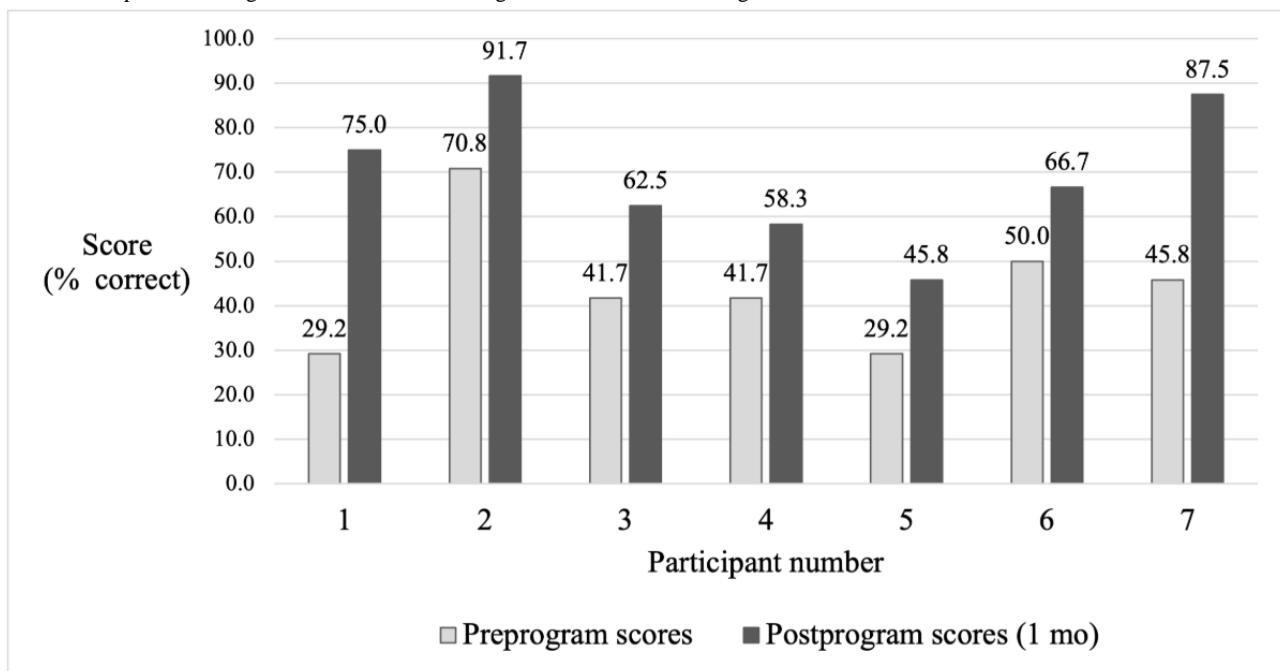
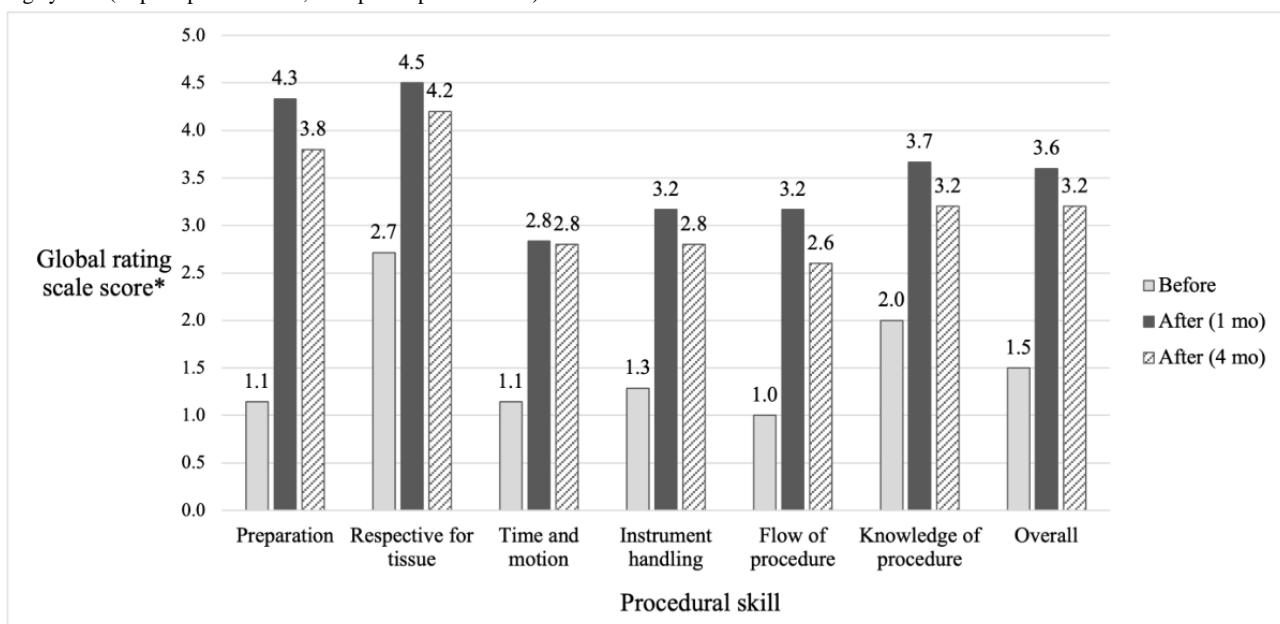
Figure 2. Participant knowledge test scores before training and 1 month after training.**Figure 3.** Change in procedural skills before and after the program at 1 and 4 months, measured by average global rating scale scores. *Using a 5-point scoring system (1=poor performance; 5=superior performance).

Table 3. Participants' self-reported changes after participating in the training program.

Survey question	Score ^a , mean (SD)
I feel that I have improved my ability to perform blocks	4.29 (0.76)
My knowledge of PNBs ^b has increased	4.43 (0.79)
I feel more comfortable using and interpreting ultrasound	4.57 (0.53)
I have a stronger desire to implement blocks in my practice	4.86 (0.38)
I plan to continue using PNBs in my practice	4.57 (0.53)
I feel that PNBs improve patient care at HNC ^c	5.00 (0)

^aResponses rated using a 5-point Likert scale (1=strongly disagree; 5=strongly agree).

^bPNB: peripheral nerve blocks.

^cHNC: Hospital Nacional de Coatepeque.

Behavior Outcomes

All participants reported high motivation to continue performing UGRA and intention to integrate PNBs into their practice (Table 3). All anesthesiologists continued to perform ultrasound-guided nerve blocks through the follow-up period after implementation.

Clinical Outcomes

A total of 54 blocks were performed in the first month under direct supervision, and 187 total procedures were recorded

during the 10 months following implementation (Table 4). Patient demographic and case information are provided in [Multimedia Appendix 11](#). The overall procedure success rate was 96.3% (180/187), with no reported patient complications. The supraclavicular block was the most frequently performed (66/187, 35.3%). General anesthesia was avoided in 95.9% (70/73) of the patients for upper limb surgery through successful brachial plexus blocks. On the postprogram survey, all 7 anesthesiologists strongly agreed that UGRA improves patient care at HNC.

Table 4. Ultrasound-guided nerve blocks performed during the first 10 months after program implementation by type (n=187).

Block type	Blocks, n (%)
Interscalene	10 (5.3)
Supraclavicular	66 (35.3)
Femoral	35 (18.7)
Adductor canal	25 (13.4)
Popliteal	40 (21.4)
Other	11 (5.9)

Discussion

Principal Findings

We describe the successful design and implementation of a hybrid training program at a resource-limited hospital in southwestern Guatemala. This program resulted in substantial growth in knowledge and technical skills, practice change, and enhanced patient care for a 10-month period. The educational model developed for this context offers a framework for introducing UGRA services in LMICs where clinical capabilities are lacking.

Educational Program Design

This program featured a unique educational setting, requiring input from local stakeholders to ensure alignment with daily clinical practice and local needs. Unlike previous initiatives, such as the original GRACE program, which aimed to expand preexisting UGRA capacity, anesthesiologists at HNC had no previous exposure to ultrasound or ultrasound-guided procedures. This lack of training presented challenges due to the steep learning curve and high-stakes environment in which

these procedures are performed, reflecting the reality for much of the anesthesia workforce in LMICs [12]. Additionally, the paucity of available Spanish-language UGRA resources necessitated the translation or development of all surveys, lectures, and training materials in Spanish.

Multiple approaches have been used to deliver education in UGRA [27]. International programs have relied on either in-person training or predominantly online or virtual formats, with variable success [13,15,28-30]. Each method has inherent advantages and limitations. In-person training provides direct interaction with learners, hands-on skill development, real-time adaptation and feedback, and access to resources and equipment but is constrained by travel, time away from clinical duties, short program duration, and higher costs [31]. Conversely, online-only training offers flexibility, continuity, lower costs, and the ability to record and revisit content. However, it is limited by reliance on visual learning, time-zone variability, and internet connectivity and lacks in-person teaching and supervision. The duality of our methodology leverages the benefits of both approaches while minimizing the challenges and limitations.

Given the limited exposure to UGRA and trainer availability, a hybrid design was selected to maximize learning and skill development. Initial self-directed learning was used to establish foundational knowledge, particularly for theoretical concepts such as ultrasound physics, pharmacology, and management of complications. Concepts were then reinforced with classroom instruction and followed by workshops and clinical practice to support technical skill acquisition. This “flipped classroom” approach prepares learners for hands-on experiential learning and has become a standard in medical education [32]. Evidence suggests that combining self-directed learning with expert clinical instruction reduces procedural errors and improves performance [33,34]. Using this multifaceted approach, HNC anesthesiologists achieved proficiency and confidence with UGRA after 1 month and, importantly, maintained these skills even after direct trainer support had ended. Blended learning has been effectively integrated into other medical contexts [35-38]. Increasing access and familiarity with e-learning tools and resources will continue to expand opportunities for student-centered medical education worldwide.

Impact on Patient Care

A notable impact of UGRA in this setting was the avoidance of general anesthesia for upper limb surgery through ultrasound-guided brachial plexus blocks. This was achieved in 70 (95.9%) of the 73 patients over the 10-month follow-up period. Regional anesthesia has well-documented advantages over general anesthesia, such as fewer perioperative respiratory events, reduced opioid use, superior pain management, and improvement in some procedure-specific outcomes [39]. These benefits are more pronounced in resource-limited settings, where constraints in supplies, equipment, and staffing reduce safety margins. Large datasets from Médecins Sans Frontières (Doctors Without Borders) facilities in LMICs similarly reported lower mortality with regional anesthesia in comparison to general endotracheal anesthesia [40]. Analgesic blocks were also performed during training and clinical practice and were perceived by both patients and surgeons to provide superior pain control compared with the existing nonopioid regimens, although formal evaluation is warranted. Nevertheless, all anesthesia participants unanimously agreed on the postprogram survey that PNBs improved patient care at their institution.

Patient safety was a priority for this pilot program. Ultrasound-based interventions require careful technique to achieve consistent and safe outcomes. Given that improper ultrasound application or technique may lead to unintended physiological effects or nerve injury, hands-on training becomes essential to ensure safe use [41,42]. In our program, no complications were observed, and the block failure rate of 3.7% was consistent or superior compared to published rates ranging from 5% to 10% [43-45].

Lessons Learned

Implementing a regional anesthesia service at a resource-limited public hospital in Guatemala required adaptability and creative problem-solving. The preoperative workflow presented challenges, as patients typically arrived directly at the operative area with minimal preparation time, limiting the window for block placement and onset. Through pharmacological

adjustments, the combination of the shorter-acting lidocaine with bupivacaine anecdotally accelerated the onset of blocks. Consent was typically not obtained by the anesthesia team; therefore, a new process was established to obtain authorization for procedures, especially parental consent in the case of minors, where parents were required to provide consent. In patients who were illiterate, we integrated the local protocol of reading the consent form information, allowing for questions, and then obtaining fingerprint stamps on forms. Technical limitations with portable ultrasound devices further complicated implementation. The prolonged procedure times during skill development led to probe overheating or battery depletion, and limited penetration of deeper structures hindered visualization in some blocks. The anesthesiologists’ unique 24-hour schedule every sixth day also required careful planning to ensure consistent practice and skill reinforcement. Despite variation in the number of blocks performed by trainees, consistent improvements in knowledge and skills were observed across all participants. Finally, establishing this new service necessitated collaboration with departments beyond anesthesiology and surgery to address issues involving supply chain, documentation processes, emergency preparedness protocols, postoperative monitoring, and nursing education on the wards.

Multiple strategic decisions contributed to the program’s success. First, the smaller cohort size, extended trainer presence on-site, and longer follow-up period increased direct interaction with experts, maximizing learning and enabling assessment of clinical practice and outcomes. Second, multidisciplinary coordination allowed for adjustment of the surgical schedule to dedicate one operating room to orthopedics, increasing block volume and creating greater opportunities for supervised practice. Third, limiting the focus to a few high-yield blocks proved effective for trainees with minimal baseline experience, as it allowed for focused application of learning and consolidation of skills. As proficiency grew with these limited blocks and ultrasound maneuvering, we then observed the adoption of new blocks among the group at HNC. Fourth, the development of data collection tools, such as the procedure log, assisted with real-time monitoring of progress, quality improvement, supply procurement, and research efforts. Finally, active engagement with hospital leadership and the Ministry of Public Health and Welfare, Guatemala not only reinforced the program’s value but also generated public recognition, laying the foundation for potential expansion across the national health system.

Future Directions

While the initial results are encouraging, future efforts must continue to prioritize sustainability, ensuring progress is both durable and scalable. This includes implementing strategies for long-term follow-up, identifying and training local experts, and maintaining buy-in from local stakeholders [12]. Our group has engaged in regular communication and quarterly visits to reinforce relationships and provide mentorship. A long-term partnership with the CU Global Anesthesia program will include ongoing peer-to-peer support, faculty engagement, research collaborations, and resident rotations. Ultimately, the shared objective of this program is to develop local experts who can

then train and support colleagues in the region. This “train-the-trainer” model has the potential to foster a self-sustaining system of local expertise and training cascades to build long-term capacity within the health system [46,47]. Efforts are underway to expand UGRA training to residency programs, equipping anesthesiologists early in their careers and integrating these techniques into both public and private practice, thereby expanding the scope and consistency of anesthesia care nationwide. Additionally, given the growing role of UGRA, future training initiatives could include collaboration with emergency physicians at HNC to integrate ultrasound-guided nerve block techniques into emergency care, allowing patients to benefit from regional anesthesia earlier in their clinical course.

Collaboration with local stakeholders, including nursing staff, surgeons, and hospital administration, is essential to the long-term success of this program. Hospital leadership engagement is critical to ensure the purchasing of supplies and medications needed to maintain a successful and safe UGRA program. To sustain interdisciplinary commitment, research should prioritize patient satisfaction, clinical outcomes, system-level impact, and cost analyses, thereby demonstrating value in patient care and resource optimization [48].

These data can also support efforts to increase access to ultrasound technology. Although donation-based models have helped bridge initial gaps, there remains a need for improved purchasing mechanisms and sustainable supply chains for portable devices in Guatemala and other Latin American countries where demand is high. Advances in technology continue to create new opportunities for training and patient safety, and recent studies demonstrate that technology-assisted models can enhance clinical decision-making and expand training capacity across multiple medical disciplines, even in resource-limited settings [49]. Technology-driven clinical training models increase accessibility, standardization, and clinical accuracy. This hybrid UGRA program, combining structured digital content with hands-on practice, offers a pragmatic approach to overcoming training barriers in LMICs.

Limitations

There are multiple limitations to this study. First, training was delivered at a single hospital site with a small cohort of anesthesiologists. While this helped ensure quality control and exposure to practice with trainers, this limits generalizability to other departments, residency programs, or hospital systems. Second, there is a risk of information biases inherent in the

observational study methods, such as recall or observer bias. Self-reporting may yield incomplete or inaccurate data; assessments of technical competency were performed by the training team with single raters for follow-up assessments, which may introduce expectancy and Hawthorne bias. Future studies should incorporate blinded, independent assessors or video adjudication to enhance objectivity and IRR. Additionally, although evaluation materials were adapted from validated sources, the translated GRACE instruments were not formally validated for cultural or psychometric equivalence, which may limit measurement accuracy. Program fidelity measures, such as detailed module completion analytics, were not available for all participants, which may limit precise quantification of exposure. However, attendance and active participation were confirmed for all anesthesiologists.

The costs of delivering in-person training can be prohibitive and should be considered when planning similar educational initiatives. Previous UGRA programs have reported costs up to US \$5000 per trainer per trip [13]. This program minimized costs by integrating the hybrid educational model. Thus, costs were lower and estimated to be an average of US \$1500 to US \$2000 per trainer per trip. We believe that this model offers a pragmatic and cost-efficient approach for future international collaborations.

Conclusions

This evaluation demonstrates the successful implementation of a hybrid training model for UGRA at a public hospital in Guatemala, advancing perioperative pain management in a resource-limited setting. By combining online modules, in-person didactics, and supervised clinical practice, the program achieved substantial improvements in anesthesiologists’ knowledge, procedural skills, and confidence. Despite a lack of previous capabilities, this program resulted in the adoption of UGRA, with sustained clinical practice observed over a 10-month follow-up period. These outcomes reflect both a tailored curriculum and strong local engagement, highlighting the model’s value, safety, and potential scalability in LMICs. The avoidance of general anesthesia in upper limb surgeries underscores UGRA’s potential to improve patient safety and optimize resources within the public health system. Future efforts should prioritize local expert development, interdisciplinary collaboration, and increased access to ultrasound devices to ensure the long-term sustainability and expansion of UGRA, thereby transforming anesthesiology care for underserved communities across Guatemala.

Acknowledgments

The authors thank Drs Rafael Menes and Juan Nájera, director and vice director of Hospital Nacional de Coatepeque, for their support in this program and Dr Kimberly Mendoza for her contributions. The authors also thank the Global Outreach Program of Butterfly Network, Inc, for their support.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this paper.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: JEG (lead), NGSD (supporting), CGS (supporting)

Formal analysis: JEG (lead), IT-D (supporting), CGS (supporting)

Investigation: JEG (lead), NGSD (supporting), CGS (supporting)

Methodology: JEG, IT-D (supporting), NGSD (supporting), ABH (supporting)

Project administration: JEG (lead), NGSD (equal), CGS (supporting)

Writing—original draft: JEG (lead), IT-D (supporting), RF (supporting), MAV (supporting), MAG (supporting), ABH (supporting), CGS (supporting)

Writing—review and editing: JEG (lead), IT-D (supporting), CGS (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1

Preprogram participant survey.

[\[DOCX File, 19 KB - mededu_v12i1e84181_app1.docx\]](#)

Multimedia Appendix 2

Curriculum outline of online learning modules.

[\[DOCX File, 18 KB - mededu_v12i1e84181_app2.docx\]](#)

Multimedia Appendix 3

Outline of educational conference topics.

[\[DOCX File, 17 KB - mededu_v12i1e84181_app3.docx\]](#)

Multimedia Appendix 4

Postprogram participant survey.

[\[DOCX File, 19 KB - mededu_v12i1e84181_app4.docx\]](#)

Multimedia Appendix 5

Ultrasound-guided regional anesthesia knowledge test.

[\[DOCX File, 8133 KB - mededu_v12i1e84181_app5.docx\]](#)

Multimedia Appendix 6

Global rating scale for procedural skills.

[\[DOCX File, 17 KB - mededu_v12i1e84181_app6.docx\]](#)

Multimedia Appendix 7

Procedure documentation form.

[\[PDF File \(Adobe PDF File\), 42 KB - mededu_v12i1e84181_app7.pdf\]](#)

Multimedia Appendix 8

Procedure consent form.

[\[DOCX File, 20 KB - mededu_v12i1e84181_app8.docx\]](#)

Multimedia Appendix 9

Results of preprogram participant survey.

[\[DOCX File, 19 KB - mededu_v12i1e84181_app9.docx\]](#)

Multimedia Appendix 10

Interrater reliability for global rating scale assessment scores.

[\[DOCX File, 16 KB - mededu_v12i1e84181_app10.docx\]](#)

Multimedia Appendix 11

Characteristics of patients and procedures performed during the ultrasound-guided regional anesthesia training program.

[[DOCX File , 17 KB - mededu_v12i1e84181_app11.docx](#)]

References

1. Albrecht E, Chin KJ. Advances in regional anaesthesia and acute pain management: a narrative review. *Anaesthesia* 2020 Jan;75 Suppl 1:e101-e110. [doi: [10.1111/anae.14868](#)] [Medline: [31903582](#)]
2. Resta F, Barcella B, Angelini V, Lago E, Santaniello A, Dedato AS, et al. Simulation-based training in ultrasound-guided regional anaesthesia for emergency physicians: insights from an Italian pre/post intervention study. *BMC Med Educ* 2024 Dec 20;24(1):1510 [FREE Full text] [doi: [10.1186/s12909-024-06500-0](#)] [Medline: [39707369](#)]
3. Tucker RV, Peterson WJ, Mink JT, Taylor LA, Leech SJ, Nagdev AD, et al. Defining an ultrasound-guided regional anesthesia curriculum for emergency medicine. *AEM Educ Train* 2021 Jul;5(3):e10557 [FREE Full text] [doi: [10.1002/aet.2.10557](#)] [Medline: [34124505](#)]
4. Di Pietro S, Maffei R, Jannelli E, Mascia B, Resta F, De Silvestri A, et al. Comparing the pericapsular nerve group block and fascia iliaca block for acute pain management in patients with hip fracture: a randomised clinical trial. *Anaesthesia* 2025 Dec;80(12):1484-1492. [doi: [10.1111/anae.16695](#)] [Medline: [40727959](#)]
5. Di Pietro S, Mascia B, Lo Bianco G, Perlini S, Iotti GA. Anterior cutaneous nerve block for analgesia in anterior chest trauma: is the parasternal approach necessary? *Clin Exp Emerg Med* 2020 Mar;7(1):67-70 [FREE Full text] [doi: [10.15441/ceem.18.089](#)] [Medline: [32252136](#)]
6. Di Pietro S, Caracciolo E, Barcella B, Perlini S, Regional Anaesthesia in Emergency Medicine Research Group. Superficial cervical plexus block in emergency departments: rationale for its use in incision and drainage of neck skin abscesses. *Intern Emerg Med* 2022 Aug;17(5):1533-1536. [doi: [10.1007/s11739-022-02992-0](#)] [Medline: [35670893](#)]
7. Schnittger T. Regional anaesthesia in developing countries. *Anaesthesia* 2007 Dec;62 Suppl 1:44-47 [FREE Full text] [doi: [10.1111/j.1365-2044.2007.05297.x](#)] [Medline: [17937713](#)]
8. Rukewe A, Fatiregun A. The use of regional anaesthesia by anaesthesiologists in Nigeria. *Anesth Analg* 2010 Jan 01;110(1):243-244. [doi: [10.1213/ANE.0b013e3181c0f00e](#)] [Medline: [19933532](#)]
9. Ho M, Livingston P, Bould MD, Nyandwi JD, Nizeyimana F, Uwineza JB, et al. Barriers and facilitators to implementing a regional anesthesia service in a low-income country: a qualitative study. *Pan Afr Med J* 2019;32:152 [FREE Full text] [doi: [10.11604/pamj.2019.32.152.17246](#)] [Medline: [31303923](#)]
10. Admassie BM, Admass BA, Melesse DY. Practice and challenges related to regional anesthesia in Amhara regional hospitals, Northwest-Ethiopia: a web-based survey study. *BMC Anesthesiol* 2024 Nov 04;24(1):398 [FREE Full text] [doi: [10.1186/s12871-024-02783-4](#)] [Medline: [39491037](#)]
11. Ranger BJ, Bradburn E, Chen Q, Kim M, Noble JA, Papageorghiou AT. Portable ultrasound devices for obstetric care in resource-constrained environments: mapping the landscape. *Gates Open Res* 2023;7:133 [FREE Full text] [doi: [10.12688/gatesopenres.15088.2](#)] [Medline: [3935587](#)]
12. Dohlman LE, Kwikiriza A, Ehie O. Benefits and barriers to increasing regional anesthesia in resource-limited settings. *Local Reg Anesth* 2020;13:147-158 [FREE Full text] [doi: [10.2147/LRA.S236550](#)] [Medline: [33122941](#)]
13. Brouillet MA, Aidoo AJ, Hondras MA, Boateng NA, Antwi-Kusi A, Addison W, et al. Regional anesthesia training model for resource-limited settings: a prospective single-center observational study with pre-post evaluations. *Reg Anesth Pain Med* 2020 Jul;45(7):528-535. [doi: [10.1136/rapm-2020-101550](#)] [Medline: [32447288](#)]
14. Dohlman LE, Thakkar N, Jivanelli B, Pakala S, Brouillet MA, Global Regional Anesthesia and Pain Management Research Group. Regional anesthesia global health collaborations- a scoping review of current intervention methods. *Curr Opin Anaesthesiol* 2022 Oct 01;35(5):647-653. [doi: [10.1097/ACO.0000000000001178](#)] [Medline: [35942724](#)]
15. Liu M, Salmon M, Zaidi R, Nagdev A, Debebe F, Muller MF, et al. Ultrasound-guided regional anesthesia: feasibility and effectiveness of teaching via telesimulation in Ethiopia. *Reg Anesth Pain Med* 2021 Aug;46(8):722-726. [doi: [10.1136/rapm-2020-102394](#)] [Medline: [33903219](#)]
16. Avila C, Bright R, Gutierrez J, Hoadley K, Manuel C, Romero N, et al. Guatemala health system assessment 2015. Health Finance & Governance Project. 2015. URL: <https://internationalmedicalrelief.org/wp-content/uploads/2019/07/Guatemala-HSA- ENG-FULL-REPORT-FINAL-APRIL-2016.pdf> [accessed 2025-12-25]
17. Zha Y, Truché P, Izquierdo E, Zimmerman K, de Izquierdo S, Lipnick MS, et al. Assessment of anesthesia capacity in public surgical hospitals in Guatemala. *Anesth Analg* 2021 Feb 01;132(2):536-544. [doi: [10.1213/ANE.0000000000005297](#)] [Medline: [33264116](#)]
18. Gamboa JE, Bolaños AG, Simmons CG. Community perceptions of accessing surgical and anesthetic care in rural Guatemala: a cross-sectional survey. *Cureus* 2025 Jun;17(6):e85783. [doi: [10.7759/cureus.85783](#)] [Medline: [40656399](#)]
19. Turner RA, Simmons CG, Ramirez S, Gamboa JE. Hypoxemia and postoperative monitoring after anesthesia: a prospective observational study using portable pulse oximetry in a resource-limited setting in Guatemala. *Cureus* 2025 Jan;17(1):e78075. [doi: [10.7759/cureus.78075](#)] [Medline: [40013179](#)]
20. Programa de estudio. Regional anestesia Guatemala. URL: <https://anestesiaregionalguatemala.com> [accessed 2025-06-01]

21. Kirkpatrick JD, Kirkpatrick WK. Kirkpatrick's Four Levels of Training Evaluation. Alexandria, VA: Association for Talent Development; 2016.
22. Anderson LN, Merkebu J. The Kirkpatrick model: a tool for evaluating educational research. Fam Med 2024 Jun;56(6):403 [FREE Full text] [doi: [10.22454/FamMed.2024.161519](https://doi.org/10.22454/FamMed.2024.161519)] [Medline: [38996342](#)]
23. Chuan A, Graham PL, Wong DM, Barrington MJ, Auyong DB, Cameron AJ, et al. Design and validation of the Regional Anaesthesia Procedural Skills Assessment Tool. Anaesthesia 2015 Dec;70(12):1401-1411 [FREE Full text] [doi: [10.1111/anae.13266](https://doi.org/10.1111/anae.13266)] [Medline: [26558857](#)]
24. Gandhi K, Patel V, Maliakal T, Xu D, Flisinski K. Universal documentation sheet for peripheral nerve blocks. J NY Sch Reg Anest 2009 May;12:23 [FREE Full text]
25. Rosenthal R. Meta-Analytic Procedures for Social Research. Thousand Oaks, CA: SAGE Publications; 1991.
26. Rosenthal R. Science and ethics in conducting, analyzing, and reporting psychological research. Psychol Sci 1994 May 01;5(3):127-134. [doi: [10.1111/j.1467-9280.1994.tb00646.x](https://doi.org/10.1111/j.1467-9280.1994.tb00646.x)] [Medline: [11652978](#)]
27. Nix CM, Margarido CB, Awad IT, Avila A, Cheung JJ, Dubrowski A, et al. A scoping review of the evidence for teaching ultrasound-guided regional anesthesia. Reg Anesth Pain Med 2013;38(6):471-480. [doi: [10.1097/AAP.0b013e3182a4ed7a](https://doi.org/10.1097/AAP.0b013e3182a4ed7a)] [Medline: [24108249](#)]
28. Moll V, Schmidt PC, Amos C, Workneh RS, Tadesse M, Mulugeta E, et al. Building regional anesthesia capacity in limited-resource settings: a pilot study evaluating a 4-week curriculum. Pain Manag 2021 Jan;11(1):29-37. [doi: [10.2217/pmt-2020-0044](https://doi.org/10.2217/pmt-2020-0044)] [Medline: [33073715](#)]
29. Burckett-St Laurent DA, Cunningham MS, Abbas S, Chan VW, Okrainec A, Niazi AU. Teaching ultrasound-guided regional anesthesia remotely: a feasibility study. Acta Anaesthesiol Scand 2016 Aug;60(7):995-1002. [doi: [10.1111/aas.12695](https://doi.org/10.1111/aas.12695)] [Medline: [26860837](#)]
30. Pfenninger EG, Tugtekin I, Stahl W, Dinse A, Gorsewski G, Vicent O, et al. "Anesthesia-focused sonography": first analysis of transferring a training from Germany to Ghana. SDRP J Anesth Surg 2018;3(1):1-9 [FREE Full text] [doi: [10.25177/jas.3.1.3](https://doi.org/10.25177/jas.3.1.3)]
31. Maloney S, Nicklen P, Rivers G, Foo J, Ooi YY, Reeves S, et al. A cost-effectiveness analysis of blended versus face-to-face delivery of evidence-based medicine to medical students. J Med Internet Res 2015 Jul 21;17(7):e182 [FREE Full text] [doi: [10.2196/jmir.4346](https://doi.org/10.2196/jmir.4346)] [Medline: [26197801](#)]
32. Phillips J, Wiesbauer F. The flipped classroom in medical education: a new standard in teaching. Trends Anaesth Crit Care 2022 Feb;42:4-8 [FREE Full text] [doi: [10.1016/j.tacc.2022.01.001](https://doi.org/10.1016/j.tacc.2022.01.001)] [Medline: [38620968](#)]
33. de Oliveira Filho GR, Mettrau FA. The effect of high-frequency, structured expert feedback on the learning curves of basic interventional ultrasound skills applied to regional anesthesia. Anesth Analg 2018 Mar;126(3):1028-1034. [doi: [10.1213/ANE.0000000000002748](https://doi.org/10.1213/ANE.0000000000002748)] [Medline: [29261545](#)]
34. Udani AD, Macario A, Nandagopal K, Tanaka MA, Tanaka PP. Simulation-based mastery learning with deliberate practice improves clinical performance in spinal anesthesia. Anesthesiol Res Pract 2014;2014:659160 [FREE Full text] [doi: [10.1155/2014/659160](https://doi.org/10.1155/2014/659160)] [Medline: [25157263](#)]
35. Koch R, Gassner L, Gerlach N, Festl-Wietek T, Hirt B, Joos S, et al. Integrated e-learning for shoulder anatomy and clinical examination skills in first-year medical students: randomized controlled trial. JMIR Med Educ 2025 Sep 17;11:e62666. [doi: [10.2196/62666](https://doi.org/10.2196/62666)] [Medline: [40961369](#)]
36. Vallée A, Blacher J, Cariou A, Sorbets E. Blended learning compared to traditional learning in medical education: systematic review and meta-analysis. J Med Internet Res 2020 Aug 10;22(8):e16504 [FREE Full text] [doi: [10.2196/16504](https://doi.org/10.2196/16504)] [Medline: [32773378](#)]
37. Tudor Car L, Kyaw BM, Dunleavy G, Smart NA, Semwal M, Rotgans JI, et al. Digital problem-based learning in health professions: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Feb 28;21(2):e12945. [doi: [10.2196/12945](https://doi.org/10.2196/12945)] [Medline: [30816846](#)]
38. 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 May 26;11:e65220. [doi: [10.2196/65220](https://doi.org/10.2196/65220)] [Medline: [40418795](#)]
39. Hughey S, Cole J, Drew B, Brust A, Stedjelarsen E. Regional anesthesia in resource-limited and disaster environments: a daring discourse. Reg Anesth Pain Med 2025 Aug 05;50(8):686-689. [doi: [10.1136/rapm-2024-105680](https://doi.org/10.1136/rapm-2024-105680)] [Medline: [38942425](#)]
40. Ariyo P, Trelles M, Helmand R, Amir Y, Hassani GH, Mftavyanka J, et al. Providing anesthesia care in resource-limited settings: a 6-year analysis of anesthesia services provided at Médecins Sans Frontières facilities. Anesthesiology 2016 Mar;124(3):561-569. [doi: [10.1097/ALN.0000000000000985](https://doi.org/10.1097/ALN.0000000000000985)] [Medline: [26881395](#)]
41. Ozsoy U, Ogut E, Sekerci R, Hizay A, Rink S, Angelov DN. Effect of pulsed and continuous ultrasound therapy on the degree of collateral axonal branching at the lesion site, polyinnervation of motor end plates, and recovery of motor function after facial nerve reconstruction. Anat Rec (Hoboken) 2019 Aug;302(8):1314-1324 [FREE Full text] [doi: [10.1002/ar.24122](https://doi.org/10.1002/ar.24122)] [Medline: [30950229](#)]
42. El-Tallawy SN, Ahmed RS, Salem GI, Alzahrani TA, Haddara MM, Ahmed RH, et al. Neurological deficits following regional anesthesia and pain interventions: reviewing current standards of care. Pain Ther 2025 Jun;14(3):817-839. [doi: [10.1007/s40122-025-00726-6](https://doi.org/10.1007/s40122-025-00726-6)] [Medline: [40146371](#)]

43. Barrington MJ, Watts SA, Gledhill SR, Thomas RD, Said SA, Snyder GL, et al. Preliminary results of the Australasian Regional Anaesthesia Collaboration: a prospective audit of more than 7000 peripheral nerve and plexus blocks for neurologic and other complications. *Reg Anesth Pain Med* 2009;34(6):534-541. [doi: [10.1097/aap.0b013e3181ae72e8](https://doi.org/10.1097/aap.0b013e3181ae72e8)] [Medline: [19916206](https://pubmed.ncbi.nlm.nih.gov/19916206/)]
44. Bottomley T, Gadsden J, West S. The failed peripheral nerve block. *BJA Educ* 2023 Mar;23(3):92-100 [FREE Full text] [doi: [10.1016/j.bjae.2022.12.002](https://doi.org/10.1016/j.bjae.2022.12.002)] [Medline: [36844443](https://pubmed.ncbi.nlm.nih.gov/36844443/)]
45. Cotter JT, Nielsen KC, Guller U, Steele SM, Klein SM, Greengrass RA, et al. Increased body mass index and ASA physical status IV are risk factors for block failure in ambulatory surgery - an analysis of 9,342 blocks. *Can J Anaesth* 2004 Oct;51(8):810-816. [doi: [10.1007/BF03018454](https://doi.org/10.1007/BF03018454)] [Medline: [15470170](https://pubmed.ncbi.nlm.nih.gov/15470170/)]
46. Mormina M, Pinder S. A conceptual framework for training of trainers (ToT) interventions in global health. *Global Health* 2018 Oct 22;14(1):100 [FREE Full text] [doi: [10.1186/s12992-018-0420-3](https://doi.org/10.1186/s12992-018-0420-3)] [Medline: [30348183](https://pubmed.ncbi.nlm.nih.gov/30348183/)]
47. Anderson CR, Taira BR. The train the trainer model for the propagation of resuscitation knowledge in limited resource settings: a systematic review. *Resuscitation* 2018 Jun;127:1-7. [doi: [10.1016/j.resuscitation.2018.03.009](https://doi.org/10.1016/j.resuscitation.2018.03.009)] [Medline: [29545135](https://pubmed.ncbi.nlm.nih.gov/29545135/)]
48. Mohamed SS, Temu R, Komba LF, Kaino MM, Olotu FI, Ndebea AS, et al. Patient satisfaction with, and outcomes of, ultrasound-guided regional anesthesia at a referral hospital in Tanzania: a cross-sectional study. *Anesth Analg* 2024 Jan 01;138(1):180-186. [doi: [10.1213/ANE.0000000000006374](https://doi.org/10.1213/ANE.0000000000006374)] [Medline: [36727852](https://pubmed.ncbi.nlm.nih.gov/36727852/)]
49. Ogut E. Artificial intelligence in clinical medicine: challenges across diagnostic imaging, clinical decision support, surgery, pathology, and drug discovery. *Clin Pract* 2025 Sep 16;15(9):169. [doi: [10.3390/clinpract15090169](https://doi.org/10.3390/clinpract15090169)] [Medline: [41002784](https://pubmed.ncbi.nlm.nih.gov/41002784/)]

Abbreviations

CU: University of Colorado

GRACE: Global Regional Anesthesia Curricular Engagement

HNC: Hospital Nacional de Coatepeque

IRR: Interrater reliability

LMIC: low- and middle-income country

PNB: peripheral nerve block

UGRA: ultrasound-guided regional anesthesia

Edited by J Gentges; submitted 16.Sep.2025; peer-reviewed by S Di Pietro, E Ogut; comments to author 09.Oct.2025; revised version received 17.Dec.2025; accepted 22.Dec.2025; published 08.Jan.2026.

Please cite as:

Gamboa JE, Tamm-Daniels I, Flores R, Sarat Diaz NG, Villasenor MA, Gist MA, Hoie AB, Kurinec C, Simmons CG

Ultrasound-Guided Regional Anesthesia in a Resource-Limited Hospital: Prospective Pilot Study of a Hybrid Training Program

JMIR Med Educ 2026;12:e84181

URL: <https://mededu.jmir.org/2026/1/e84181>

doi: [10.2196/84181](https://doi.org/10.2196/84181)

PMID: [41428829](https://pubmed.ncbi.nlm.nih.gov/41428829/)

©Jakob E Gamboa, Inge Tamm-Daniels, Roland Flores, Nancy G Sarat Diaz, Mario A Villasenor, Mitchell A Gist, Aidan B Hoie, Christopher Kurinec, Colby G Simmons. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 08.Jan.2026. 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.

Original Paper

Exploring the Icarus Paradox in Indonesia's Specialist Medical Education System Using the Public Perspective From Online Media: Convergent Mixed Methods Study

Faisal Binsar¹, DR (AM); Mohammad Hamsal², Prof Dr

¹Management Department, Binus Online Learning, Binus University, Jakarta, Jakarta, Indonesia

²Binus University, Binus Business School Doctor of Research in Management, Binus University, Jakarta, Jakarta, Indonesia

Corresponding Author:

Faisal Binsar, DR (AM)

Management Department

Binus Online Learning

Binus University

Jl K H Syahdan No 9

Kemanggisan, Palmerah

Jakarta, Jakarta, 11480

Indonesia

Phone: 62 215345830

Fax: 62 215300244

Email: faisal.binsar@binus.ac.id

Abstract

Background: The Icarus Paradox in health care refers to the tension between the ambition to succeed as a specialist doctor and the limitations of the medical education system. Indonesia aspires to produce quality doctors, yet limited infrastructure and resources hinder the educational journey of prospective specialists.

Objective: This study aimed to identify the Icarus Paradox in Indonesia's specialist medical education by examining prospective specialist medical students and the quality of health services and by analyzing how this paradox is reflected in society's perspectives.

Methods: Using a convergent mixed methods design, this study integrated quantitative content analysis of 5047 online reviews across multiple platforms with qualitative thematic and cognitive analysis using NVivo 14, combining sentiment classification and topic coding.

Results: Twitter contributed 573 (11.3%) of 5047 reviews, with 218 (38%) negative, 251 (43.8%) neutral, and 104 (18.2%) positive entries. TikTok generated 282 (5.6%) reviews, the majority being neutral (n=225, 79.5%). YouTube produced 96 (1.9%) reviews, with 89 (92.7%) neutral entries. News platforms exhibited the largest volume (n=3040, 60.2%) of reviews, with 2885 (94.9%) neutral, 105 (3.5%) positive, and 50 (1.6%) negative entries. Blogs and websites contributed 353 (7%) and 692 (11.3%) reviews, respectively, with neutral sentiment dominating (n=329, 93.2%, for blogs and n=599, 86.6%, for websites). Three cognitive perspectives demonstrated the Icarus Paradox in the Indonesian medical education system: education system, society's views of students, and health care services. Although there are aspirations to improve education and health care quality, these ambitions often collide with structural challenges, such as resource shortages, heavy workloads, and limited accessibility, which link directly to cognitive themes of stress, resilience, and ethical dilemmas. We proposed a conceptual model to illustrate these dynamics.

Conclusions: Our findings offer insights into the Icarus Paradox in Indonesia's medical education system, highlighting its complexity and reinforcing the need for systemic reform. Beyond academic relevance, the findings also emphasize the importance of strengthening student mental health support, ensuring equitable access to health care, and enhancing regulatory oversight of training. This was not a clinical trial. Although limited by reliance on online reviews, the results underscore the urgent need for targeted policy interventions in medical education and health care services.

Trial Registration: ClinicalTrials.gov registration: NCT123456

(*JMIR Med Educ* 2026;12:e60452) doi:[10.2196/60452](https://doi.org/10.2196/60452)

KEYWORDS

Icarus Paradox; aspirations; realities; medical education; student well-being; community perspective; online reviews

Introduction

Background

The Icarus Paradox, originally used in management studies to explain how strengths can become liabilities, offers a useful lens to examine paradoxical dynamics in specialist medical education [1]. Although such programs are designed to produce highly competent professionals, they often generate conditions that undermine performance and well-being. These paradoxes arise from three interconnected sources.

First, systemic educational design—marked by rigid curricula, competitive entry pathways, and limited infrastructure—restricts flexibility and adaptability. Second, entrenched institutional cultures characterized by hierarchy, conformity, and normalization of bullying shape learning environments where silence is often preferred over expression [2-6]. Third, the expectation that trainees endure prolonged working hours, intensive workloads, and psychological pressure with minimal support places substantial demands on individual resilience. These interactions illustrate how a system striving for excellence may inadvertently create vulnerabilities that undermine its goals.

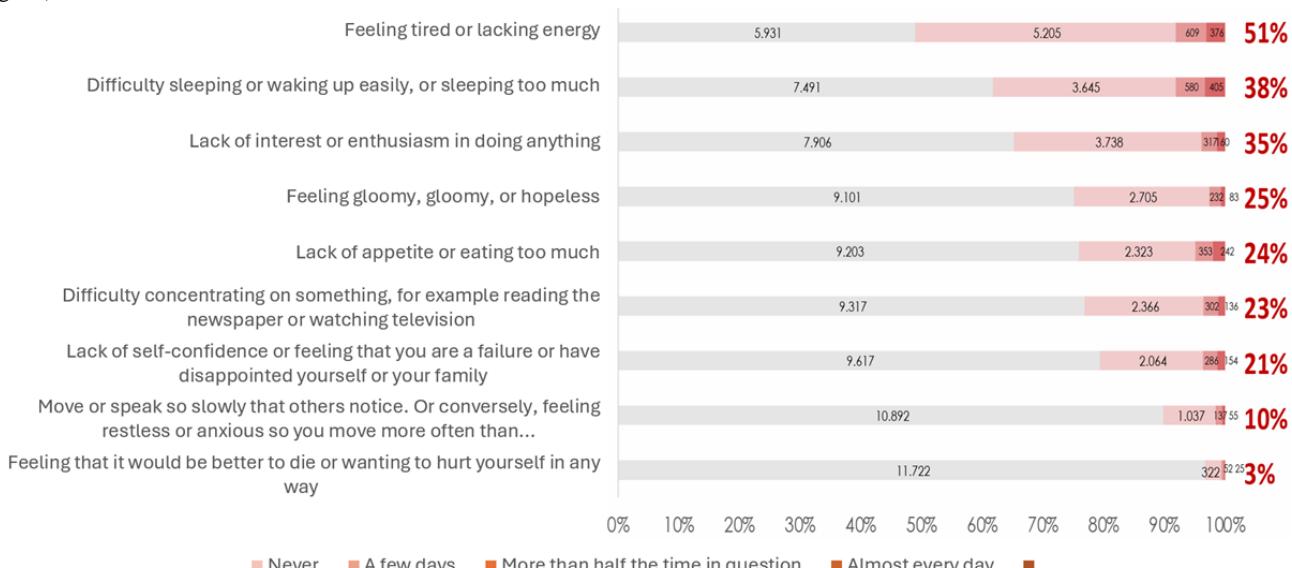
Paradox theory underpins this study, emphasizing the simultaneous coexistence of mutually contradictory yet interdependent elements that persist over time [7,8]. A paradox is defined as “contradictory yet interrelated elements that exist simultaneously and persist over time” [9]. In the context of specialist medical education, these tensions emerge when

academic rigor, resilience, and institutional pride—typically considered strengths—lead to burnout, intimidation, or compromised well-being.

Medical students and residents frequently face extreme academic pressure that may lead to burnout or depression [10]. Burnout is characterized by emotional exhaustion, depersonalization, and reduced personal accomplishment [11], while depression comprises persistent sadness, loss of interest, or suicidal ideation. Although overlapping, the two conditions are conceptually distinct. Resident doctors are particularly vulnerable to both, and their compromised well-being negatively affects patient care quality [11]. Media coverage frequently portrays emergency conditions in specialist medical education, illustrating the severity of burnout and distress among trainees [12-14]. Policymakers have expressed growing concern over declining patient care quality linked to these systemic issues [15].

A national survey conducted by the Ministry of Health of Indonesia in 2024 reported that among 12,121 participants in the *Program Pendidikan Dokter Spesialis* (PPDS; Specialist Doctor Education Program) at 28 vertical education hospitals, 2716 (22.4%) showed depressive symptoms [16]. Additionally, 6182 (51%) reported fatigue or low energy, 4606 (38%) experienced sleep disturbances, 4242 (35%) reported reduced interest in daily activities, and 400 (3.3%) indicated suicidal ideation or self-harm thoughts (Figure 1) [16]. These results underscore the significant mental health challenges facing trainees.

Figure 1. PPDS mental health screening at vertical education hospitals. PPDS: *Program Pendidikan Dokter Spesialis* (Specialist Doctor Education Program).



The same survey revealed that the specialties with the highest depressive symptoms were pediatrics, with 1672 (13.8%) trainees; orthopedics and traumatology, with 1491 (12.3%) trainees; and internal medicine, with 1091 (9.0%) trainees [16]. This situation raises concerns regarding both trainee well-being and patient care quality. Media reports frequently highlight

psychological crises among specialist medical trainees, echoing public worries about inadequate institutional support. High academic expectations combined with insufficient psychosocial resources leave many trainees struggling with anxiety, exhaustion, and a profound sense of entrapment.

These challenges reflect worldwide patterns. In the United States, a prospective study found that depressive symptoms increase progressively across medical training, with fourth-year students having more than eight times the odds of emotional exhaustion compared to first-year students [17]. More than one-third of US family medicine residents report burnout, particularly women and those with high educational debt [18]. In Canada, over half of resident doctors experience burnout, especially those working over 80 hours per week [19]. A scoping review reported a wide variation in the prevalence of burnout and depression among physicians and residents, with the evidence consistently indicating greater vulnerability among women [20].

Regional data reveal similar patterns: Iranian medical students experience significant burnout associated with academic pressures [21]; Hong Kong students commonly report burnout linked to poor sleep and limited exercise [22]; and Thai clinical students report high emotional exhaustion and depersonalization, particularly among males and those with lower academic performance [23]. The British Medical Association highlights excessive demands and poor working conditions as major contributors to declining physician well-being worldwide [24].

Although these issues occur worldwide, Indonesia's case is particularly severe due to rigid systemic design, hierarchical cultures, and limited psychosocial support. These characteristics intensify paradoxical tensions consistent with broader paradox scholarship, which argues that unresolved tensions often generate both innovation and dysfunction [7,8].

Aims of the Study

This study had two primary aims. Theoretically, it used the Icarus Paradox to explain how exclusivity and academic excellence within Indonesia's specialist medical education system can inadvertently produce negative outcomes for trainee well-being and service quality. Practically, it examined public perspectives captured through online media to identify institutional gaps in responsiveness to societal expectations, health care market needs, regulatory dynamics [25-28], and medical technology advancement [29]. In doing so, the study asked how the Icarus Paradox manifests within Indonesia's specialist medical education system from the viewpoint of the public, and how these insights can guide policymakers and training institutions toward reforms that balance quality, responsiveness, and trainee well-being [29].

Theoretical Background

Paradoxical dynamics have become increasingly prominent in contemporary organizations as scholars examine how contradictory yet interrelated elements coexist and shape behavior over time. Such tensions surface when organizations must respond to opposing demands that are both legitimate, generating cycles of alignment, resistance, and adaptation. These dynamics appear across institutional logics, identity structures, and organizational processes, with senior leaders experiencing them as strategic contradictions and employees encountering them in daily routines and socioemotional relationships [30]. Although paradox scholarship has generated numerous influential studies [31-33], its rapid adoption also raises concerns

about conceptual stretching, overconfidence in prevailing interpretations, and labeling practices that may reinforce dominant theoretical positions [34].

This study adopted the Icarus Paradox as its core theoretical lens. Rooted in Greek mythology, the metaphor describes how Icarus's success in flying ultimately led to failure when he ignored warnings and flew too close to the sun [35-37]. In strategic management, the Icarus Paradox explains how organizations' strengths can become vulnerabilities when past success fosters overconfidence, rigidity, or reluctance to adapt [35,38]. Institutions that excel may inadvertently undermine their own survival by persisting with established routines despite environmental shifts [39]. This phenomenon, often referred to as "vicious inertia," reflects a broader tension between stability and adaptation—an enduring hallmark of paradox theory. Comparable dynamics emerge in education, where aspirations for excellence collide with resource scarcity, rigid structures, and uneven institutional preparedness [40].

Paradoxical tensions are also evident in medical education. High academic standards and professional expectations are essential for producing competent physicians, yet they simultaneously generate psychological burden among trainees. Medical students frequently report stress, burnout, and depression [41,42], and Indonesia's specialist medical trainees experience particularly elevated levels of distress, with consequences for well-being and professional performance [10]. These tensions represent *strategic paradoxes*—institutions prioritizing excellence and productivity at the expense of trainee mental health. Identity paradoxes also emerge as students' aspirations to become empathetic, capable physicians conflict with the exhaustion, self-doubt, and emotional strain encountered during training. Institutional logic paradoxes further arise when cultures valorize endurance and sacrifice, while health systems increasingly demand empathy, patient-centeredness, and clinician well-being.

These psychological strains carry serious implications. Burnout and depressive symptoms reduce clinical performance, increase the likelihood of errors, and diminish empathy toward patients [43]. If unaddressed, these conditions often persist into residency and professional practice, reinforcing negative cycles within the health care workforce [11]. Medical education paradoxes therefore represent tangible, lived realities that directly influence patient safety and service quality.

Addressing these tensions requires intentional institutional support. Access to mental health services, resilience education, and curricula integrating well-being are foundational [42,43]. Another paradox emerges from the gap between expected graduate competencies and actual system-level outcomes: training aspires to produce highly capable specialists, but health care performance does not always match societal expectations [44]. Continuous reform is therefore essential to align educational structures with evolving health care needs, while recognizing human limitations.

Social media has recently played a dual role in illuminating these paradoxes. On the one hand, it facilitates collaboration [27], resource sharing [45], knowledge exchange [46], and peer support that strengthens students' sense of belonging [26,47,48]. On the other hand, it can propagate misinformation, amplify

stress, and distract from academic focus [49]. Online engagement also provides a space for public commentary [50,51], revealing lived experiences of medical trainees and societal expectations of the health care system. Thus, social media functions both as a catalyst and a mirror of the paradoxes embedded within specialist medical education.

Methods

Study Design

This study was registered at ClinicalTrials.gov (NCT123456). The research used a convergent mixed methods design [52], in which quantitative and qualitative approaches were conducted simultaneously and coalesced to provide a rich understanding of the Icarus Paradox within the specialist medical trainee education system in Indonesia. In this design, quantitative sentiment analysis and qualitative thematic analysis were conducted independently but merged at the interpretation stage to facilitate triangulation of findings.

Data Collection

Data collection was carried out over 30 days (March 24-April 23, 2024) using Brand24 [50,53,54], a social listening and monitoring tool that enables focused and comprehensive analysis of conversations on platforms such as Twitter [55], online news portals, blogs, videos, discussion forums, and websites [56,57]. Brand24 was selected over other monitoring tools because it provides reliable access to Indonesian language data, allows automated tracking of multiple keywords, and has been previously validated in academic studies [50].

The search was conducted using a set of keywords relevant to Indonesia's specialist medical education system, including "PPDS," "resident doctor," "specialist candidate," "collegium," "medical student," and "medical education." The data collected were stored in Microsoft Excel format, containing metadata, such as the posting time, ID or username, source, review content, sentiment category, number of replies, number of likes, number of reposts, number of followers, and influence score.

To ensure the relevance of the dataset, irrelevant or duplicate content was systematically removed. Posts were excluded if they were (1) purely promotional or advertisements, (2) unrelated to medical education despite containing a keyword, (3) spam or bot-generated content, or (4) duplicates of

previously captured entries. This filtering step ensured that only meaningful and contextually relevant discussions were retained for analysis.

Data Analysis

The analysis process involved both quantitative sentiment analysis and qualitative thematic analysis.

For quantitative sentiment analysis, Python was used with text mining and natural language processing (NLP) libraries, such as *NLTK*, *spaCy*, *Pandas*, and *Scikit-learn* [58,59]. Sentiment classification grouped posts into three categories: positive, neutral, and hostile. Additional preprocessing steps included removing hashtags, emojis, hyperlinks, and user mentions, as these elements did not add meaningful semantic value for sentiment classification. The output of this analysis included statistics, graphs, proportions, and trends.

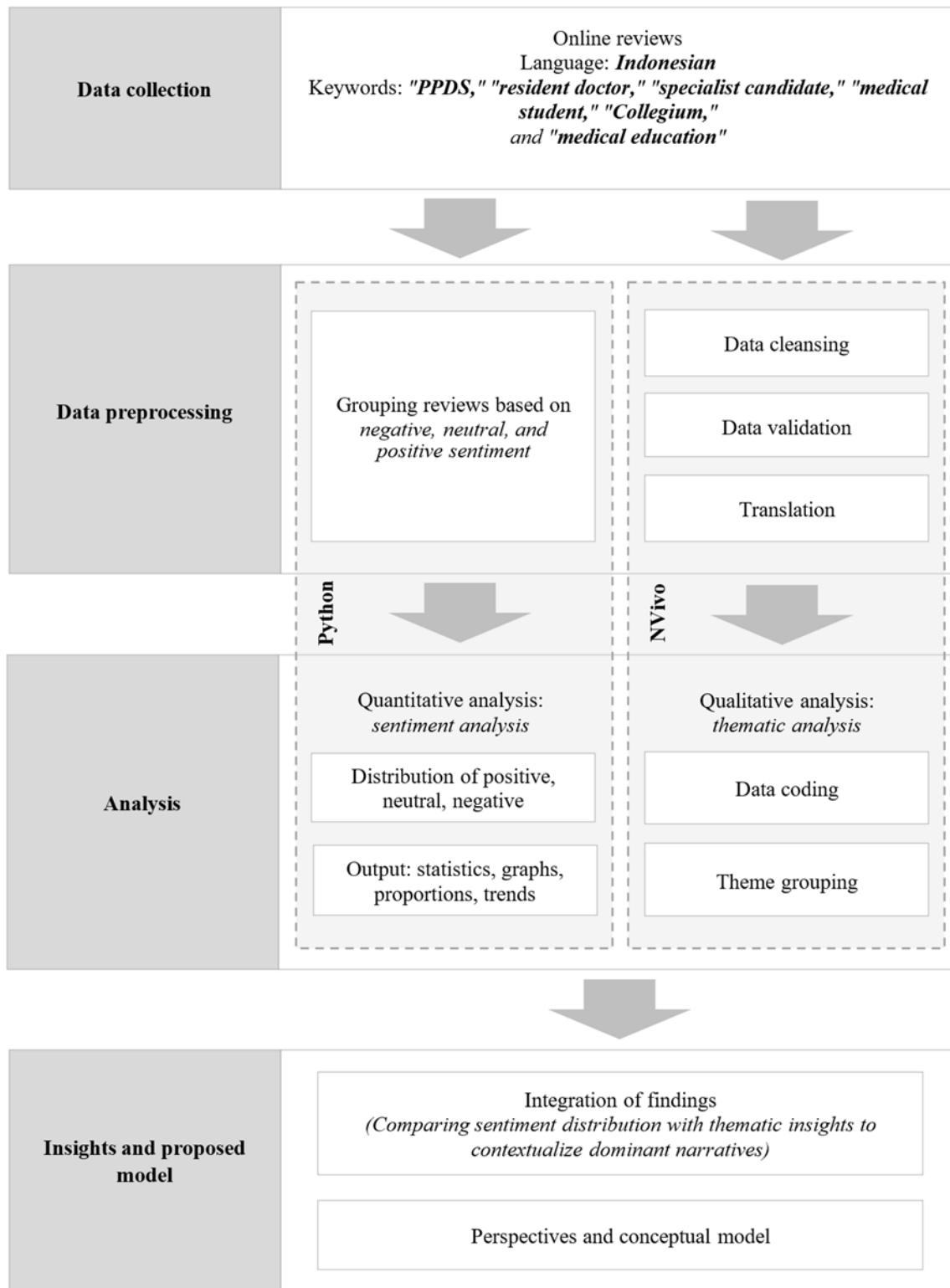
Since the dataset was originally in Indonesian, translation into English was necessary for subsequent qualitative coding. Translation was performed using Google Translate, combined with manual verification by bilingual members of the research team to ensure contextual and semantic accuracy. This step reduced the risk of misinterpretation and enhanced the reliability of findings.

For qualitative thematic analysis, data were imported into NVivo 14 [60]. Coding was performed using a hybrid approach: inductive coding allowed themes to emerge directly from the data, while deductive coding was informed by paradox theory and the Icarus Paradox framework. To enhance intercoder reliability, multiple researchers were involved in the coding process, and discrepancies were resolved through discussion until a consensus was reached. Codes with similar content were grouped into broader themes to identify dominant narratives.

Integration of Findings

The final stage involved integrating the results from both strands of analysis. Quantitative sentiment trends were compared with qualitative thematic insights to contextualize dominant narratives about specialist medical education. This integration enabled the development of a perspectives-based conceptual model, summarizing how the Icarus Paradox manifests in Indonesia's specialist medical education system and how public perceptions highlight areas requiring reform. A summary of all stages of the method is shown in Figure 2.

Figure 2. Convergent mixed methods study design. PPDS: *Program Pendidikan Dokter Spesialis* (Specialist Doctor Education Program).



Ethical Considerations

This study used publicly available, nonidentifiable online data and did not involve interaction with human participants; therefore, Institutional Review Board (IRB) approval was not required. The authors adhered to applicable ethical standards for research using public online data and removed any potentially identifying information during analysis.

Results

Online Review Data

Data were collected for 30 days (March 24-April 23, 2024) using Brand24 as the primary social listening and analytics platform [50,53,54]. This tool was selected because of its ability to capture Indonesian language data across multiple platforms, provide sentiment classification, and export structured datasets.

for further quantitative and qualitative analyses. In total, 5047 reviews were retrieved based on the specified keywords, distributed across digital platforms such as Twitter (X), TikTok, video-sharing sites, online news portals, podcasts, forums, blogs, and websites.

The platform with the highest number of reviews was news portals, with 3040 (60.2%) reviews; followed by websites, with 692 (13.7%), Twitter with 573 (11.3%), and blogs with 353 (7%) reviews. TikTok and video systems additionally contributed significantly, with 282 (5.6%) reviews and 96 (1.9%) opinions, respectively. Podcasts accounted for the best 6 (0.1%) evaluations, and forums contributed 5 (0.1%) reviews, making them the least represented platforms. These percentages provide a clearer picture of the dominance of such structures, with news portals rising to become the number one space for public discourse on professional clinical education in Indonesia. The variation highlights how exclusive audiences use more than one online channel to share their perspectives and worries, underscoring the subject's relevance in numerous segments of digital media.

Table 1. Distribution of review data by platform.

Platform	Negative reviews (n=312), n (%)	Neutral reviews (n=4389), n (%)	Positive reviews (n=346), n (%)	Total (N=5047), n (%)
Twitter (X)	218 (69.9)	251 (5.7)	104 (30.1)	573 (11.3)
TikTok	19 (6.1)	225 (5.1)	38 (11.0)	282 (5.6)
Videos	1 (0.3)	89 (2.0)	6 (1.7)	96 (1.9)
News portals	50 (16.0)	2885 (65.7)	105 (30.3)	3040 (60.2)
Podcasts	0	6 (0.1)	0	6 (0.1)
Forums	0	5 (0.1)	0	5 (0.1)
Blogs	11 (3.5)	329 (7.5)	13 (3.8)	353 (7.0)
Websites	13 (4.2)	599 (13.6)	80 (23.1)	692 (13.7)

News platforms exhibited the largest volume of data, amounting to 3040 (60.2%) reviews; however, most were neutral (n=2885, 94.9%), accompanied by 105 (3.5%) positive and 50 (1.6%) negative sentiments. Blogs and websites contributed 353 (7%) and 692 (11.3%) reviews, respectively, with neutral sentiment dominating both categories (n=329, 93.2%, for blogs and n=599, 86.6%, for websites).

Overall, these findings indicated substantial cross-platform variation in sentiment distribution. Despite differences in volume, most platforms show a predominance of neutral sentiment, suggesting that public discourse on specialist medical education tends to be descriptive or informational rather than overtly evaluative.

These data show that Twitter has the highest proportion of negative sentiment, whereas TikTok has the highest proportion of positive sentiment relative to its volume. Meanwhile, news portals are overwhelmingly neutral, reflecting the reporting

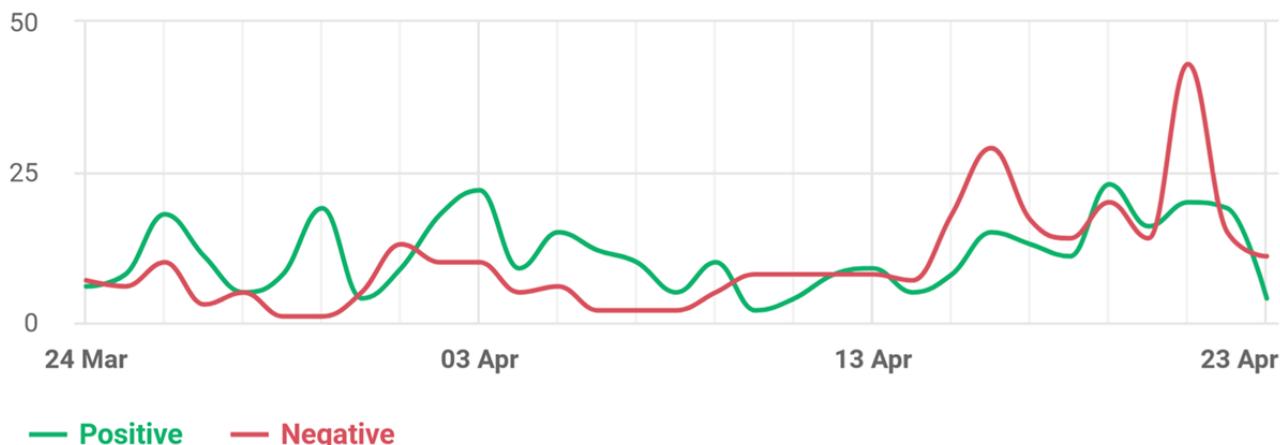
Sentiment Analysis

Sentiment classification of the collected data was conducted using the Brand24 platform [61]. Brand24 integrates artificial intelligence (AI)-driven deep learning with rule-based lexicon techniques, enabling both high accuracy and methodological transparency. Its multilingual sentiment model supports analysis in more than 100 languages, including Indonesian, and its performance has been validated using the F_1 -score metric. Based on evaluations of 50,000 mentions, overall accuracy increased [61].

The processed data reflected a mixture of negative, neutral, and positive sentiments across platforms, as summarized in **Table 1**. The online review dataset showed substantial variation in the volume and distribution of sentiments between media sources. Of the 573 (11.3%) reviews contributed by Twitter, 218 (38%) were categorized as negative, 251 (43.8%) as neutral, and 104 (18.2%) as positive entries. Of the 282 (5.6%) reviews generated by TikTok, the majority (n=225, 79.5%) were categorized as neutral. For video-based platforms, such as YouTube, which produced 96 (1.9%) reviews, 89 (92.7%) reviews were categorized as neutral.

style of professional media outlets. Podcasts and forum platforms do not show any negative or positive sentiment, while video and web-based platforms have a low number of negative- and positive-sentiment reviews. Furthermore, news platforms show a fairly low number of negative- and positive-sentiment reviews, with neutral sentiment dominating. Overall, the percentage of negative sentiment of around 6.2% (n=312 reviews) is smaller than the percentage of positive sentiment of 6.9% (n=346 reviews).

Figure 3 shows the trend in daily positive and negative reviews for 1 month. At the beginning of the month, the number of positive and negative reviews was relatively low and stable. However, starting April 16, there was a significant increase in the number of positive and negative reviews. This is in line with reports in the mass media regarding the results of PPDS mental health screening [16] on that date [12], which may have triggered an increase in online activity related to the topics discussed in the research.

Figure 3. Sentiment-based review history for 1 month.

Coding Review Topics

After carrying out data cleaning, 292 (5.8%) reviews with negative sentiment and 317 (6.3%) reviews with positive sentiment were obtained. These data were then analyzed using NVivo 14.

Negative-Review Topic Coding

As the initial aim of this research, each sentiment in the review was coded to find review themes. We used the *autocode* feature provided by NVivo 14 to automatically recognize and mark text or other qualitative data based on predetermined rules. NVivo applies techniques such as text analysis and content analysis to identify patterns in text and tag or code data accordingly [62].

To ensure the credibility of the coding process, the results of *autocode* generation were followed by manual validation carried out by two independent coders. An intercoder agreement check was conducted, and discrepancies were resolved through discussion to strengthen the reliability of the coding.

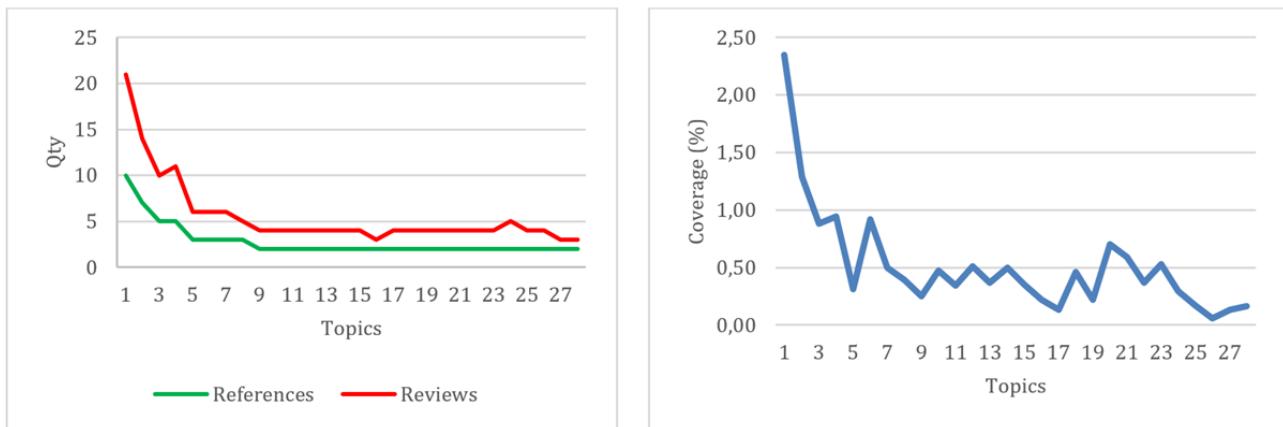
The *autocode* results produced 28 negative-review topic codes, as presented in [Table 2](#). To improve interpretability, these topics were subsequently grouped into six broader clusters: *medical workforce*, *educational system and training*, *health care institutions and patients*, *financial and resource constraints*, *media and public discourse*, and *other student life factors*.

The largest cluster was “*educational system and training*” (n=8, 28.6%, topics), followed by “*medical workforce*” (n=7, 25%, topics). This highlights that criticisms are not only about individuals (doctors, students, residents) but also about structural issues in programs and curricula. Smaller clusters, such as “*media and public discourse*” and “*student life factors*,” capture contextual influences that extend beyond institutional boundaries, underscoring the multifaceted nature of the challenges.

Discussion topic data generated by NVivo 14 related to negative community reviews revealed a variety of topics highlighted in these reviews. Of the 28 topics identified, based on the number of references and reviews ([Figure 4a](#)), the data showed that the topics of discussion tended to be spread out, with significant variations in the number of references and reviews. Most topics had an equal number of references and reviews, with some topics, such as “*doctors*” and “*specialists*,” having a slight difference in the number of references and reviews. However, when looking from a coverage perspective ([Figure 4b](#)), most topics had low scores, indicating that these topics were not widely discussed in the data. Nevertheless, it was still important to pay attention to them because some topics could affect the validity and representativeness of analysis results.

Table 2. Topics in negative reviews.

Topic number	Topic	Topic feature words	Cluster
1	Doctors	intern doctors, resident doctors, alone specialist doctors, emergency room doctors, previous doctors, procuring specialist doctors, prospective doctors, teachers, specialist doctors	Medical workforce
2	Specialists	lung specialist, alone specialist doctors, dermatology specialists, otolaryngology specialists, procuring specialist doctors, teachers, specialist doctors	Medical workforce
3	Students	medical student status, boarding student, dental students, nursing students	Educational system and training
4	Hospitals	vertical hospitals, hospital environment, hospital management, large hospitals	Health care institutions and patients
5	Teaching	teaching staff, teachers, specialist doctors	Medical workforce
6	Medicine	internal medicine, traditional medicine	Educational system and training
7	Programs	8th pregnancy program, expert master education program, study program	Educational system and training
8	Lungs	lung specialist, lung disease	Medical workforce
9	Nursing	nursing student, senior nurse	Medical workforce
10	Patients	sleep patients, trigger patients	Health care institutions and patients
11	Sleep	patient sleep, sleeping pills	Student life factors
12	Media	media briefings, social media	Media and public discourse
13	Vertical hospitals	vertical hospitals	Health care institutions and patients
14	Health	community health center, first-class health clinic	Health care institutions and patients
15	Education	expert master education program, medical education providers	Educational system and training
16	Expert master	expert master candidates, expert master education program	Educational system and training
17	Suitable	following suit, really suitable	Student life factors
18	Money	received money, scientific money	Financial and resource constraints
19	Medical student status	medical student status	Educational system and training
20	Internal doctors	internal doctors	Medical workforce
21	Results	good results, solid results	Educational system and training
22	Community	community health center, local community	Health care institutions and patients
23	Working	work strikes, working hours	Financial and resource constraints
24	Resident doctors	resident doctor	Medical workforce
25	Urine	urine bag, urine hunter	Financial and resource constraints
26	Smells	formalin smell, n't smell	Financial and resource constraints
27	Stages	stage compositor, stage rotation	Educational system and training
28	House	boarding house, halfway house	Financial and resource constraints

Figure 4. References and reviews (a) and coverage (b) for each negative-review topic.

Positive-Review Topic Coding

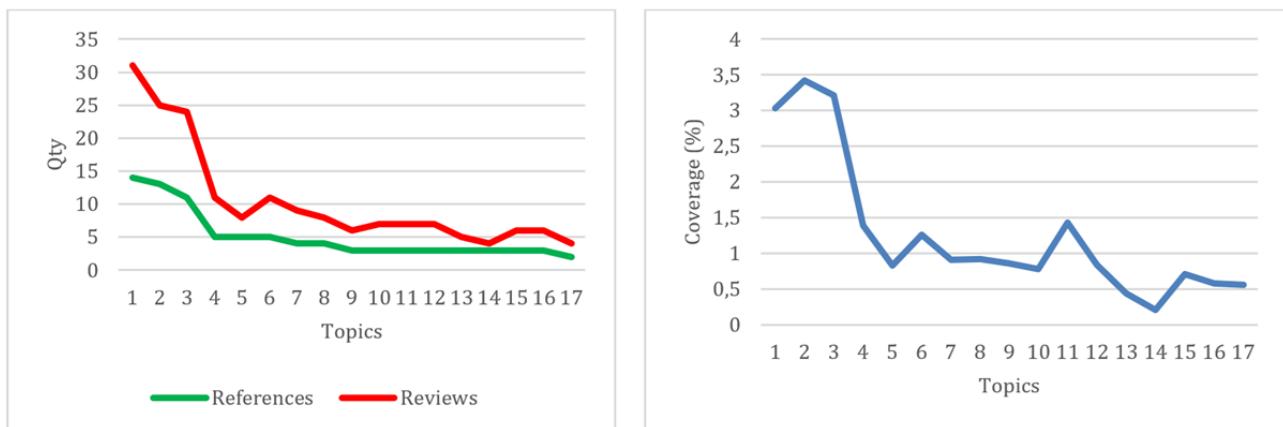
Autocoding carried out by NVivo 14 on positive reviews produced 17 topic codes, as summarized in [Table 3](#).

Based on the number of references and reviews ([Figure 5a](#)), the data showed significant variations in the number of references and reviews for each topic. Most topics had an equal number of references and reviews, but some had quite striking differences, such as the topics “doctors” and “services,” which

had a higher number of reviews than the number of references. However, from a coverage perspective ([Figure 5b](#)), the majority of topics had quite high scores, indicating that these topics were widely discussed in the data. Topics such as “doctors” and “services” even had coverage values above 3% (17 reviews), indicating that they are the main focus of positive community reviews. Even though some topics had lower coverage values, the majority of them were still quite high, showing the variety of positive topics discussed in public conversations.

Table 3. Topics in positive reviews.

Topic number	Topic	Topic feature words	Operational definition
1	Doctors	professional doctor candidates, resident doctors, aspiring doctors, general doctors, eye specialists, online doctor apps	Mentions of doctors in a positive sense, reflecting professionalism, expertise, or accessibility
2	Services	superior service, pediatrician services, integrated services, rehab services	References to the quality, diversity, or efficiency of health care services
3	Students	medical students, students	Positive perception of medical students' role, contribution, or progress
4	Health	various health problems, health centers, health fields, health practitioners	General statements about health promotion, prevention, or community health benefits
5	Surgery	cardiac, cataract, orthopedic, preventive surgery	Praise for successful surgical procedures or advanced surgical options
6	Specialists	pediatric specialists, beauty specialists, eye specialists	Recognition of specialist expertise or availability in various fields
7	Treatment	beauty treatment, canal treatment, treating skin, treatment techniques	Appreciation of effective or innovative treatment modalities
8	Practitioners	general practitioner, health practitioner, health care practitioner	Positive portrayal of practitioners' skills and dedication
9	Facilities	equivalent facilities, sophisticated facilities, surgical facilities	Highlighting the adequacy, modernity, or comfort of health care facilities
10	Consultation	direct consultation, online doctor consultation application, psychological consultation services	Positive feedback on ease of access and quality of medical consultations
11	Superior service	superior service	Reinforcement of outstanding service quality in medical practice
12	Fields	health fields, medical fields, various fields	References to the diversity of medical and health disciplines
13	Beauty	beauty specialist, beauty treatment, well-maintained beauty	Positive associations with medical aesthetics and wellness
14	Rooms	physiotherapy rooms, reflection rooms, speech therapy rooms	Praise for the availability and functionality of medical or therapeutic rooms
15	Practice	specialist pediatrician practice services, practice schedule	Mentions of structured and accessible medical practices
16	Programs	23 study programs, male pregnancy program, professional program	Positive recognition of medical or educational programs
17	Choices	menu choices, the right choice	Appreciation of having diverse and appropriate health care or program options

Figure 5. References and reviews (a) and coverage (b) for each positive-review topic.

Cognitive Perspectives and Themes

Cognitive perspectives were established to help explain how society responds to the Icarus Paradox and how these perceptions influence their behavior regarding medical education. The cognitive perspective in the research context refers to the way individuals understand, interpret, and give meaning to information obtained from their environment. This includes how people process the information they obtain from

online media, such as news articles, comments on social media, blogs, and discussion forums, as well as how their views are formed, influenced, and expressed through these platforms [27]. The combination of all topics in positive and negative reviews is then sorted to develop appropriate themes. Each corresponding theme is then grouped into a cognitive perspective. Based on all the reviews collected in this research, we defined four cognitive perspectives: *education system*, *policy*, *society's views of students*, and *health care services* (Table 4).

Table 4. Cognitive perspectives and themes by review topic.

Cognitive perspective and themes	Negative topics	Positive topics
Education system		
Professional challenges	<ul style="list-style-type: none"> Topic 1: doctors Topic 24: resident doctors Topic 25: urine 	<ul style="list-style-type: none"> Topic 1: doctors Topic 12: fields
Economic challenges	<ul style="list-style-type: none"> Topic 18: money 	— ^a
Workload	<ul style="list-style-type: none"> Topic 19: medical student status Topic 20: internal doctors 	—
Medical understanding	<ul style="list-style-type: none"> Topic 8: lungs 	—
Teacher qualifications	<ul style="list-style-type: none"> Topic 5: teaching Topic 17: suitable 	—
Medical education	<ul style="list-style-type: none"> Topic 7: programs Topic 13: vertical hospitals Topic 15: education Topic 16: expert master 	<ul style="list-style-type: none"> Topic 4: health Topic 16: programs
Policy		
Government policy	<ul style="list-style-type: none"> Topic 23: working Topic 27: stages 	—
Society's views of students		
About students	<ul style="list-style-type: none"> Topic 3: students Topic 28: houses 	<ul style="list-style-type: none"> Topic 3: students
Community involvement	<ul style="list-style-type: none"> Topic 12: media 	<ul style="list-style-type: none"> Topic 6: specialists
Health care services		
Quality of health services	<ul style="list-style-type: none"> Topic 2: specialists Topic 4: hospitals Topic 6: medicine Topic 9: nursing Topic 14: health Topic 21: results Topic 26: smells 	<ul style="list-style-type: none"> Topic 2: services Topic 5: surgery Topic 8: practitioners Topic 9: facilities Topic 11: superior service Topic 13: beauty Topic 15: practice
Patient experience	<ul style="list-style-type: none"> Topic 10: patients Topic 11: sleep Topic 22: community 	<ul style="list-style-type: none"> Topic 7: treatment Topic 10: consultation Topic 14: rooms Topic 17: choices

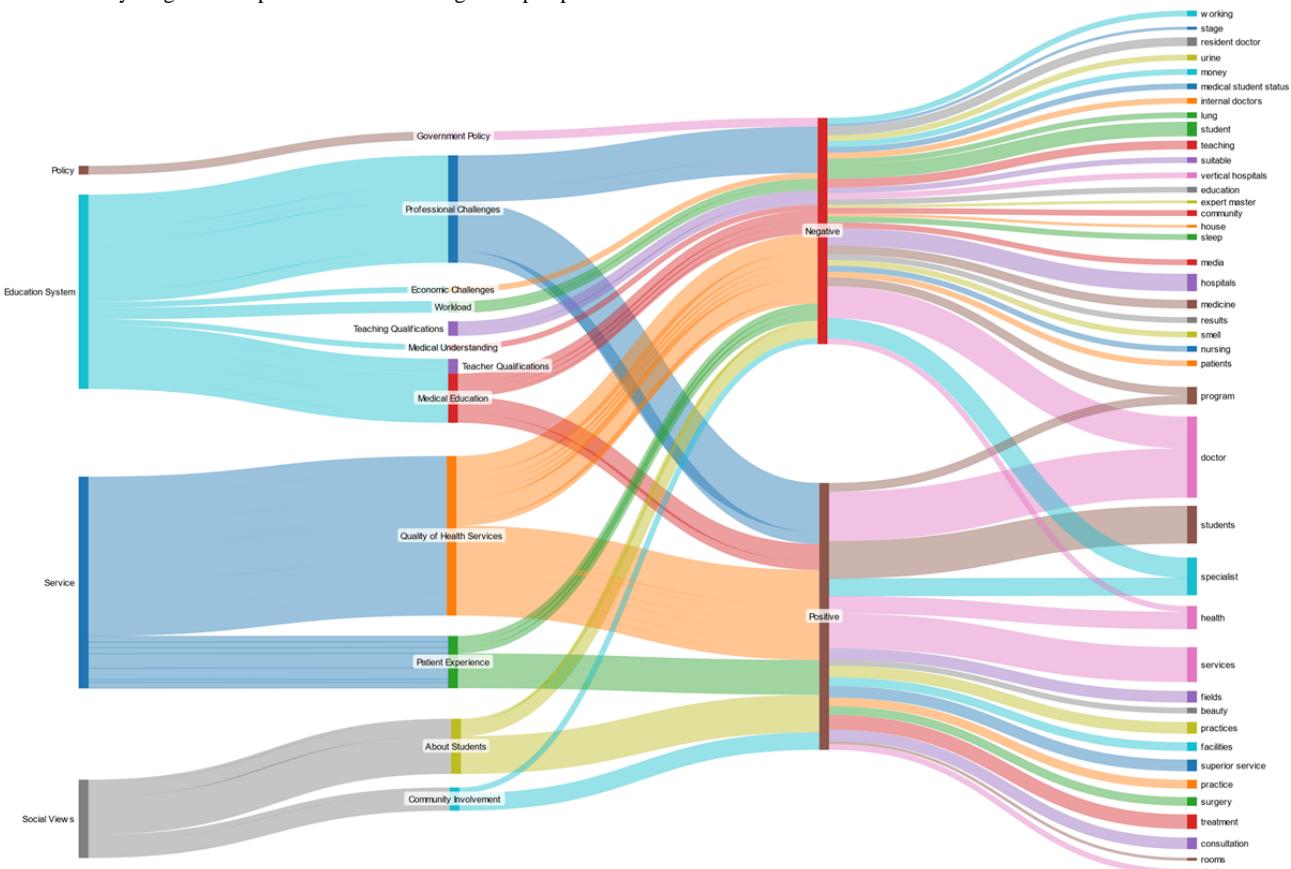
^aNot applicable.

The “education system” perspective explains prospective doctors’ direct experience with the medical education system and health facilities, which will shape their perceptions and assessments of the quality and effectiveness of the system. The

information they obtain from online media can influence how they interpret and remember their own experiences. This perspective explains several themes, such as professional challenges, economic challenges, workload, medical

understanding, teaching qualifications, and medical education. The “policy” perspective explains the public’s response to assessing the quality of medical education and health services, which is linked to the impact of government policy. Information about government policies, societal responses, and changes in the health system may influence the public’s perceptions of the medical education system. This perspective focuses solely on the theme of government policy. Furthermore, the “society’s views of students” perspective influences perceptions of medical education and the medical profession, shaped by the social support individuals receive from their environment. The perspectives of friends, family, and the medical community can influence how individuals understand and respond to the challenges they face in their medical education and careers. This perspective explains themes such as student views, as well as the themes of community engagement and public health. The “health care services” perspective explains public reviews

Figure 6. Sankey diagram of topic distribution for cognitive perspectives and themes.



The topic *program* centered on the theme of “medical education” within the “education system” perspective. Positive reviews emphasized the value of educational opportunities, such as clinical clerkships and diverse study programs offered by Indonesia’s medical schools. These were seen as mechanisms to broaden competence and strengthen professional readiness. At the same time, negative sentiment highlighted recurring concerns, including bullying cases in the PPDS and the closure of certain specialties, such as radiology. These contradictory views—progressive ambition versus internal dysfunction—reflect the Icarus Paradox, where efforts to elevate educational excellence may simultaneously generate harmful systemic pressures [35,36].

<https://mededu.jmir.org/2026/1/e60452>

regarding the experience of receiving services and the results obtained when visiting a doctor. This perspective explains themes such as the quality of health services and the patient experience. The entire cognitive perspective, along with its respective themes, is further explained in the coding process of the negative and positive reviews next.

Visualization of topic flow across cognitive perspectives, themes, sentiments, and issues was produced using a Sankey diagram (Figure 6). The diagram shows that five topics (*program*, *doctor*, *student*, *specialist*, and *health*) contained both positive and negative sentiments, indicating contradictory reviews and the presence of paradoxical dynamics. Three topics (*program*, *doctor*, *student*) displayed opposing sentiments within the same theme and perspective, whereas two topics (*specialist* and *health*) exhibited sentiment divergence across different themes or perspectives.

The topic *doctors* related to “professional challenges” within the “education system” perspective. Positive sentiment acknowledged the achievements of doctors and medical students, recognition of expertise, and optimism surrounding innovations, such as online consultation platforms. However, negative sentiment pointed to structural barriers, such as shortages of specialist lecturers, harsh working conditions, low compensation for interns, and ethical concerns related to patient rights. These tensions illustrate the conflict between professional ideals and the realities imposed by institutional limitations.

The topic *students* was discussed under the “society’s views of students” perspective. Positive sentiment highlighted recognition

of student accomplishments, the use of simulators and models in training, and hopes for producing compassionate and resilient physicians. Conversely, negative reviews drew attention to workload imbalance, ethical dilemmas when shifting roles, and perceptions of injustice in training environments. These competing narratives reflect the paradox of aspiring to excellence, while navigating a system that can undermine fairness and well-being.

Sentiment patterns for the topic *specialists* also revealed contrasting perspectives. Positive sentiment, categorized under “community involvement,” underscored the indispensable role of specialists in delivering quality care, especially in underserved areas. Public comments highlighted trust in specialists’ ability to accelerate recovery and address complex conditions. In contrast, negative sentiment, linked to the “health care services” perspective, emphasized limited specialist availability, particularly in remote regions, as well as funding and resource constraints. These shortages translate directly into service gaps, as illustrated by cases in which patients cannot obtain urgently needed specialist care.

Finally, the topic *health* contained both positive and negative sentiment across differing themes. Positive reviews, aligned with the “medical education” theme under the “education system” perspective, highlighted improvements in training quality, innovations in medical research, and the contribution of diverse specialists to system enhancement [63]. Negative sentiment, viewed from the “health care services” perspective,

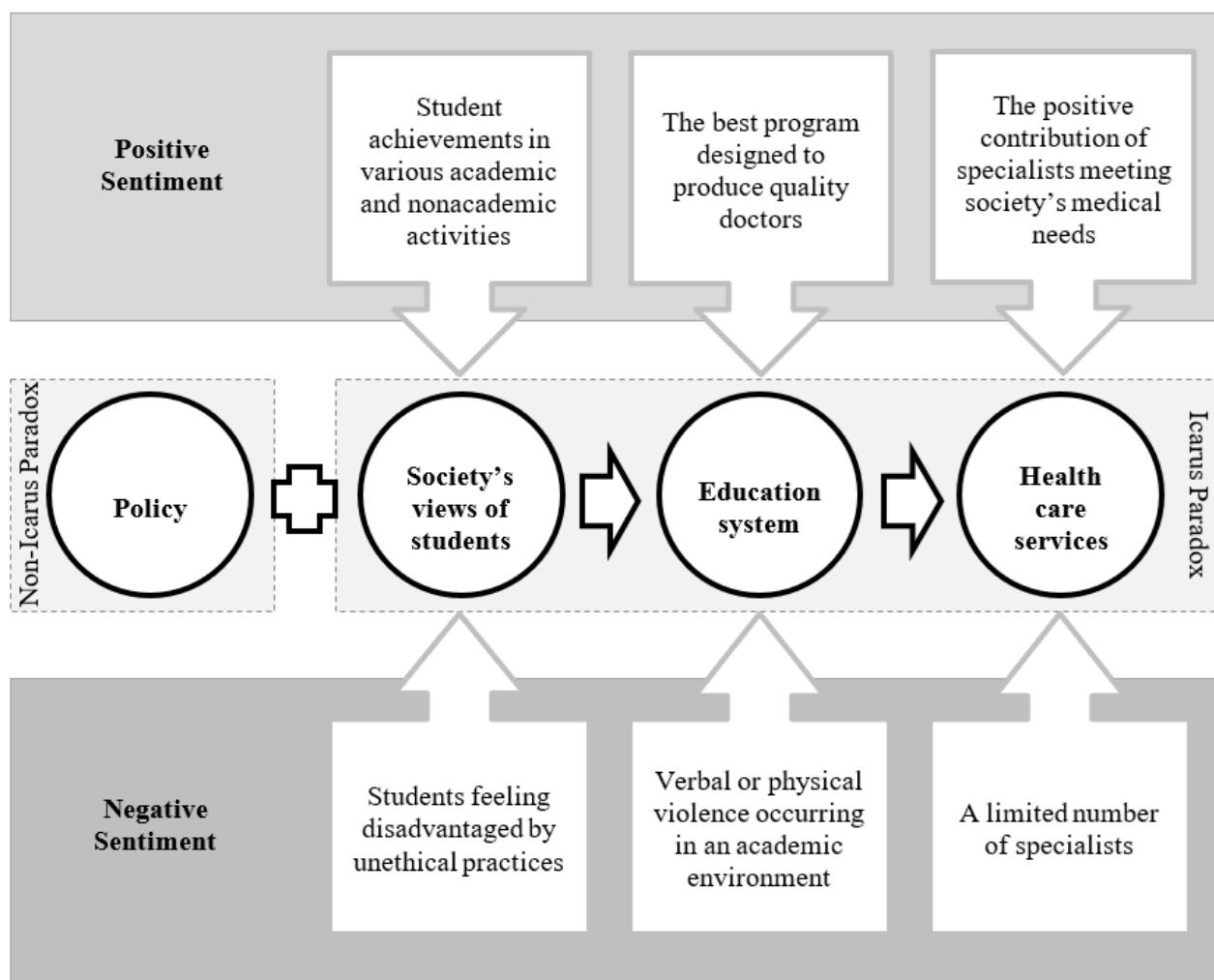
reflected dissatisfaction with access barriers, cost concerns, and perceived shortcomings in primary-level care. Public comments described challenges in securing treatment for family members and skepticism toward provider competence in community clinics.

Collectively, these topic-based findings revealed how paradoxical tensions manifest simultaneously across multiple perspectives in Indonesia’s specialist medical education ecosystem, providing critical insight into structural, educational, and service-related challenges.

Conceptual Model

Based on the results of our analysis, we proposed a conceptual model that links these four cognitive perspectives. This model can serve as a basis for formulating comprehensive policies and strategies to enhance medical education, improve public perception of the medical profession, and overall health services. With an integrated and holistic approach, it is hoped that the Icarus Paradox identified in this study can be overcome, thereby improving the overall quality of the health system. The proposed conceptual model is shown in Figure 7. Perspectives with white circles have the Icarus Paradox effect, while perspectives marked with gray circles do not have the Icarus Paradox effect. The “education system” perspective in this conceptual model is the primary component in producing quality health services, which is largely determined by the input and role of regulations, as well as “society’s views of students” regarding prospective medical students.

Figure 7. Conceptual model linking the four cognitive perspectives in this study: (1) education system, (2) policy, (3) society's views of students, and (4) health care services.



Discussion

Principal Findings

This study identified three major cognitive perspectives—education system, health care services, and society's views of students—frequently reflected in public reviews, while policy issues appeared less prominently. Findings illustrate the Icarus Paradox as it emerges across these perspectives, where ambition and the pursuit of excellence coexist with constraints and systemic limitations. Rooted in classical interpretations of overreach and human limitation [37,38], the paradox becomes visible in how communities describe tensions within Indonesia's specialist medical education system. In contrast, the policy perspective generated comparatively uniform sentiment, suggesting limited public engagement with regulatory matters.

The Icarus Paradox in Specialist Medical Education

The Icarus Paradox describes how attributes that drive success can, when overstretched, evolve into vulnerabilities [7,37,38]. Indonesia's specialist medical education system reflects this dynamic. Programs such as clinical clerkships and PPDS training offer essential experiential learning and uphold high academic

and professional standards [11,44,64]. These strengths, however, may escalate into risks when demands become excessive. Heavy workloads, hierarchical structures, and competitive environments contribute to burnout, intimidation, and uncertainty, exemplified by bullying cases and program closures, such as in radiology [10,11]. Public discourse captured in this study also shows admiration for educational excellence, alongside accounts of program instability and insufficient institutional support, illustrating how the system's strongest features can simultaneously generate tension.

Addressing these paradoxes requires structural and cultural reforms, including more inclusive learning environments, transparent governance, and stronger mentoring systems grounded in empathy and collaboration [65]. Updating curricula to align with evolving scientific and educational developments is equally vital. Such measures may help sustain academic quality, while minimizing the unintended consequences of overly rigid standards.

The Icarus Paradox in the Student Context

Among students, the paradox emerges when determination, resilience, and idealism—traits celebrated at the start of medical training—become sources of vulnerability under excessive pressure. Students often begin their education with strong

motivation and aspirations to contribute meaningfully to society [66]. Yet system demands, extended working hours, ethical dilemmas, and limited supervision due to dual lecturer-clinician roles can erode well-being and moral sensitivity [41]. Review data reveal disillusionment when expectations of professional fairness conflict with realities, such as inequitable pathways into practice or nonformal short-course alternatives.

Such tensions reflect how ambition without adequate support can lead to burnout and ethical distress. Strengthening mental health support, promoting fairness in training systems, and integrating ethical guidance are essential to prevent resilience from becoming a liability [65].

The Icarus Paradox in the Health Care Context

Paradoxes also materialize within health care delivery [67]. Advances in training, the expansion of simulations, and improvements in health facilities have contributed to better-prepared professionals and broader service coverage. Government initiatives, such as infrastructure development, *Jaminan Kesehatan Nasional* (JKN; National Health Insurance Program) implementation, and adoption of digital health and telemedicine, have improved access in many regions. However, unchecked growth and technological progress can introduce new weaknesses, including practitioner burnout, ethical compromises, or neglect of basic primary care, particularly in underserved communities [8,25].

Persistent disparities between urban and rural areas reflect this tension. Although capacity has increased nationally, specialist-to-population ratios and hospital bed availability remain considerably lower in remote regions [2,68-70]. These gaps demonstrate how reforms designed to elevate care quality can inadvertently reinforce inequality.

Addressing such paradoxes requires integrating community-based training, strengthening support for rural deployment, and ensuring equitable resource allocation [26]. Civil society involvement can help promote fairness and public accountability.

The “education system” perspective illustrates how efforts to sustain high academic standards and broaden learning opportunities in specialist training can inadvertently generate strain. Although diverse programs and research initiatives aim to enhance the quality and relevance of specialist medical education, these strengths may create vulnerabilities when demands exceed student capacity. Heavy academic loads, tight schedules, and competitive selection for placements or specialties can intensify stress and disrupt students’ balance between academic, social, and personal lives. Limited institutional attention to student well-being, insufficient support systems, and cultures that tolerate intimidation or academic mistreatment may further exacerbate these pressures, reinforcing the system’s paradoxical nature.

These dynamics reveal the tension between the ideal of academic excellence and the structural challenges embedded in training environments. Understanding this paradox is essential when evaluating how effectively the educational system prepares future doctors for real-world health care demands.

The “society’s views of students” perspective similarly reflects tensions between public expectations of medical professionalism and students’ lived realities. High academic pressure, limited access to qualified supervisors, financial burdens, and long working hours often conflict with students’ aspirations, contributing to fatigue and reduced well-being. These constraints show how ambition in medical training may collide with structural and ethical limitations, creating conditions consistent with the Icarus Paradox.

Although the “policy” perspective does not directly display paradoxical tension, it remains crucial in mitigating risks emerging from the other three perspectives. Well-designed regulations can support improvements in medical training quality, shape public perceptions of the profession, and promote equitable access to health services.

The “health care services” perspective represents the system’s final output and illustrates tensions between the need for high-quality health care and persistent resource constraints. Limited numbers and uneven distribution of specialists, especially in remote regions, hinder timely access to appropriate care. Funding shortages, inadequate human resources, and complex bureaucratic processes further complicate service delivery. These factors generate a paradox in which aspirations for excellent, equitable care conflict with systemic limitations [67].

Overall, the conceptual model highlights how progress within education, professional expectations, and health care delivery can unintentionally produce new vulnerabilities. Recognizing and managing these paradoxes is essential for creating a more resilient and equitable medical education and health care system.

Implications

The findings of this study carry significant implications for medical education, health service quality, and the role of specialist medical trainees [26]. First, within medical education, the Icarus Paradox, reflecting tensions between ambition and structural limitations, shapes student experiences and underscores the urgency of curriculum reform. Greater emphasis is needed on clinical skill development [71], communication, and leadership, which are essential for effective future practice. Educational institutions must also enhance academic and emotional support systems to help students manage stressors that arise during training.

Second, gaps in health service quality between urban and rural regions remain substantial. This highlights the need for increased investment in rural health infrastructure and the equitable placement of medical personnel. Elevating service standards nationwide requires consistent quality assurance, adequate staffing, and measures ensuring equal access to care.

Third, specialist medical trainees play a vital role as agents of change within both the education system and the broader health care landscape. Students must be equipped to navigate contemporary medical challenges and encouraged to engage in system improvement. Sufficient institutional support will enable them to contribute effectively to positive transformation in health services.

Fourth, the findings emphasize that oversight and governance in specialist medical education must be strengthened. Historically, weaknesses in program supervision have received attention only after issues were identified by the Ministry of Health. This implies a strong need for more rigorous monitoring by both government and universities.

Fifth, reliance solely on university-based specialist medical training may limit the number and diversity of specialists needed nationwide. Expanding hospital-based specialist medical training is therefore essential. Since the number of hospitals exceeds the number of medical schools, such an expansion could broaden access, improve specialist distribution to underserved regions, and reduce the training burden on universities. Increasing practitioner involvement in specialist medical education would also enrich training quality. Strengthening hospital-based pathways represents an important step toward addressing disparities and enhancing the national medical education system.

Overall, these implications highlight the need for systemic adjustments in medical education and health care delivery. Recognizing the presence of the Icarus Paradox helps identify areas for improvement and guides targeted reforms. Positive aspects, such as the growing role of hospital-based training, should be sustained and expanded, while structural weaknesses, such as inadequate governance and uneven specialist distribution, must be addressed. With joint commitment from the government, health care institutions, academics, and society, Indonesia can build a more inclusive, sustainable, and community-oriented health system [26], ultimately improving medical education and health service outcomes nationwide.

Limitations and Further Research

Despite offering valuable insights into the Icarus Paradox within medical education and health services, this study has several limitations. First, the findings rely heavily on online review data, which may not represent the full population. Individuals without internet access or those who do not use online review platforms may be underrepresented, and the anonymous nature of such reviews may affect data reliability.

Second, the study primarily captures community perceptions of the medical education system and specialist medical trainees, which introduces potential subjective bias. Different individuals may express varying experiences and expectations, which may influence sentiment data.

Third, this study does not directly measure clinical performance indicators or objective metrics of service quality. The data analyzed are predominantly qualitative and descriptive. To obtain a more comprehensive understanding of health service quality, future research should integrate quantitative measures with qualitative insights.

Further research can address these limitations in several ways. First, studies may incorporate broader respondent groups, including patients [25], medical practitioners, and students, through structured surveys or interviews to enhance representativeness. Second, advanced analytical techniques, such as deeper text mining or longitudinal sentiment tracking, can identify emerging patterns in digital public discourse. Third, expanding the study scope to examine interactions between

variables, such as health care access, medical training quality, and patient satisfaction, over extended time periods (beyond 30 days) will enrich understanding.

Finally, empirical testing of the proposed conceptual model can offer deeper insights into the complex dynamics of Indonesia's health system. Such research can guide policymakers and educational institutions in designing interventions that enhance system effectiveness and mitigate paradoxical tensions. Ultimately, further investigation will contribute to a more comprehensive understanding of strategies needed to strengthen medical education and health care delivery in Indonesia.

Conclusion

Indonesia's specialist medical education system continues to face complex challenges that can be understood through the lens of the Icarus Paradox, where strengths that drive success may simultaneously give rise to failure. Ambition, resilience, and rigorous training enable medical students to pursue excellence, yet these same strengths can lead to burnout, intimidation, and psychological strain when not supported by adequate institutional resources. As Eriksson et al [7] describe, such dynamics reflect an "inverted Icarus" pattern in which progress carries inherent vulnerabilities. Prospective specialist medical students therefore stand at a critical juncture, symbolizing both the promise of advancement and the risks embedded in systemic pressure.

Health service quality also reflects this tension. Although national initiatives, such as infrastructure expansion and the JKN universal health coverage program, have increased access, disparities remain significant. For example, although the doctor-to-population ratio in Jakarta is approximately 1:1000, in some rural provinces, such as East Nusa Tenggara, the ratio drops below 1:7000 [72]. These inequalities illustrate how the push to expand services may exceed the capacity to ensure equitable quality of care. In education, national surveys indicate that more than half of PPDS students report mental health distress and excessive workloads, demonstrating that high standards often translate into substantial personal cost.

Addressing this paradox requires coordinated and targeted reforms. At the policy level, priority actions include (1) redistributing specialist doctors through structured incentives for underserved regions; (2) revising PPDS curricula to incorporate mental health resilience, mentorship, and leadership training; (3) expanding investment in simulation-based learning to reduce overreliance on hospital workloads; and (4) strengthening regulations against intimidation and unsafe learning environments.

At the same time, progress is attainable. Initiatives, such as scholarships for PPDS candidates from remote regions, the development of eLearning platforms in medical schools, and improvements in provincial teaching hospitals, indicate ongoing reform efforts. These programs show that although paradoxical tensions persist, Indonesia is gradually moving toward solutions that balance ambition with sustainable practice.

With collaborative efforts involving government, universities, hospitals, and society, the Icarus Paradox in medical education and health services can be mitigated. This will support

Indonesia's movement toward a more inclusive, equitable, and resilient health care system, one in which the aspiration to produce excellent medical professionals does not compromise their well-being or widen disparities in health service delivery.

Acknowledgments

The authors received no financial support for the research, authorship, or publication of this manuscript. This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' Contributions

FB was responsible for conceptualization, methodology, data curation, formal analysis, writing—original draft, resources, supervision, and visualization and MH for conceptualization, validation, writing—review and editing, supervision, and formal analysis. All authors have read and approved the final manuscript.

Generative artificial intelligence (ChatGPT) was used only for language refinement and did not contribute to the analysis, interpretation, or generation of research data.

Conflicts of Interest

None declared.

References

1. Sulphey MM. How Icarus paradox doomed Kingfisher Airlines. *Vision: J Bus Perspect* 2020 Feb 28;24(1):118-124 [FREE Full text] [doi: [10.1177/0972262919897646](https://doi.org/10.1177/0972262919897646)]
2. Mahendradhata Y, Trisniantoro L, Listyadewi S, Soewondo P, Marthias T, Harimurti P, et al. The Republic of Indonesia health system review. In: Hort K, Patcharanarumol W, editors. *Health Systems in Transition*, Vol-7, Number-1. New Delhi: World Health Organization (Asia Pacific Observatory on Health Systems and Policies); 2017.
3. Wallin G. An introduction to R and Python for data analysis: a side-by-side approach. *Am Stat* 2024 Apr 17;78(2):265 [FREE Full text] [doi: [10.1080/00031305.2024.2320949](https://doi.org/10.1080/00031305.2024.2320949)]
4. Bondestam F, Lundqvist M. Sexual harassment in higher education – a systematic review. *Eur J High Educ* 2020 Feb 18;10(4):397-419 [FREE Full text] [doi: [10.1080/21568235.2020.1729833](https://doi.org/10.1080/21568235.2020.1729833)]
5. Heffernan T, Bosetti L. Incivility: the new type of bullying in higher education. *Cambridge J Educ* 2021 Mar 21;51(5):641-652 [FREE Full text] [doi: [10.1080/0305764x.2021.1897524](https://doi.org/10.1080/0305764x.2021.1897524)]
6. Yousaf R, Schmiede R. Harassment Act implementation in higher education institutions. *Open J Leadersh* 2016;05(01):8-19 [FREE Full text] [doi: [10.4236/ojl.2016.51002](https://doi.org/10.4236/ojl.2016.51002)]
7. Eriksson K, Lakomaa E, Nykvist R, Sandström C. Introducing the inverted Icarus paradox in business history – evidence from David and Goliath in the Swedish telecommunications industry 1981–1990. *Bus Hist* 2024 Mar 05;67(2):551-576 [FREE Full text] [doi: [10.1080/00076791.2023.2292134](https://doi.org/10.1080/00076791.2023.2292134)]
8. Desveaux L. Rethinking healthcare: why paradox science is core to the future of health and health leadership. *JHL* 2025 Feb 21;17:45-48 [FREE Full text] [doi: [10.2147/jhl.s512053](https://doi.org/10.2147/jhl.s512053)]
9. Smith WK, Lewis MW. Toward a theory of paradox: a dynamic equilibrium model of organizing. *Acad Manag Rev* 2011 Apr;36(2):381-403 [FREE Full text] [doi: [10.5465/amr.2009.0223](https://doi.org/10.5465/amr.2009.0223)]
10. Lili R, Molodynki A, Farrell SM, Citraningtyas T, Kloping NA. Wellbeing and mental health among medical students in Indonesia: a descriptive study. *Int J Soc Psychiatry* 2022 Sep 25;68(6):1277-1282 [FREE Full text] [doi: [10.1177/00207640211057709](https://doi.org/10.1177/00207640211057709)] [Medline: [34823405](https://pubmed.ncbi.nlm.nih.gov/34823405/)]
11. Dyah Perwitasari V, Hidayat R. Depression level among neurology resident doctors in the Faculty of Medicine, Universitas Indonesia. *Acta Neurol Indones* 2024 Apr 17;2(01):60-71 [FREE Full text] [doi: [10.69868/ani.v2i01.25](https://doi.org/10.69868/ani.v2i01.25)]
12. Arlinta D. Depresi, 3,3 Persen Calon Dokter Spesialis Ingin Akhiri Hidup atau Lukai Diri. *Kompas*. Jakarta: Kompas; 2024 Apr 15. URL: <https://www.kompas.id/artikel/224-persen-mahasiswa-pendidikan-dokter-spesialis-alami-gejala-depresi-sebagian-merasa-lebih-baik-mengakhiri-hidup> [accessed 2026-01-22]
13. Kesaksian calon dokter spesialis yang sempat berusaha bunuh diri – “perundungan dijustifikasi atas nama pendidikan mental. *BBC News Indonesia*. 2024 Apr. URL: <https://www.bbc.com/indonesia/articles/cyj3mkp7jl0o> [accessed 2026-01-20]
14. 2.716 Calon dokter spesialis mengalami gejala depresi, 3,3% ingin akhiri hidup. *Metro TV News*. 2024 Apr. URL: <https://www.metrotvnews.com/play/b3JCr9d2-2-716-calon-dokter-spesialis-mengalami-gejala-depresi-3-3-ingin-akhiri-hidup> [accessed 2026-01-20]
15. Widayanti AW, Green JA, Heydon S, Norris P. Health-seeking behavior of people in Indonesia: a narrative review. *J Epidemiol Glob Health* 2020 Mar;10(1):6-15 [FREE Full text] [doi: [10.2991/jegh.k.200102.001](https://doi.org/10.2991/jegh.k.200102.001)] [Medline: [32175705](https://pubmed.ncbi.nlm.nih.gov/32175705/)]
16. Martina Ws Nasrun. Depression Phenomenon in PPDS, Fact or Mirage? *Kompas*. Kompas: Kompas; 2024 May 10. URL: <https://www.kompas.id/artikel/en-fenomena-depresi-di-ppds-fakta-atau-fatamorgana-1> [accessed 2026-01-21]

17. Ranasinghe PD, Owusu JT, Bertram A, Michalik H, Yeh H, Cofrancesco J, et al. Depressive symptoms and burnout among medical students: a prospective study. *J Gen Intern Med* 2022 Jan 26;37(1):64-69 [FREE Full text] [doi: [10.1007/s11606-021-06765-x](https://doi.org/10.1007/s11606-021-06765-x)] [Medline: [34037922](#)]
18. Doe S, Coutinho AJ, Weidner A, Cheng Y, Sanders K, Bazemore AW, et al. Prevalence and predictors of burnout among resident family physicians. *Fam Med* 2024 Mar 1;56(3):148-155 [FREE Full text] [doi: [10.22454/fammed.2024.875388](https://doi.org/10.22454/fammed.2024.875388)]
19. Shalaby R, Oluwasina F, Eboreime E, El Gindi H, Agyapong B, Hrabok M, et al. Burnout among residents: prevalence and predictors of depersonalization, emotional exhaustion and professional unfulfillment among resident doctors in Canada. *Int J Environ Res Public Health* 2023 Feb 19;20(4):3677 [FREE Full text] [doi: [10.3390/ijerph20043677](https://doi.org/10.3390/ijerph20043677)] [Medline: [36834373](#)]
20. Obeng Nkrumah S, Adu MK, Agyapong B, da Luz Dias R, Agyapong VIO. Prevalence and correlates of depression, anxiety, and burnout among physicians and postgraduate medical trainees: a scoping review of recent literature. *Front Public Health* 2025 Jul 8;13:1537108 [FREE Full text] [doi: [10.3389/fpubh.2025.1537108](https://doi.org/10.3389/fpubh.2025.1537108)] [Medline: [40697832](#)]
21. 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 May 15;11:e58008 [FREE Full text] [doi: [10.2196/58008](https://doi.org/10.2196/58008)] [Medline: [40372966](#)]
22. Lee KP, Yeung N, Wong C, Yip B, Luk LHF, Wong S. Prevalence of medical students' burnout and its associated demographics and lifestyle factors in Hong Kong. *PLoS One* 2020 Jul 10;15(7):e0235154 [FREE Full text] [doi: [10.1371/journal.pone.0235154](https://doi.org/10.1371/journal.pone.0235154)] [Medline: [32649681](#)]
23. Thamissarakul S, Hongkan W, Wannapaschayieng P. Factors associated with burnout syndrome among clinical medical students at Chonburi Hospital, Thailand. *J Health Sci Med Res* 2024;42(3):1-6 [FREE Full text] [doi: [10.31584/jhsmr.20231020](https://doi.org/10.31584/jhsmr.20231020)]
24. Caring for the mental health of the medical workforce. British Medical Association. 2019. URL: <https://www.bma.org.uk/media/ckshvkzc/bma-mental-health-survey-report-september-2024.pdf> [accessed 2026-01-20]
25. Berger S, Saut AM, Berssaneti FT. Using patient feedback to drive quality improvement in hospitals: a qualitative study. *BMJ Open* 2020 Oct 23;10(10):e037641 [FREE Full text] [doi: [10.1136/bmjopen-2020-037641](https://doi.org/10.1136/bmjopen-2020-037641)] [Medline: [33099495](#)]
26. Bismantara H, Ahern S, Teede HJ, Liew D. Academic health science centre models across the developing countries and lessons for implementation in Indonesia: a scoping review. *BMJ Open* 2022 Sep 06;12(9):e051937 [FREE Full text] [doi: [10.1136/bmjopen-2021-051937](https://doi.org/10.1136/bmjopen-2021-051937)] [Medline: [36691121](#)]
27. Abid A, Harrigan P, Roy S. A relationship marketing orientation in politics: young voters' perceptions of political brands' use of social media. *Journal of Strategic Marketing* 2020 Jun 12;29(4):359-374 [FREE Full text] [doi: [10.1080/0965254x.2020.1777457](https://doi.org/10.1080/0965254x.2020.1777457)]
28. Ashraf MA, Khan MN, Chohan SR, Khan M, Rafique W, Farid MF, et al. Social media improves students' academic performance: exploring the role of social media adoption in the open learning environment among international medical students in China. *Healthcare (Basel)* 2021 Sep 26;9(10):1272 [FREE Full text] [doi: [10.3390/healthcare9101272](https://doi.org/10.3390/healthcare9101272)] [Medline: [34682952](#)]
29. Noya FC, Carr SE, Thompson SC. Attracting, recruiting, and retaining medical workforce: a case study in a remote province of Indonesia. *Int J Environ Res Public Health* 2023 Jan 12;20(2):1435 [FREE Full text] [doi: [10.3390/ijerph20021435](https://doi.org/10.3390/ijerph20021435)] [Medline: [36674191](#)]
30. Schad J, Lewis MW, Raisch S, Smith WK. Paradox research in management science: looking back to move forward. *Acad Manag Ann* 2016 Jan;10(1):5-64 [FREE Full text] [doi: [10.5465/19416520.2016.1162422](https://doi.org/10.5465/19416520.2016.1162422)]
31. West J, Gallagher S. Challenges of open innovation: the paradox of firm investment in open-source software. *R D Manag* 2006 Jun;36(3):319-331 [FREE Full text] [doi: [10.1111/j.1467-9310.2006.00436.x](https://doi.org/10.1111/j.1467-9310.2006.00436.x)]
32. Oliveira P, e Cunha MP. Centralized decentralization, or distributed leadership as paradox: the case of the Patient Innovation's COVID-19 portal. *J Chang Manag* 2021 May 09;21(2):203-221 [FREE Full text] [doi: [10.1080/14697017.2021.1917493](https://doi.org/10.1080/14697017.2021.1917493)]
33. Klein SP, Spieth P, Söllner M. Employee acceptance of digital transformation strategies: a paradox perspective. *J Prod Innov Manag* 2024 Jan 10;41(5):999-1021 [FREE Full text] [doi: [10.1111/jpim.12722](https://doi.org/10.1111/jpim.12722)]
34. Cunha MPE, Putnam LL. Paradox theory and the paradox of success. *Strateg Organ* 2017 Oct 31;17(1):95-106 [FREE Full text] [doi: [10.1177/1476127017739536](https://doi.org/10.1177/1476127017739536)]
35. Miller D. The Icarus paradox: how exceptional companies bring about their own downfall. *Bus Horizons* 1992 Jan;35(1):24-35 [FREE Full text] [doi: [10.1016/0007-6813\(92\)90112-M](https://doi.org/10.1016/0007-6813(92)90112-M)]
36. Dowds B. Depression and the Erosion of the Self in Late Modernity. London, New York: Routledge, Taylor and Francis; 2018.
37. Sherrer D, Franklin A, Kimatian S, Black I, Tsai M. The Icarus paradox and the future of anesthesiology. *Anesth Analg* 2023 Jan 01;136(1):185-189 [FREE Full text] [doi: [10.1213/ANE.0000000000006253](https://doi.org/10.1213/ANE.0000000000006253)] [Medline: [36534720](#)]
38. Biçer C. The Icarus paradox in management: how to be a well-balanced leader? *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Derg* 2021;11(4):1891-1905 [FREE Full text] [doi: [10.30783/nevsosbilen.892797](https://doi.org/10.30783/nevsosbilen.892797)]
39. Vermeulen F. Businesses and the Icarus paradox. *Harvard Business Review*. 2009 Mar 5. URL: <https://hbr.org/2009/03/businesses-and-the-icarus-paradox> [accessed 2026-01-20]

40. Zreik M. The paradox of educational inequality in Indonesia: socioeconomic implications and paths towards inclusion. *Socio-Economic Implic Glob Educ Inequalities* 2024;69-85 [FREE Full text] [doi: [10.4018/979-8-3693-0693-2.ch004](https://doi.org/10.4018/979-8-3693-0693-2.ch004)]

41. Winzer R, Lindberg L, Guldbrandsson K, Sidorchuk A. Effects of mental health interventions for students in higher education are sustainable over time: a systematic review and meta-analysis of randomized controlled trials. *PeerJ* 2018;6:e4598 [FREE Full text] [doi: [10.7717/peerj.4598](https://doi.org/10.7717/peerj.4598)] [Medline: [29629247](#)]

42. Rotenstein LS, Ramos MA, Torre M, Segal JB, Peluso MJ, Guille C, et al. Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: a systematic review and meta-analysis. *JAMA* 2016 Dec 06;316(21):2214-2236 [FREE Full text] [doi: [10.1001/jama.2016.17324](https://doi.org/10.1001/jama.2016.17324)] [Medline: [27923088](#)]

43. Shapiro J. The paradox of teaching empathy in medical education. In: Decety J, editor. *Empathy: From Bench to Bedside*. Cambridge, MA: MIT Press; 2011:275-290.

44. Brown T, Williams B, McKenna L, Palermo C, McCall L, Roller L, et al. Practice education learning environments: the mismatch between perceived and preferred expectations of undergraduate health science students. *Nurse Educ Today* 2011 Nov;31(8):e22-e28 [FREE Full text] [doi: [10.1016/j.nedt.2010.11.013](https://doi.org/10.1016/j.nedt.2010.11.013)] [Medline: [21131106](#)]

45. Tess PA. The role of social media in higher education classes (real and virtual) – a literature review. *Comput Hum Behav* 2013 Sep;29(5):A60-A68 [FREE Full text] [doi: [10.1016/j.chb.2012.12.032](https://doi.org/10.1016/j.chb.2012.12.032)] [Medline: [21131106](#)]

46. Sivakumar A, Jayasingh S, Shaik S. Social media influence on students' knowledge sharing and learning: an empirical study. *Educ Sci* 2023 Jul 20;13(7):745 [FREE Full text] [doi: [10.3390/educsci13070745](https://doi.org/10.3390/educsci13070745)]

47. Wilson C, McDarby V. Social media and mental health. *Clin Child Psychol Psychiatry* 2023 Jan 28;28(1):157-160 [FREE Full text] [doi: [10.1177/13591045221144926](https://doi.org/10.1177/13591045221144926)] [Medline: [36579367](#)]

48. Firdos S, Almulla S, Aldossary S, Al Hassan S, Aldhaif L. Exploring the attitudes of medical students towards social media and e-professionalism in Al-Ahsa, Saudi Arabia. *Cureus* 2023 Nov;15(11):e48718 [FREE Full text] [doi: [10.7759/cureus.48718](https://doi.org/10.7759/cureus.48718)] [Medline: [38094527](#)]

49. Terry K, Yang F, Yao Q, Liu C. The role of social media in public health crises caused by infectious disease: a scoping review. *BMJ Glob Health* 2023 Dec 28;8(12):e013515 [FREE Full text] [doi: [10.1136/bmjgh-2023-013515](https://doi.org/10.1136/bmjgh-2023-013515)] [Medline: [38154810](#)]

50. Binsar F, Arief M, Tjhin VU, Susilowati I. Exploring consumer sentiments in telemedicine and telehealth services: towards an integrated framework for innovation. *J Open Innov Technol Mark Complex* 2025 Mar;11(1):100453 [FREE Full text] [doi: [10.1016/j.joitmc.2024.100453](https://doi.org/10.1016/j.joitmc.2024.100453)]

51. Kim D, Jung W, Jiang T, Zhu Y. An exploratory study of medical journal's Twitter use: metadata, networks, and content analyses. *J Med Internet Res* 2023 Jan 19;25:e43521 [FREE Full text] [doi: [10.2196/43521](https://doi.org/10.2196/43521)] [Medline: [36656626](#)]

52. Creswell JW, Creswell JD. In: Fetts DC, editor. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Fifth Edition. Thousand Oaks, CA: Sage Publications; 2018.

53. Hutagalung SS, Kartika T, Suciska W. Media monitoring analysis of government image in infrastructure development in Indonesia. *J Komun* 2023 Jul 27;15(1):212-227 [FREE Full text] [doi: [10.24912/jk.v15i1.20605](https://doi.org/10.24912/jk.v15i1.20605)]

54. Hadi SP, Ibrahim MH, Prabawani B, Hamdani RS. Environmental dimension of pandemic COVID-19: case studies of Indonesia. *IOP Conf Ser Earth Environ Sci* 2021 Jan 01;623(1):012051 [FREE Full text] [doi: [10.1088/1755-1315/623/1/012051](https://doi.org/10.1088/1755-1315/623/1/012051)]

55. Punziano G, De Falco CC, Trezza D. Digital mixed content analysis for the study of digital platform social data: an illustration from the analysis of COVID-19 risk perception in the Italian Twittersphere. *J Mix Methods Res* 2023 Apr 21;17(2):143-170 [FREE Full text] [doi: [10.1177/15586898211067647](https://doi.org/10.1177/15586898211067647)] [Medline: [38603067](#)]

56. Chhabra J, Pilkington V, Benakovic R, Wilson MJ, La Sala L, Seidler Z. Social media and youth mental health: scoping review of platform and policy recommendations. *J Med Internet Res* 2025 Jun 20;27:e72061 [FREE Full text] [doi: [10.2196/72061](https://doi.org/10.2196/72061)] [Medline: [40540734](#)]

57. Zhou Z, Jin D, He J, Zhou S, Wu J, Wang S, et al. Digital health platform for improving the effect of the active health management of chronic diseases in the community: mixed methods exploratory study. *J Med Internet Res* 2024 Nov 18;26:e50959 [FREE Full text] [doi: [10.2196/50959](https://doi.org/10.2196/50959)] [Medline: [39556830](#)]

58. Binsar F, Mauritsius T. Mining of social media on Covid-19 big data infodemic in Indonesia. *J Comput Sci* 2020 Nov 01;16(11):1598-1609 [FREE Full text] [doi: [10.3844/jcssp.2020.1598.1609](https://doi.org/10.3844/jcssp.2020.1598.1609)]

59. Lee YU, Chung SH, Park JY. Online review analysis from a customer behavior observation perspective for product development. *Sustainability* 2024 Apr 24;16(9):3550 [FREE Full text] [doi: [10.3390/su16093550](https://doi.org/10.3390/su16093550)]

60. Limna P. The impact of NVivo in qualitative research: perspectives from graduate students. *J Appl Learn Teach* 2023 Aug 2;6(2):271-282 [FREE Full text] [doi: [10.37074/jalt.2023.6.2.17](https://doi.org/10.37074/jalt.2023.6.2.17)]

61. Dereń K, Rajda K. How to master AI-powered sentiment analysis in 2025? Brand24. 2025. URL: <https://brand24.com/blog/sentiment-analysis/> [accessed 2025-08-19]

62. Jackson K, Bazeley P. Qualitative Data Analysis with NVivo. 3rd Edition. London, California, New Delhi, Singapore: Sage Publications; 2019:1-346.

63. Binsar F, Legowo N. Design of cloud computing outpatient registration model through SMS messages at hospitals using TOGAF ADM. *Int J Recent Technol Eng* 2020 Jan 30;8(5):3857-3865 [FREE Full text] [doi: [10.35940/ijrte.E4885.018520](https://doi.org/10.35940/ijrte.E4885.018520)]

64. Binsar F, Kartono R, Bandur A, Kosasih W. Digital transformation of information fulfillment and patient engagement for health service safety. 2022 Jul 18 Presented at: 4th International Conference on Management Science and Industrial Engineering; April 28-30, 2022; Chiang Mai, Thailand p. 229-236. [doi: [10.1145/3535782.3535813](https://doi.org/10.1145/3535782.3535813)]
65. Mustika R, Nishigori H, Ronokusumo S, Scherbier A. The odyssey of medical education in Indonesia. Asia Pacific Sch 2019 Jan 2;4(1):4-8 [FREE Full text] [doi: [10.29060/taps.2019-4-1/gp1077](https://doi.org/10.29060/taps.2019-4-1/gp1077)]
66. Zechariah S, Ansa BE, Johnson SW, Gates AM, Leo GD. Interprofessional education and collaboration in healthcare: an exploratory study of the perspectives of medical students in the United States. Healthcare (Basel) 2019 Oct 15;7(4):117 [FREE Full text] [doi: [10.3390/healthcare7040117](https://doi.org/10.3390/healthcare7040117)] [Medline: [31618920](https://pubmed.ncbi.nlm.nih.gov/31618920/)]
67. Hofmann B. The paradox of health care. Health Care Anal 2001 Dec;9(4):369-386 [FREE Full text] [doi: [10.1023/a:1013854030699](https://doi.org/10.1023/a:1013854030699)]
68. Sutrisno MD. Shortages of medical doctors in indonesia, is it true? Asian J Heal Res 2023 Aug 11;2(2):2828-2829 [FREE Full text] [doi: [10.55561/ajhr.v2i2.121](https://doi.org/10.55561/ajhr.v2i2.121)]
69. Risky YPA, Projo NWK. The effect of the digital economy on indecent work in Indonesia 2019. Jakarta; 2022 Jan 04 Presented at: 2021 International Conference on Data Science and Official Statistics (ICDSOS); November 13-14, 2021; Jakarta, Indonesia. [doi: [10.34123/icdsos.v2021i1.88](https://doi.org/10.34123/icdsos.v2021i1.88)]
70. Awofeso N, Rammohan A, Asmaripa A. Exploring Indonesia's "low hospital bed utilization-low bed occupancy-high disease burden" paradox. J Hosp Adm 2012 Nov 22;2(1):49 [FREE Full text] [doi: [10.5430/jha.v2n1p49](https://doi.org/10.5430/jha.v2n1p49)]
71. Binsar F, Riantono I, Kartono R, Bandur A, Kosasih W. Gamification to grow motivation for interactive engagement of health nurses in using health information systems: a conceptual framework. 2023 Presented at: Asia Pacific Computer Systems Conference 2021; February 18, 2023; North Carolina, USA p. 99-120. [doi: [10.1007/978-981-19-7904-0_10](https://doi.org/10.1007/978-981-19-7904-0_10)]
72. Putri LP, Russell DJ, O'Sullivan BG, Kippen R. Factors associated with working in remote Indonesia: a national cross-sectional study of early-career doctors. Front Med (Lausanne) 2021 May 13;8:594695 [FREE Full text] [doi: [10.3389/fmed.2021.594695](https://doi.org/10.3389/fmed.2021.594695)] [Medline: [34055819](https://pubmed.ncbi.nlm.nih.gov/34055819/)]

Abbreviations

JKN: Jaminan Kesehatan Nasional (National Health Insurance Program)

PPDS: *Program Pendidikan Dokter Spesialis* (Specialist Doctor Education Program)

Edited by J Gentges; submitted 11.May.2024; peer-reviewed by PH Sebong, NA Pambudi; comments to author 15.May.2025; accepted 17.Sep.2025; published 26.Jan.2026.

Please cite as:

Binsar F, Hamsal M

Exploring the Icarus Paradox in Indonesia's Specialist Medical Education System Using the Public Perspective From Online Media: Convergent Mixed Methods Study

JMIR Med Educ 2026;12:e60452

URL: <https://mededu.jmir.org/2026/1/e60452>

doi: [10.2196/60452](https://doi.org/10.2196/60452)

PMID:

©Faisal Binsar, Mohammad Hamsal. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 26.Jan.2026. 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.

Investigating the Impact of a Virtual Reality Experience on Medical Student Empathy: Mixed Methods Study

Allen G Mundok^{1*}, BS; Vivian N Ho^{1*}, BS; Lauren A Fowler^{2*}, PhD; Ann Blair Kennedy^{1*}, DrPH; Shannon Stark-Taylor^{1*}, PhD

¹Department of Biomedical Sciences, University of South Carolina School of Medicine Greenville, 607 Grove Rd, Greenville, SC, United States

²Department of Translational Neuroscience, Wake Forest University School of Medicine Charlotte, Charlotte, NC, United States

* all authors contributed equally

Corresponding Author:

Ann Blair Kennedy, DrPH

Department of Biomedical Sciences, University of South Carolina School of Medicine Greenville, 607 Grove Rd, Greenville, SC, United States

Abstract

Background: Physician empathy is important not only for improving patient satisfaction and health outcomes but also for increasing physician job satisfaction and protecting against burnout. However, amid concerns over declining empathy levels in medical education, there is a need for innovative teaching approaches that address the empathy gap, a critical element in patient-centered care.

Objective: This study aimed to use a mixed-methods analysis to explore the effectiveness of a virtual reality (VR) intervention versus traditional lecture methods in enhancing empathy among medical students.

Methods: Overall, 50 first- and second-year medical students were randomized to either a VR intervention, which simulated patient experiences, or a control group receiving traditional empathy lectures. Both groups watch 2 videos with reflections gathered after each video to capture students' experiential learning. Empathy was measured using the Jefferson Scale of Empathy-Student Version before and after the intervention.

Results: Quantitative analysis revealed significant increases in empathy scores post intervention for both groups (lecture group: mean increase 4.71, SD 11.01; VR group: mean increase 5.6, SD 10.02; $P<.001$), indicating that both interventions enhanced empathy. The VR group exhibited a significant difference in qualitative empathy coding after the second video ($U=165.5$; $P<.001$) compared to the lecture group. Qualitative feedback from the VR group emphasized a more profound emotional and cognitive engagement with the patient perspective than the lecture group.

Conclusions: This study supports the integration of VR into medical education as a complementary approach to traditional teaching methods for empathy training. VR immersion provides a valuable platform for students to develop a deeper, more nuanced understanding of empathy. These findings advocate for further exploration into VR's long-term impact on empathy in clinical practice.

(*JMIR Med Educ* 2026;12:e76504) doi:10.2196/76504

KEYWORDS

medical education; virtual reality; medical students; mixed-methods; empathy

Introduction

Empathy can be interpreted as the emotional engagement of the observer with the circumstances of another [1]. In medical settings, higher physician empathy has been associated with improved patient outcomes, including greater patient satisfaction and increased adherence to medical recommendations and treatments [2]. For example, individuals with diabetes whose physicians score higher on measures of empathy demonstrate better glucose and cholesterol control than those whose physicians score lower, controlling for other physician and patient factors [3]. For physicians, empathy has also been linked

to higher supervisor ratings of clinical competence and fewer malpractice claims [4,5]. Despite the known benefits of empathy, research also shows that empathy tends to decline during medical school, particularly during clinical training, with male students showing larger decreases [5]. These studies highlight the importance of understanding how empathy develops and changes throughout the course of one's medical education and suggest a need for interventions to support its maintenance.

Studies have demonstrated the effectiveness of many different modes of empathy training for increasing empathy ratings in medical students [6,7]. To train individuals in empathy, traditional interventions have often taken the form of

perspective-taking. These interventions involve having participants imagine another's situation and put themselves "in that person's shoes." Perspective-taking has also been shown to promote empathy in a range of nonmedical contexts [8-10].

Compared to traditional perspective-taking, virtual reality (VR) interventions offer a fully immersive experience, giving the user the visual and emotional feeling of being present for the interaction and embodying another [11,12]. One study compared traditional perspective-taking to a VR perspective-taking intervention using a virtual environment to promote increased empathy for homeless persons. Participants who underwent the VR intervention showed more positive, longer-lasting attitudes toward the homeless and signed a petition supporting the homeless at a significantly higher rate than those who engaged in a traditional perspective-taking task [11]. Studies also suggest VR interventions are less cognitively taxing than imagining another's perspective and allow for improved methodology as all participants engage in the same exercise [13-15].

Despite this evidence, few studies have applied VR technology to teaching medical students and physicians. With medical students increasingly using technology resources to enhance their classroom learning [16,17], finding ways to integrate technology into empathy training may be an effective method to engage medical students [18,19]. Given the benefits of immersive perspective-taking in other contexts, we designed an innovative VR experience to facilitate patient perspective-taking among medical students and compared it to traditional teaching methods (a video lecture). Based on prior findings promoting empathy for several different human conditions in lay populations and the theoretical advantages of VR for perspective-taking, we hypothesize that engaging in patient perspective-taking through VR will result in greater increases in empathy among medical students than the traditional teaching method.

Methods

Study Design and Video Development

To compare the efficacy of a VR experience to traditional lecture learning, we used a randomized controlled mixed-methods convergent parallel design. In this study design, both qualitative (structured reflection responses) and quantitative data (Jefferson Scale of Empathy for medical students [JSE-S] pre- and posttest

scores) were collected simultaneously in a single phase. Each type of data were analyzed separately, then merged to identify points of convergence [20]. This design allows us to integrate the strengths of both approaches, providing a more comprehensive understanding of how the VR intervention influenced empathy, while also allowing the qualitative data to contextualize and enrich the quantitative findings [20].

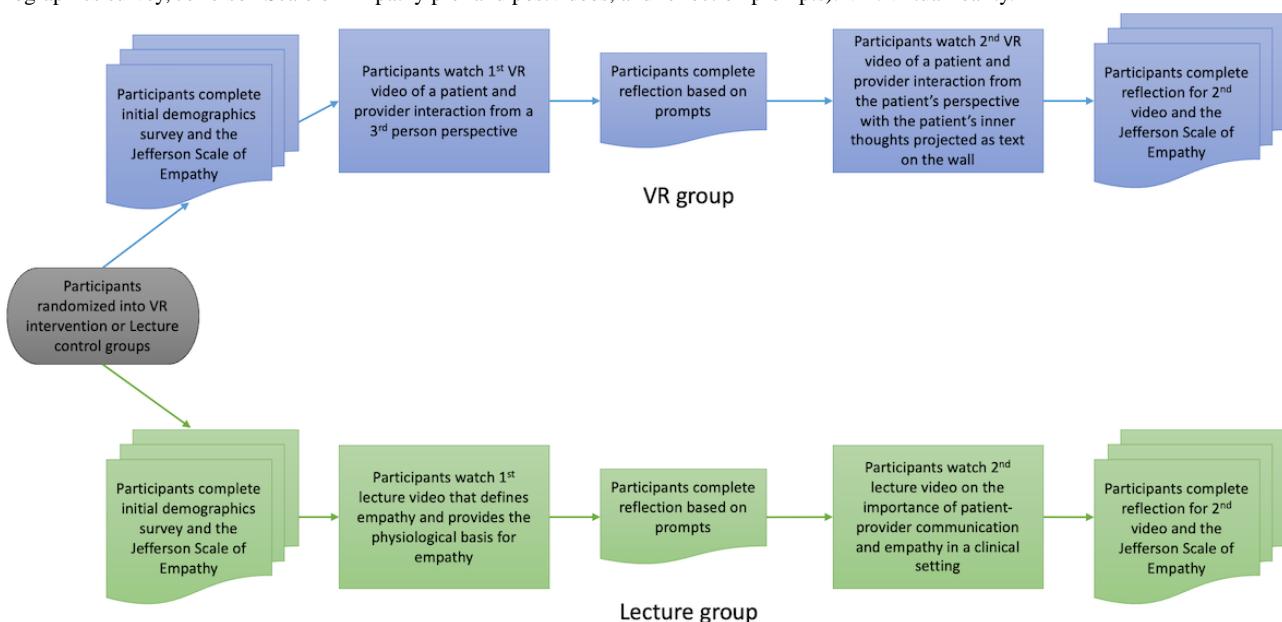
The lecture videos were PowerPoint lectures designed and narrated by a biomedical sciences professor and research team member (LAF) with expertise in empathy. Lecture video 1 (V1) described the science behind empathy and its importance in health care. Lecture video 2 (V2) provided strategies for students to enhance empathy in clinical practice.

To make the VR videos and reflection questions as authentic as possible, these items were presented to a panel of patients and providers trained in giving feedback to researchers. The panel members made suggestions for script and question improvements. After incorporating the received feedback into the revised script, the VR videos were filmed using a Garmin VIRB Ultra 360 camera (Garmin, Ltd.) with hired actors playing the roles of a patient and physician. VR Video 1 presented a third-person perspective of a "normal" patient-physician interaction during a routine office visit. VR Video 2 showed the same interaction but through the first-person perspective of the patient, and the patient's inner thoughts were projected as text on the screen for students to read. An example of the patient's thoughts is, "How am I going to afford a dietitian? Will my insurance pay for that?" Videos for the lecture and VR group were designed to be the same length to control for time.

Data Collection

Using a random number generator, students were randomly assigned to either the lecture group (n=25) or VR group (n=25). In the lecture group, students were given a tablet to watch the 2 PowerPoint lecture videos on empathy. In the VR group, students wore a VR headset to watch the 2 VR videos depicting a patient-physician interaction. Prior to watching their respective videos, students completed an initial demographics survey and the Jefferson Scale of Empathy for medical students (JSE-S). After watching the first video, students answered 3 open-ended reflection questions on their learning experience. Students then watched the second video and again completed the 3 reflection prompts and the JSE-S. [Figure 1](#) depicts this methodology.

Figure 1. Study design overview. Flowchart depicting participant randomization into virtual reality or lecture video group and timeline of assessment (demographics survey, Jefferson Scale of Empathy pre- and postvideos, and reflection prompts). VR: virtual reality.



Throughout their viewing of the videos, students in both groups were also equipped with Empatica E4 wristbands (Empatica, Inc.) to measure physiological markers of empathy such as galvanic skin response and heart rate variability. While physiological markers have been used in prior research to complement self-reported empathy measures, technical artifacts in our dataset rendered the recordings uninterpretable; these data were therefore not included in our final results [21].

Instruments

The validity of the JSE in evaluating empathy in physicians and medical students has been well-established [22,23]. We used the student version (JSE-S), which was developed to measure medical students' attitudes toward empathy in patient-physician interactions. The JSE comprises 20 items, rated on a 7-point Likert-type scale (1=strongly disagree, 7=strongly agree), with higher total scores indicating greater empathy (maximum score=140). An example of a typical item is, "Patients feel better when their health care providers understand their feelings." The JSE-S has demonstrated strong internal consistency, construct validity, and predictive associations with clinical competence and patient-centered outcomes [24,25]. Although the JSE-S primarily assesses attitudes toward empathy rather than direct empathic behavior, it is widely accepted as a valid proxy for empathy in medical education research [26]. Empathy in clinical contexts is understood as a multidimensional construct encompassing cognitive, affective, and attitudinal components [22]. The attitudinal dimension captured by the JSE-S reflects physicians' and students' willingness and value orientation toward understanding patients' experiences, which are key determinants of empathetic engagement in practice. Thus, while the JSE-S measures empathy-related attitudes, its theoretical foundation and empirical validation support its continued use as a reliable instrument for assessing empathy development in medical students.

To complement the JSE-S and provide qualitative insight into students' experiences, participants also responded to structured

reflection questions. These served as our qualitative data collection tool, analogous to an interview guide. The lecture group questions asked students to describe the video, to describe how the video made them feel, and to explain how they would apply the information in future practice. The VR group questions asked students to describe the video, to describe how the video made them feel, and to explain how they thought the patient felt during the interaction. While these reflections are self-reported and may be influenced by social desirability or awareness of study aims, they provide rich, contextual data that complement the quantitative JSE-S measures.

Statistical Analysis

We conducted qualitative analysis based on the participants' responses to the reflection prompts using a thematic analysis approach as outlined by Braun and Clarke [27-29]. Written responses were imported into Microsoft Excel for data management. Given the dataset in this study was relatively small (n=50 participants, 100 reflections), Microsoft Excel was sufficient for systematically organizing codes and themes, while analytic rigor was ensured through team-based coding and consensus review. Initial codes were created by reading the text provided by the participants and assigning a code identifier. Coding was an iterative process in which each instance of a code was compared to previous instances to confirm or modify the code and its definition. Once all the data were coded, emergent themes were abstracted by grouping similar codes together. Codes and themes were each reviewed by members of the research team (AM, VH, ABK, and SST). Any discrepancies were resolved through discussion until consensus was reached. This process represents a "modified" application of thematic analysis, and codes were later used to derive quantitative empathy scores as described below. Major themes and illustrative quotes from the participants' responses are reported in the results section to provide further context and to support our findings.

We then determined quantitative scores based on the level of empathy present in the participants' responses. We assigned empathy scores to each participant based on their responses to video 1 (V1) and responses to video 2 (V2), and we also assigned an overall empathy score based on their responses to both videos 1 and 2 (V3). Empathy was scored on a 3-point scale (1=no signs of empathy, 2=some signs of empathy, 3=high signs of empathy). We used our codes to assign these empathy scores. High signs of empathy were indicated by codes such as "patient focused" and "emotive words," while low signs of empathy were indicated by codes like "surface level" and "misperception." We also used a 3-point scale to assign each participant a score for level of change in assigned empathy scores from video 1 to video 2 (1=negative change, 2=no change, and 3=positive change).

The data were quantitatively analyzed by LAF using IBM SPSS Statistics (version 29.0). Descriptive statistics (counts and frequencies for categorical data; means and SD values for continuous data) were performed. We used Mann-Whitney tests to compare the assigned empathy scores at V1, V2, and V3 between the lecture and VR groups. JSE-S scores pre- and posttest were analyzed using paired-sample 2-tailed *t* tests to determine whether JSE-S scores changed over time for all participants. A factorial analysis of variance assessed whether time (pre or posttest) and group (lecture or VR) affected JSE-S scores. To confirm a relationship between the JSE-S and our assigned empathy scores, we ran a Spearman correlation between posttest JSE-S score and V3 for both lecture and VR groups. We considered *P*<.05 as statistically significant.

Ethical Considerations

From January to February of 2020, 50 first- and second-year medical students at the University of South Carolina School of Medicine Greenville participated in our study. The study was reviewed and approved by the University of South Carolina Office of Research Compliance (Institutional Review Board approval number: Pro00089391). Participation was voluntary, and students received a \$30 gift card as compensation. All

participants provided written informed consent prior to participation. Upon arrival, study personnel explained the study procedures and reviewed the informed consent document with each participant, answering all questions before consent was obtained. To minimize response bias, recruitment materials and the consent process described the study as an investigation of the role of virtual reality in medical education, which was accurate but did not fully disclose that empathy was a primary outcome of interest. Participants were verbally debriefed immediately following completion of their participation and informed of the study's full purpose. The study posed minimal risk to participants. Potential risks included mild skin irritation from the polyurethane E4 wristband and nausea or headache associated with use of the virtual reality headset. Participants were informed they could discontinue use of either device or withdraw from the study at any time without penalty. Identifying information was collected solely for recruitment purposes and was not linked to study data. Recruitment information was stored on a password-protected Microsoft Teams site accessible only to study personnel. All study data were deidentified, assigned a unique participant identification number, and stored in a HIPAA-compliant REDCap database on a secure network accessible only to authorized members of the research team.

Results

Participant Characteristics

Of the 50 preclinical medical students that participated in our study, the overall mean age was 24.14 (SD 2.66) years. The majority of participants were female (37/50, 74%) and non-Hispanic White (31/50, 62%). There were similar numbers of participants in each class year, with 24 of 50 (48%) M1 students and 26 of 50 (52%) M2 students. Eight of 50 (16%) participants identified as first-generation students, and 14 of 50 (28%) participants were from rural backgrounds. The specialty of interest varied among the participants, but the highest number of students (11/50, 22%) chose obstetrics and gynecology. Additional demographic information is depicted in [Table 1](#).

Table . Participant demographic data.

Variables	Lecture, (n=25)	VR ^a , (n=25)	Total, (N=50)
Age (years), mean (SD)	23.92 (1.53)	24.36 (3.46)	24.14 (2.66)
Sex, n (%)			
Male	8 (32)	4 (16)	12 (24)
Female	16 (64)	21 (84)	37 (74)
Transgender	1 (4)	0 (0)	1 (2)
Race and ethnicity ^b , n (%)			
Asian	2 (8)	7 (28)	9 (18)
Black or African American	4 (16)	5 (20)	9 (18)
Hispanic or Latino	1 (4)	0 (0)	1 (2)
Non-Hispanic White	18 (72)	13 (52)	31 (62)
Another or not listed	2 (7.7)	0 (0)	2 (4)
Decline to answer	0 (0)	1 (4)	1 (2)
Class year, n (%)			
M1	13 (52)	11 (44)	24 (48)
M2	12 (48)	14 (56)	26 (52)
Marital status, n (%)			
Single or never married	23 (92)	21 (84)	44 (88)
Married	2 (8)	4 (16)	6 (12)
First generation ^c , n (%)	4 (16)	4 (16)	8 (16)
Rural, n (%)	7 (28)	7 (28)	14 (28)
Specialty, n (%)			
Dermatology	0 (0)	2 (8)	2 (4)
Emergency medicine	2 (8)	1 (4)	3 (6)
Family medicine	1 (4)	2 (8)	3 (6)
General surgery	6 (24)	2 (8)	8 (16)
Internal medicine	2 (8)	1 (4)	3 (6)
Internal medicine-pediatrics	1 (4)	0 (0)	1 (2)
Interventional radiology	1 (4)	0 (0)	1 (2)
Obstetrics and Gynecology	5 (20)	6 (24)	11 (22)
Ophthalmology	1 (4)	0 (0)	1 (2)
Orthopedic surgery	2 (8)	2 (8)	4 (8)
Otolaryngology	1 (4)	1 (4)	2 (4)
Pediatrics	1 (4)	5 (20)	6 (12)
Undecided	2 (8)	3 (12)	5 (10)

^aVR: virtual reality.^bResponse options allowed respondents to check more than 1 option.^cIn our study, defined as a student whose parents or guardians did not complete a four-year degree.

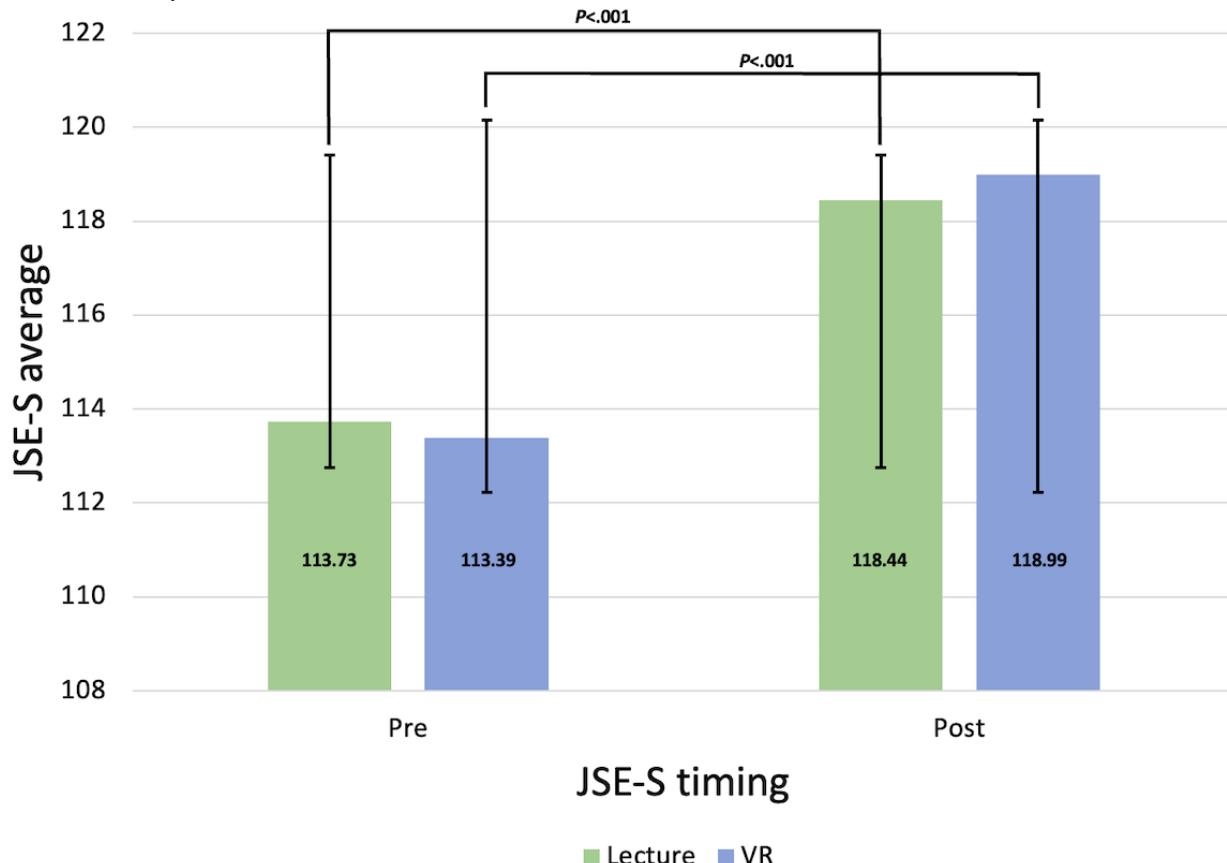
Quantitative Analysis: JSE-S

JSE-S scores were assessed prior to and following each testing session (pre or posttest) for both the lecture and VR groups. Prevideo JSE-S scores were not significantly different between the groups, with the lecture group averaging 113.73 (SD 10.57)

and the VR group averaging 113.39 (SD 10.24). Postvideo averages of the JSE scores between the 2 groups were again not significantly different, showing that the lecture group averaged 118.44 (SD 11.01) and the VR group averaged 118.99 (SD 10.02). A paired-samples *t* test was used to determine whether JSE-S scores changed across time for all participants. Results

indicate that JSE scores significantly increased after the empathy training sessions compared to before the sessions ($t_{49}=7.28$; $P<.001$). A factorial analysis of variance assessed whether time (pre or post) and group (lecture or VR) affected JSE-S scores, with results indicating no significant effect of either ($F_{1,48}=0.17$; $P=.68$).

Figure 2. Average scores across time (pre or post) by group (lecture or VR). The Jefferson Scale of Empathy-Student Version was administered to participants prior to viewing both videos (pre) and after viewing both videos (post). Jefferson Scale of Empathy-Student Version score was compared across time in lecture versus VR groups using a paired-sample t test ($***P<.001$). SD bars are shown. JSE-S: Jefferson Scale of Empathy-Student Version; VR: virtual reality.

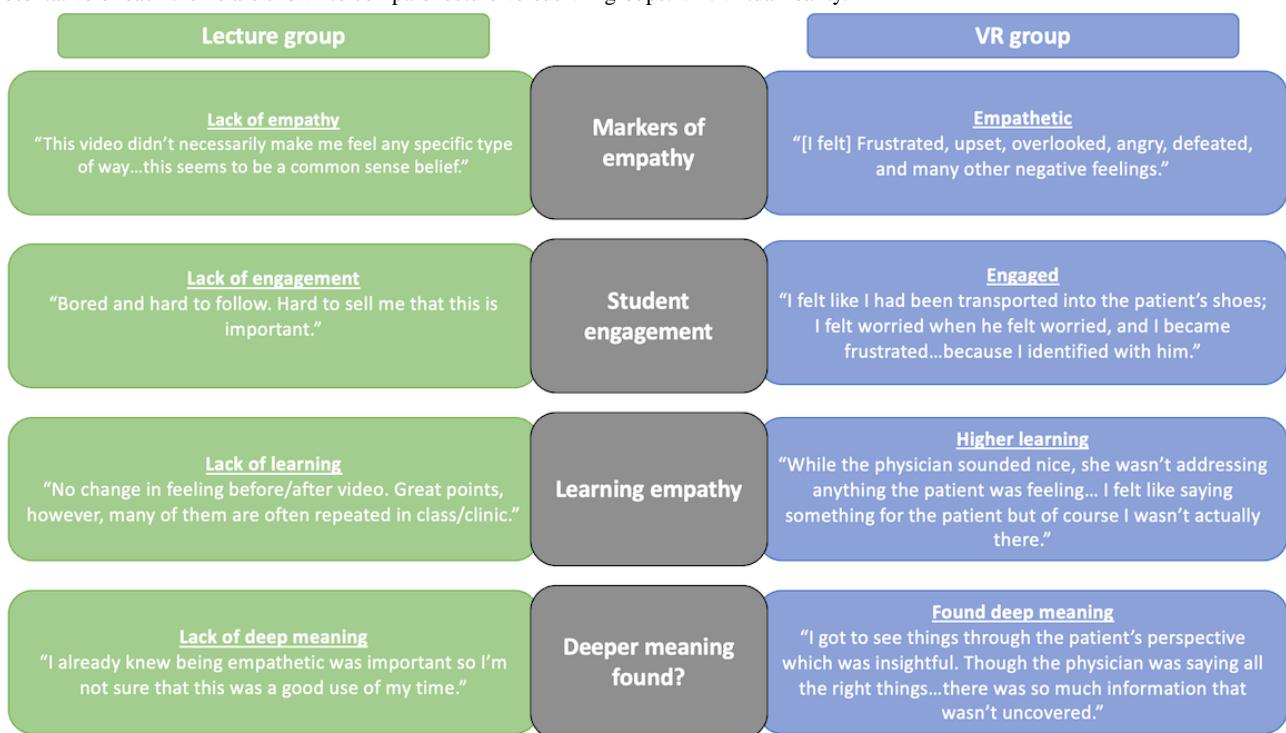


Qualitative Analysis of Reflection Responses

Qualitative analysis of participants' reflection responses in the lecture group versus the VR group revealed 4 major themes relating to empathy and its learning that diametrically opposed each other: (1) markers of empathy, (2) student engagement, (3) learning empathy, and (4) deeper meaning found (Figure 3). The lecture group responses overall demonstrated lack of empathy, lack of engagement, lack of learning, and lack of deeper meaning. While lecture group students seemed to cognitively acknowledge the importance of empathy, their responses were generally sarcastic and demonstrated low levels of empathy. For example, when asked how the lecture video made them feel, 1 student responded simply with, "bored." Another stated, "this video made me feel sleepy...this is all stuff we have heard before." When asked how they would apply the lecture content to future clinical practice, 1 student responded with "be empathetic like I was supposed to be before?" In contrast, the VR group responses showed empathy, engagement, higher learning, and deeper meaning. VR group

students experienced a range of empathetic emotions after the VR experience, with 1 student stating that they felt "frustrated, upset, overlooked, angry, defeated, and many other negative feelings" on behalf of the patient. The VR group participants were highly engaged, commenting, "I felt like stepping in and asking the patient for input," and "I felt like I wanted to be Sam's advocate and ask questions...the provider should've asked." Because the VR videos promoted perspective-taking, the students found deeper meaning. One individual stated, "I think the patient...needed to talk about his feelings just as much, if not more than his abdominal pain." The VR group also demonstrated learned empathy in their responses. After watching the first video, 1 participant commented, "I felt that the physician was doing a good job being pleasant." After the second video, the same participant stated, "I felt like I had been transported into the patient's shoes; I felt worried when he felt worried, and I became frustrated when the doctor didn't give him any feasible solutions because I identified with him at that point."

Figure 3. Major themes and representative quotations. Qualitative coding of participant reflection responses revealed 4 major themes. Example quotations representative of each theme are shown to compare lecture versus VR groups. VR: virtual reality.

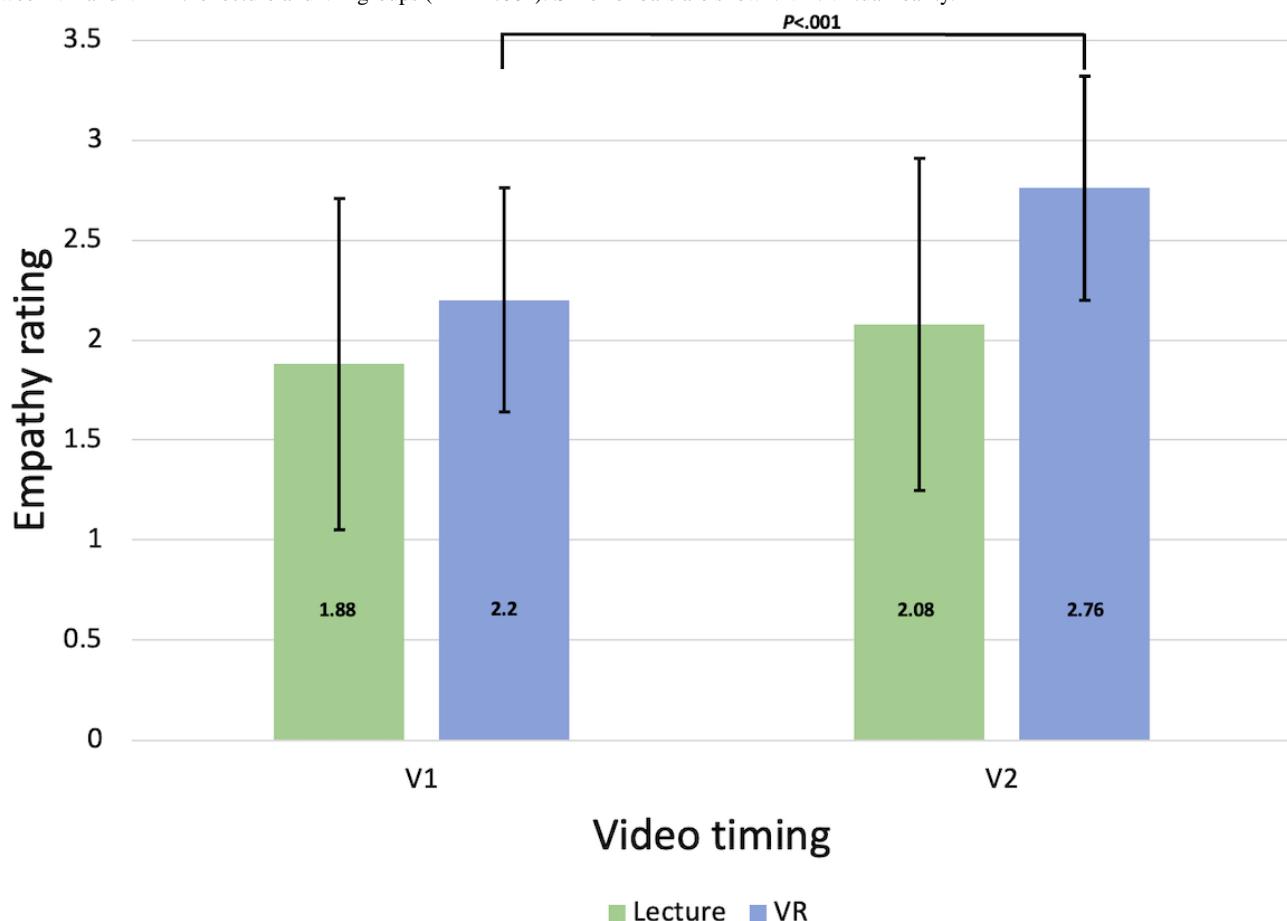


Quantitative Scoring of Qualitative Coding and Comparison to JSE-S Scores

A Mann-Whitney analysis was used to compare the assigned empathy scores at V1 (after video 1) and V2 (after video 2) between the lecture and VR groups. Results indicate that there was no significant difference in qualitative coding for empathy scores between the 2 groups after the first video (V1). However, after the second video (V2), there was a significant difference between qualitative scores, with those in the VR group having significantly higher empathy coding scores than the lecture group ($U=165.5$; $P<.001$). A Mann-Whitney analysis also

showed that overall qualitative empathy coding for the participants (V3) was significantly higher for those in the VR group as compared to those in the lecture group ($U=202.5$; $P=.03$), based on empathy scores of 1.96 and 2.52 in the lecture and VR groups, respectively.. The overall qualitative coding for empathy (V3) was assessed with the post-JSE-S scores to determine if the 2 scores were related to each other (a quantitative rating of empathy in the JSE-S compared to the quantitative scoring of the qualitative coding). Spearman correlation between posttest JSE-S and V3 for both the lecture and VR groups was significantly, positively correlated ($\rho_{49}=0.29$; $P=.03$). Refer to [Figure 4](#) for more information.

Figure 4. Qualitative scoring of comments after video 1 (V1) and video 2 (V2). V1 is defined as participants' assigned empathy score after viewing Video 1. V2 is defined as participants' assigned empathy score after viewing Video 2. A Mann-Whitney analysis compared the assigned empathy scores between V1 and V2 in the lecture and VR groups ($***P<.001$). SD error bars are shown. VR: virtual reality.



Discussion

Principal Findings

This study presents early findings that the VR experience uniquely engaged students and elicited more empathetic responses compared to the traditional lecture. In our mixed methods evaluation, the intervention or VR group demonstrated qualitatively higher markers of empathy, more student engagement, found deeper meaning, and demonstrated more learning of empathy than the lecture group. The intervention or VR group also showed significantly higher overall assigned empathy scores than the lecture group. These findings indicate that immersive VR experiences may enhance empathy development beyond conventional lecture-based methods.

Quantitative and Qualitative Findings

The JSE-S is widely used and has been validated by numerous empirical studies as an instrument to measure empathy [22,30]. Our quantitative analysis conducted using this tool demonstrated a significant increase in JSE-S scores after both the VR and traditional lecture training sessions compared to before the sessions. Seeing an increase in JSE-S scores in both groups posttest is not a surprising result considering students were primed on the topic of physician empathy while completing the questionnaire and since traditional teaching methods have already been proven effective in teaching empathy to some degree [6,31,32]. However, the VR intervention's efficacy in

fostering empathy demonstrates noninferiority to traditional lecture-based learning and underscores the importance of experiential learning, as outlined in adult learning theory, and provides a new dimension to empathy theory in medical education. Adult learning theory emphasizes the significance of self-directed, experiential learning processes, where adults learn best when engaged in activities that relate closely to real-life scenarios or challenges they might face [33]. The immersive nature of VR, by providing medical students with a first-person perspective of patient experiences, leverages these principles effectively, offering a profound, emotive learning experience that traditional didactic methods may lack. This approach aligns with the constructs of empathy theory, which posits that empathy involves not just cognitive understanding but also affective sharing and emotional resonance with others [34].

Qualitative analysis helped us better understand the experiential differences between the groups. Themes from the qualitative responses relate clearly to the literature on empathy and adult learning. Individuals in the VR group described experiencing a variety of emotions while watching the VR video, while the lecture group described their experience using more cognitive language. This suggests that the VR intervention met our intent to promote perspective taking. Furthermore, adult learning theorists emphasize the role emotions can play in helping learners connect a construct with prior knowledge and past experiences [35]. Beyond enriching their learning through

perspective taking, the VR intervention appeared to uniquely engage students more than the lecture. The theme of deeper meaning found in the VR group's reflections aligns with the concept of "presence" in VR environments, the true feeling of "being there" in the virtual space. This sense of presence is critical for generating impactful, empathetic responses, as it bridges the gap between knowing about another's experience and feeling with them. Literature on VR and empathy underscores the role of presence in enhancing the emotional impact of VR experiences, suggesting that it is this immersion that enables learners to connect with content on a deeper emotional level, fostering a genuine understanding and appreciation for the patient's perspective [36].

By connecting the quantitative increases in JSE-S scores with the rich qualitative reflections, a clear pattern emerges: immersive VR experiences not only improve self-reported empathy but also foster deeper emotional engagement, perspective-taking, and understanding of patient experiences. This combined evidence highlights the main contribution of the study, showing that VR interventions can meaningfully complement and enhance traditional teaching methods.

Implications for Medical Education

For this topic in particular, embodied and applied experience seems to provide much more nuance to the concept of empathy versus the cognitive acknowledgment that empathy is important in practice. Through the comments, the intervention appeared to help many students more fully understand how empathy can promote better health care. Furthermore, this study's findings resonate with previous research indicating the potential of VR to overcome some of the limitations associated with traditional empathy training methods. For instance, studies have demonstrated VR's effectiveness in reducing cognitive load and allowing for a more standardized and immersive learning experience, characteristics that are particularly advantageous in the context of medical training where the cognitive demand is high and the need for consistent, reproducible training experiences is paramount [11].

Limitations and Future Directions

Our review consisted of a single-institution study at 1 academic center, meaning more work is needed to make more widespread application of our results possible. We note that our participants were predominantly female (37/50, 74%). This reflects the voluntary nature of our participant recruitment, as a greater number of female students elected to enroll compared to their male counterparts. Prior literature has documented sex and gender differences in empathy [37]; thus, we acknowledge the overrepresentation of female participants in our study as a factor that may have influenced the observed results. Furthermore, we did not follow up with participants after the end of the study to assess empathy during clinical rotations and in practice. We feel this would have provided a better understanding of the long-term impact of our empathy training intervention on participants' empathy throughout their medical training. We also recognize some limitations in our use of the JSE-S, including the fact that we may have applied the scale too early following our intervention, thus potentially limiting its reliability. Compared to other interventions to promote empathy

in health care learners, this intervention was quite brief; a systematic review of such interventions found that the average length was 10 hours [38].

While traditional teaching methods have been shown to help increase empathy in medical students, the profound differences noted in our qualitative data, however, lead us to question whether full immersion through a VR experience may be a helpful additional tool in allowing students to develop a deeper understanding. Future directions warrant testing a combined approach of traditional lecture alongside a VR intervention to see if a combination provides an even greater increase in empathy than either one alone. We would also like to assess the long-term impact of these interventions on students as they progress through their clinical years.

Although the JSE-S is widely used and validated, it primarily assesses attitudes toward empathy rather than direct empathetic behavior, and participants may respond in a socially desirable manner. No filler items were included to mask the study's purpose, but participants were not explicitly informed of the study hypotheses to minimize expectancy effects. We acknowledge the Hawthorne effect, in which participants modify behavior because they know they are being observed, could have influenced responses. However, the randomized controlled design, balanced group assignments, and inclusion of both qualitative and quantitative data help mitigate potential bias and strengthen the validity of the findings. Future studies might incorporate validated infrequency or social desirability scales, such as the Infrequency Scale from the Zuckerman-Kuhlman Personality Questionnaire to further account for response bias.

Conclusions

Our study underscores the transformative potential of VR interventions in enhancing empathy among medical students. By allowing learners to experience the patient's perspective firsthand, VR offers an emotionally engaging, experiential form of learning that complements traditional didactic methods. Our findings, demonstrating higher JSE-S scores and richer qualitative reflections in the VR group, indicate that immersive learning environments can promote deeper perspective taking and maintain empathic growth. These findings illuminate the importance of adopting innovative, technology-enhanced educational strategies to cultivate a more empathetic health care workforce. They also suggest that technology-enhanced approaches may help counter the well-documented decline in empathy during medical training. Integrating VR into medical curricula could provide a scalable, standardized way to strengthen the humanistic foundations of physician education while aligning with contemporary adult learning principles. Future research should examine the long-term effects of such interventions and explore combinations of VR with reflective or clinical teaching to reinforce empathetic skills in practice. Our work lays a foundational step toward reimaging empathy training in medical education by leveraging immersive technology to cultivate emotional understanding and advance a more holistic model of physician training that integrates scientific expertise with the empathy essential to compassionate, patient-centered care.

Acknowledgments

Special thanks to our student research assistants: Kurestin Miller, Dan Strat, Meghan Dubose, Sarah Barto, and Anne Kirby; our actors Jenn Maness and Chris Freeland; and the Patients from the University of South Carolina's Patient Engagement Studio who reviewed the project and provided feedback for improvement on the patient/provider scenario for the video. Additionally, a game of Rochambeau (rock, paper, and scissors) was used to determine senior and corresponding author positions.

Funding

Partial financial support was received from a Prisma Health-Upstate seed grant, The University of South Carolina Center for Teaching Excellence teaching innovation grant, and the University of South Carolina School of Medicine SOARing initiative.

Conflicts of Interest

None declared.

References

1. Singer T, Lamm C. The social neuroscience of empathy. *Ann N Y Acad Sci* 2009 Mar;1156(1):81-96. [doi: [10.1111/j.1749-6632.2009.04418.x](https://doi.org/10.1111/j.1749-6632.2009.04418.x)] [Medline: [19338504](https://pubmed.ncbi.nlm.nih.gov/19338504/)]
2. Kim SS, Kaplowitz S, Johnston MV. The effects of physician empathy on patient satisfaction and compliance. *Eval Health Prof* 2004 Sep;27(3):237-251. [doi: [10.1177/0163278704267037](https://doi.org/10.1177/0163278704267037)] [Medline: [15312283](https://pubmed.ncbi.nlm.nih.gov/15312283/)]
3. Hojat M, Louis DZ, Markham FW, Wender R, Rabinowitz C, Gonnella JS. Physicians' empathy and clinical outcomes for diabetic patients. *Acad Med* 2011 Mar;86(3):359-364. [doi: [10.1097/ACM.0b013e3182086fe1](https://doi.org/10.1097/ACM.0b013e3182086fe1)] [Medline: [21248604](https://pubmed.ncbi.nlm.nih.gov/21248604/)]
4. Hojat M, Gonnella JS, Nasca TJ, Mangione S, Vergare M, Magee M. Physician empathy: definition, components, measurement, and relationship to gender and specialty. *Am J Psychiatry* 2002 Sep;159(9):1563-1569. [doi: [10.1176/appi.ajp.159.9.1563](https://doi.org/10.1176/appi.ajp.159.9.1563)] [Medline: [12202278](https://pubmed.ncbi.nlm.nih.gov/12202278/)]
5. Hojat M, Mangione S, Nasca TJ, et al. An empirical study of decline in empathy in medical school. *Med Educ* 2004 Sep;38(9):934-941. [doi: [10.1111/j.1365-2929.2004.01911.x](https://doi.org/10.1111/j.1365-2929.2004.01911.x)] [Medline: [15327674](https://pubmed.ncbi.nlm.nih.gov/15327674/)]
6. Batt-Rawden SA, Chisolm MS, Anton B, Flickinger TE. Teaching empathy to medical students: an updated, systematic review. *Acad Med* 2013 Aug;88(8):1171-1177. [doi: [10.1097/ACM.0b013e318299f3e3](https://doi.org/10.1097/ACM.0b013e318299f3e3)] [Medline: [23807099](https://pubmed.ncbi.nlm.nih.gov/23807099/)]
7. Wündrich M, Schwartz C, Feige B, Lemper D, Nissen C, Voderholzer U. Empathy training in medical students - a randomized controlled trial. *Med Teach* 2017 Oct;39(10):1096-1098. [doi: [10.1080/0142159X.2017.1355451](https://doi.org/10.1080/0142159X.2017.1355451)] [Medline: [28749198](https://pubmed.ncbi.nlm.nih.gov/28749198/)]
8. Batson CD, Eklund JH, Chermok VL, Hoyt JL, Ortiz BG. An additional antecedent of empathic concern: valuing the welfare of the person in need. *J Pers Soc Psychol* 2007 Jul;93(1):65-74. [doi: [10.1037/0022-3514.93.1.65](https://doi.org/10.1037/0022-3514.93.1.65)] [Medline: [17605589](https://pubmed.ncbi.nlm.nih.gov/17605589/)]
9. Vescio TK, Sechrist GB, Paolucci MP. Perspective taking and prejudice reduction: the mediational role of empathy arousal and situational attributions. *Euro J Social Psych* 2003 Jul;33(4):455-472. [doi: [10.1002/ejsp.163](https://doi.org/10.1002/ejsp.163)]
10. Galinsky AD, Ku G. The effects of perspective-taking on prejudice: the moderating role of self-evaluation. *Pers Soc Psychol Bull* 2004 May;30(5):594-604. [doi: [10.1177/0146167203262802](https://doi.org/10.1177/0146167203262802)] [Medline: [15107159](https://pubmed.ncbi.nlm.nih.gov/15107159/)]
11. 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](https://doi.org/10.1371/journal.pone.0204494)] [Medline: [30332407](https://pubmed.ncbi.nlm.nih.gov/30332407/)]
12. Ahn SJG, Bostick J, Ogle E, Nowak KL, McGillicuddy KT, Bailenson JN. Experiencing nature: embodying animals in immersive virtual environments increases inclusion of nature in self and involvement with nature. *J Comput-Mediat Comm* 2016 Nov;21(6):399-419. [doi: [10.1111/jcc4.12173](https://doi.org/10.1111/jcc4.12173)]
13. Zaki J. Empathy: a motivated account. *Psychol Bull* 2014 Nov;140(6):1608-1647. [doi: [10.1037/a0037679](https://doi.org/10.1037/a0037679)] [Medline: [25347133](https://pubmed.ncbi.nlm.nih.gov/25347133/)]
14. Macrae CN, Bodenhausen GV. Social cognition: thinking categorically about others. *Annu Rev Psychol* 2000;51:93-120. [doi: [10.1146/annurev.psych.51.1.93](https://doi.org/10.1146/annurev.psych.51.1.93)] [Medline: [10751966](https://pubmed.ncbi.nlm.nih.gov/10751966/)]
15. Oh SY, Bailenson J, Weisz E, Zaki J. Virtually old: embodied perspective taking and the reduction of ageism under threat. *Comput Human Behav* 2016 Jul;60:398-410. [doi: [10.1016/j.chb.2016.02.007](https://doi.org/10.1016/j.chb.2016.02.007)]
16. Sirkunen E, Väätäjä H, Uskali T, Rezaei PP. Journalism in virtual reality. Presented at: AcademicMindtrek'16; Oct 17-18, 2016; Tampere, Finland p. 297-303. [doi: [10.1145/2994310.2994353](https://doi.org/10.1145/2994310.2994353)]
17. Benbassat J, Baumal R. What is empathy, and how can it be promoted during clinical clerkships? *Acad Med* 2004 Sep;79(9):832-839. [doi: [10.1097/00001888-200409000-00004](https://doi.org/10.1097/00001888-200409000-00004)] [Medline: [15326005](https://pubmed.ncbi.nlm.nih.gov/15326005/)]
18. Gallop R, Lancee WJ, Garfinkel PE. The empathic process and its mediators. a heuristic model. *J Nerv Ment Dis* 1990 Oct;178(10):649-654. [doi: [10.1097/00005053-199010000-00006](https://doi.org/10.1097/00005053-199010000-00006)] [Medline: [2230750](https://pubmed.ncbi.nlm.nih.gov/2230750/)]
19. Coll MP, Viding E, Rütgen M, et al. Are we really measuring empathy? Proposal for a new measurement framework. *Neurosci Biobehav Rev* 2017 Dec;83:132-139. [doi: [10.1016/j.neubiorev.2017.10.009](https://doi.org/10.1016/j.neubiorev.2017.10.009)] [Medline: [29032087](https://pubmed.ncbi.nlm.nih.gov/29032087/)]

20. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research, 3rd edition: Sage Publications; 2017. URL: <https://collegepublishing.sagepub.com/products/designing-and-conducting-mixed-methods-research-3-241842> [accessed 2025-12-30]
21. Deuter CE, Nowacki J, Wingenfeld K, et al. The role of physiological arousal for self-reported emotional empathy. *Auton Neurosci* 2018 Nov;214:9-14. [doi: [10.1016/j.autneu.2018.07.002](https://doi.org/10.1016/j.autneu.2018.07.002)] [Medline: [30104144](https://pubmed.ncbi.nlm.nih.gov/30104144/)]
22. Hojat M, DeSantis J, Shannon SC, et al. The Jefferson Scale of Empathy: a nationwide study of measurement properties, underlying components, latent variable structure, and national norms in medical students. *Adv Health Sci Educ Theory Pract* 2018 Dec;23(5):899-920. [doi: [10.1007/s10459-018-9839-9](https://doi.org/10.1007/s10459-018-9839-9)] [Medline: [29968006](https://pubmed.ncbi.nlm.nih.gov/29968006/)]
23. Hojat M, Gonnella JS, Nasca TJ, Mangione S, Veloksi JJ, Magee M. The Jefferson Scale of Physician Empathy: further psychometric data and differences by gender and specialty at item level. *Acad Med* 2002 Oct;77(10 Suppl):S58-S60. [doi: [10.1097/00001888-200210001-00019](https://doi.org/10.1097/00001888-200210001-00019)] [Medline: [12377706](https://pubmed.ncbi.nlm.nih.gov/12377706/)]
24. Yu J, Ganesh S, Lancastle D. Psychometric evaluation of the 'Jefferson Scale of Empathy' in a sample of nursing students in the United Kingdom. *Nurs Open* 2024 Jan;11(1):e2054. [doi: [10.1002/nop2.2054](https://doi.org/10.1002/nop2.2054)]
25. Piumatti G, Abbiati M, Baroffio A, Gerbase MW. Empathy trajectories throughout medical school: relationships with personality and motives for studying medicine. *Adv Health Sci Educ Theory Pract* 2020 Dec;25(5):1227-1242. [doi: [10.1007/s10459-020-09965-y](https://doi.org/10.1007/s10459-020-09965-y)] [Medline: [32095990](https://pubmed.ncbi.nlm.nih.gov/32095990/)]
26. Hojat M, Shannon SC, DeSantis J, Speicher MR, Bragan L, Calabrese LH. Empathy in medicine National norms for the Jefferson Scale of Empathy: a nationwide project in osteopathic medical education and empathy (POMEE). *J Am Osteopath Assoc* 2019 Aug 1;119(8):520-532. [doi: [10.7556/jaoa.2019.091](https://doi.org/10.7556/jaoa.2019.091)] [Medline: [31355891](https://pubmed.ncbi.nlm.nih.gov/31355891/)]
27. Braun V, Clarke V, Terry G, Hayfield N. Thematic analysis. In: Liamputpong P, editor. *Handbook of Research Methods in Health and Social Sciences*: Springer; 2019:843-860. [doi: [10.1007/978-981-10-5251-4_103](https://doi.org/10.1007/978-981-10-5251-4_103)]
28. Braun V, Clarke V. *Successful Qualitative Research: A Practical Guide for Beginners*: SAGE Publications; 2013. URL: <https://uk.sagepub.com/en-gb/eur/successful-qualitative-research/book233059> [accessed 2025-12-17]
29. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006 Jan;3(2):77-101. [doi: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)]
30. Tavakol S, Dennick R, Tavakol M. Psychometric properties and confirmatory factor analysis of the Jefferson Scale of Physician Empathy. *BMC Med Educ* 2011 Aug 2;11:54. [doi: [10.1186/1472-6920-11-54](https://doi.org/10.1186/1472-6920-11-54)] [Medline: [21810268](https://pubmed.ncbi.nlm.nih.gov/21810268/)]
31. Doyen S, Klein O, Simons DJ, Cleeremans A. On the other side of the mirror: priming in cognitive and social psychology. *Soc Cogn* 2014 Jun;32:12-32. [doi: [10.1521/soco.2014.32.supp.12](https://doi.org/10.1521/soco.2014.32.supp.12)]
32. Chartrand TL, Bargh JA. Automatic activation of impression formation and memorization goals: nonconscious goal priming reproduces effects of explicit task instructions. *J Pers Soc Psychol* 1996;71(3):464-478. [doi: [10.1037/0022-3514.71.3.464](https://doi.org/10.1037/0022-3514.71.3.464)]
33. Knowles MS, Holton EF, Swanson RA. *The Adult Learner*, 8th edition: Routledge; 2015. [doi: [10.4324/9781315816951](https://doi.org/10.4324/9781315816951)]
34. Decety J, Jackson PL. The functional architecture of human empathy. *Behav Cogn Neurosci Rev* 2004 Jun;3(2):71-100. [doi: [10.1177/1534582304267187](https://doi.org/10.1177/1534582304267187)] [Medline: [15537986](https://pubmed.ncbi.nlm.nih.gov/15537986/)]
35. Merriam SB. Adult learning theory for the twenty - first century. *New Dir Adult Contin Educ* 2008 Sep;2008(119):93-98. [doi: [10.1002/ace.309](https://doi.org/10.1002/ace.309)]
36. Riva G, Wiederhold BK, Mantovani F. Neuroscience of virtual reality: from virtual exposure to embodied medicine. *Cyberpsychol Behav Soc Netw* 2019 Jan;22(1):82-96. [doi: [10.1089/cyber.2017.29099.gri](https://doi.org/10.1089/cyber.2017.29099.gri)] [Medline: [30183347](https://pubmed.ncbi.nlm.nih.gov/30183347/)]
37. Christov-Moore L, Simpson EA, Coudé G, Grigaityte K, Iacoboni M, Ferrari PF. Empathy: gender effects in brain and behavior. *Neurosci Biobehav Rev* 2014 Oct;46 Pt 4(Pt 4):604-627. [doi: [10.1016/j.neubiorev.2014.09.001](https://doi.org/10.1016/j.neubiorev.2014.09.001)] [Medline: [25236781](https://pubmed.ncbi.nlm.nih.gov/25236781/)]
38. Winter R, Issa E, Roberts N, Norman RI, Howick J. Assessing the effect of empathy-enhancing interventions in health education and training: a systematic review of randomised controlled trials. *BMJ Open* 2020 Sep 25;10(9):e036471. [doi: [10.1136/bmjopen-2019-036471](https://doi.org/10.1136/bmjopen-2019-036471)] [Medline: [32978187](https://pubmed.ncbi.nlm.nih.gov/32978187/)]

Abbreviations

JSE-S: Jefferson Scale of Empathy for medical students

V1: lecture video 1

V2: lecture video 2

V3: lecture video 3

VR: virtual reality

Edited by J Eriksen; submitted 24.Apr.2025; peer-reviewed by R Aro, S Bidmon; accepted 30.Nov.2025; published 04.Feb.2026.

Please cite as:

Mundok AG, Ho VN, Fowler LA, Kennedy AB, Stark-Taylor S

Investigating the Impact of a Virtual Reality Experience on Medical Student Empathy: Mixed Methods Study

JMIR Med Educ 2026;12:e76504

URL: <https://mededu.jmir.org/2026/1/e76504>

doi:10.2196/76504

© Allen G Mundok, Vivian N Ho, Lauren A Fowler, Ann Blair Kennedy, Shannon Stark-Taylor. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 4.Feb.2026. 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.

Blended Learning Compared With Face-to-Face Learning Among Family Medicine Residents: Randomized Controlled Trial

Pierre-Yves Meunier^{1,2,3}, MD; Sophie Schlatter^{2,4}, PhD; Juliette Macabrey^{2,3,5,6}, MD; Frédéric Zorzi³, MD; Thomas Colleony^{3,7}, MD; Rémy Boussageon^{3,8}, MD, PhD; Hubert Maisonneuve^{3,9,10}, MD, MHPE; Marion Lamort-Bouché^{2,3}, MD, PhD

¹Maison de santé pluriprofessionnelle et universitaire Beauvisage, Lyon, France

¹⁰Mermoz Primary Health Center, Lyon, France

²Research on Healthcare Performance RESHAPE, Inserm U1290, Université Claude Bernard Lyon 1, Lyon, France

³Collège universitaire de médecine générale, Université Claude Bernard Lyon 1, 8 avenue Rockefeller, Lyon, France

⁴Healthcare Simulation Center, SIMULYON, Hospices Civils de Lyon, Université Claude Bernard Lyon 1, Lyon, France

⁵Maison de santé pluriprofessionnelle et universitaire Bel Air, Saint Priest, France

⁶Faculty of Education, University of Ottawa, Ottawa, ON, Canada

⁷French Military Medical School, Bron, France

⁸Laboratoire de biométrie et biologie évolutive UMR CNRS 5558, Université Claude Bernard Lyon 1, Université de Lyon, Lyon, France

⁹University Institute for Primary Care (IuMFE), Faculty of Medicine, University of Geneva, Geneva, Switzerland

Corresponding Author:

Pierre-Yves Meunier, MD

Maison de santé pluriprofessionnelle et universitaire Beauvisage, Lyon, France

Abstract

Background: The medical education of French family medicine residents involves active, socioconstructivist-inspired small-group courses useful for skill acquisition. This is challenged by the increasing gap between the growing number of residents and the limited number of teachers. Blended courses have the potential to address this issue by reducing the duration of face-to-face sessions while preserving small-group courses.

Objective: This study aimed to compare the effects of blended vs traditional, face-to-face, active, socioconstructivist learning on the acquisition of knowledge and skills by family medicine residents.

Methods: We conducted a randomized controlled trial to compare a blended course and a traditional course. The blended course involved 2.5 hours of asynchronous e-learning and a 3-hour face-to-face session. The traditional course involved 5.5 hours of face-to-face teaching. Both courses were grounded in socioconstructivist principles and actively engaged residents. The primary outcome was residents' self-assessment of knowledge and skills. Secondary outcomes included satisfaction with knowledge- or skill-related learning objectives and academic achievement at 6 months.

Results: We included 155 family medicine residents (n=78, 50.3% in the blended course and n=77, 49.7% in the traditional course). There was no significant difference between groups regarding the primary outcome (mean difference 0.40 [maximum mean difference 20] points, 95% CI -0.21 to 1.02; $P=.19$; Cohen $d=0.21$). No significant differences were observed for the secondary outcomes except for knowledge self-assessment, which was higher in the blended course but not educationally meaningful (mean difference 0.40 [maximum possible 10] points, 95% CI 0.07-0.71; $P=.02$; Cohen $d=0.39$).

Conclusions: Blended courses can help sustain socioconstructivist small-group teaching methods while accommodating a growing family medicine resident population, with no deleterious impact on knowledge and skill self-assessments.

Trial Registration: ClinicalTrials.gov NCT06409273; <https://clinicaltrials.gov/ct2/show/NCT06409273>

(*JMIR Med Educ* 2026;12:e86387) doi:10.2196/86387

KEYWORDS

primary health care; family practice; education; medical; graduate; internship and residency; distance; computer-assisted instruction; educational technology; self-assessment; educational measurement; randomized controlled trial

Introduction

The World Health Organization emphasizes the critical role of efficient primary care systems in ensuring population health [1]. However, many countries face a shortage of family physicians, which limits access to primary care services [2-4]. To address this issue, some countries, including France, have recently increased the number of medical students [5-7]. In France, family medicine residents are medical graduates enrolled in a 4-year family medicine residency program [8]. Their curriculum is competency-based, guided by a national framework [9]. The constructive alignment of competency-based educational objectives, learning activities, and assessment implies, among other things, the provision of active, socioconstructivist-inspired courses [10,11]. Social constructivism emphasizes the role of social interactions and collaboration for effective learning [12]. As a result, residents engage in a variety of collaborative learning activities with 3 to 25 participants per group. Reports of these activities, individual narrative medicine work, and assessments are collected in a portfolio. However, from the perspective of health profession educators, it is difficult to maintain feasibility with increasing class sizes [13]. This difficulty is particularly acute for family medicine residents because their growing number has outpaced the availability of teaching hours, threatening the quality of their education.

Face-to-face traditional learning relies on nondigital materials and human interaction [14]. Blended learning (BL) combines face-to-face education with asynchronous or synchronous e-learning [15]. BL is often used to shift from a teacher-centered, behaviorist instructional model to a student-centered, socioconstructivist instructional model [16]. In blended courses, face-to-face sessions reinforce knowledge and develop skills through the collective mobilization of knowledge acquired during e-learning [17]. Two meta-analyses found that BL had a more positive effect on knowledge acquisition compared to traditional education in health professions (including medical students) but with large heterogeneity in the results [18,19]. This heterogeneity can be explained by differences in student populations and missing details about the instructional approach. Moreover, these studies did not evaluate the effect of BL on skill acquisition. While BL appears to improve satisfaction among both students and teachers, no meta-analysis has specifically evaluated this outcome. Teachers have reported that BL has the potential to increase interactions with students [20] and allow educators to develop more content [21] and they felt greater job satisfaction with this modality [22]. Students have generally been satisfied [23].

While existing studies suggest that BL is generally effective, its effectiveness compared with an already active, socioconstructivist, face-to-face course remains unexplored. Moreover, its effect on skill acquisition among postgraduate medical trainees (residents) remains underexplored. In this study, we compared the effects of a blended course

(asynchronous active e-learning with shorter socioconstructivist face-to-face sessions) vs a fully face-to-face socioconstructivist course on the acquisition of knowledge and skills by family medicine residents. While implementing BL could mitigate the shortage of teachers by shortening face-to-face sessions, reduced in-person interaction may limit feedback and engagement, potentially leading to lower knowledge and skill acquisition.

Methods

Study Design

We conducted a single-blind randomized controlled trial with 2 parallel arms. The experimental arm received a blended course comprising an asynchronous e-learning module (2.5 hours; 45% of the course) followed by a face-to-face session (3 hours; 55% of the course). The control arm received the usual face-to-face traditional course (5.5 hours; 100% of the course). The trial was conducted between May 2024 and November 2024 at Université Claude Bernard Lyon 1 (France). It followed the CONSORT (Consolidated Standards of Reporting Trials) guidelines [24] and was prospectively registered on ClinicalTrials.gov (NCT06409273).

Sample and Setting

All first-year family medicine residents at Université Claude Bernard Lyon 1 were administratively enrolled, with no exclusion criteria. The course was delivered over a single day. Residents were assigned to 1 of 2 available dates. Participants were told that the study compared different teaching methods, with no further detail. They were block randomized using random sequences of block sizes of 4 and 8 in a 1:1 allocation ratio. Randomization was performed by the first author using random permuted blocks generated online [25]. On each teaching date, 4 groups were formed (2 per arm), with a maximum of 23 participants per group. Classrooms were configured similarly, with tables arranged in a U shape and spaces reserved for breakout work.

The 5 teachers were randomized via coin toss to 1 of the 2 study arms. Teachers who taught on both dates crossed over to the other arm on the second day. For each arm, we developed detailed, time-stamped scripts to ensure instructional fidelity and equivalence of content and duration.

Educational Intervention and Conceptual Framework

The course prepared residents to complete the thesis required for full qualification in France. This is a professionally oriented doctoral thesis publicly defended before an examination board. It is required for the award of the national Doctor of Medicine degree, which permits registration with the French Medical Council for independent practice. Distinct from a PhD, it is a concise, supervised scholarly work demonstrating the candidate's ability to design, conduct, and communicate research relevant to medical practice. We specified 5 process-oriented learning objectives: 3 for knowledge acquisition and 2 for skill development (Table 1).

Table . Learning objectives across study arms.

Learning objective	Study arm	
	Blended course	Traditional course
Knowledge oriented		
Identify the practical organization and regulatory requirements of the medical thesis	e-Learning	Face-to-face
Understand the different steps of a research project	e-Learning	Face-to-face
Summarize the medical thesis application form requirements	e-Learning	Face-to-face
Skill oriented		
Frame the research question for a thesis project	Face-to-face	Face-to-face
Assess the feasibility of a medical thesis project	Face-to-face	Face-to-face

The blended course sequenced active knowledge construction online with supervised skill development in person. We made the alignment between the 2 sessions explicit to residents to underscore the coherence between the learning objectives and the learning methods. The blended course arm began with an e-learning session grounded in self-regulated learning (SRL). SRL has been defined as self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goals [26]. Process-oriented learning objectives (Table 1) for knowledge acquisition were designed to activate the SRL forethought phase (goal setting and strategy selection) [27]. We aimed to help residents identify immediate next steps (securing a supervisor, conducting a literature review, and framing a research question), whereas strategy selection focused on learning how to choose an appropriate research method. Formative checkpoints and immediate feedback were embedded through interactive HTML5 Package activities—videos with decision points (“crossroads”), multiple-choice questions, drag-the-words items, and navigation hot spots. This design supported ongoing self-monitoring and adaptive control (comparison to standards and adjustment), thereby reducing discrepancies between current understanding and target performance. The accompanying explanatory text and curated external links to websites or documents enabled resource management and self-paced elaboration, further supporting SRL. The e-learning module was implemented online in the course management system Moodle (Moodle HQ; [Multimedia Appendix 1](#)).

The traditional course arm began with the collaborative construction of a knowledge base anchored in the knowledge-oriented learning objectives. Teachers acted as facilitators, providing key information and externalizing group thinking in mind maps that made success criteria and progress visible for monitoring.

Skill-oriented learning objectives were addressed identically in both arms during a face-to-face session. For the socioconstructivist activities, peer evaluation of thesis projects was organized in breakout groups, which discussed the research question and feasibility. These activities provided external feedback according to standards explicitly set by teachers,

closing the control loop and supporting the SRL reflection phase (self-evaluation and calibration). To stimulate motivational processes and goal setting, we highlighted the timeline leading up to the thesis [28].

Data Collection and Outcomes

Data were collected at the end of the teaching day via individualized links to an online form ([Multimedia Appendix 2](#)). Participants provided demographics (gender, age, research experience, and progress stage of the thesis project before teaching) and completed two 7-point Likert questionnaires: self-assessed learning (3 - 8 items per objective) and satisfaction (1 item per objective). To ensure content validity and constructive alignment, all items were copied verbatim from prespecified intended learning outcomes ([Multimedia Appendix 3](#)), yielding a criterion-referenced measure. Item order was randomized.

The primary outcome was a composite score of 20 points reflecting residents' self-assessment of knowledge and skills. It was the sum of 2 secondary outcomes: a knowledge self-assessment score (10 points) and a skill self-assessment score (10 points).

The secondary outcomes were as follows:

1. Overall satisfaction (20 points); sum of secondary outcomes 4 and 5
2. Self-assessment of knowledge acquisition (10 points), with equal weight given to each knowledge-oriented learning objective
3. Self-assessment of skill acquisition (10 points), with equal weight given to each skill-oriented learning objective
4. Satisfaction with knowledge-oriented learning objectives (10 points)
5. Satisfaction with skill-oriented learning objectives (10 points)
6. Academic achievement (submission of a thesis proposal form at 6 months; yes or no)

Statistical Analysis

Analysis was performed using R (version 4.4.3; R Foundation for Statistical Computing) in RStudio (version 2024.12.1; Posit

PBC) [29]. Because the primary outcome was assessed only at the end of the intervention, participants with missing outcome data were excluded from the analysis, resulting in a modified intention-to-treat approach with a complete-case analysis. Missing outcomes were not imputed, and reasons for missing data were documented. Multiplicity across secondary outcomes was controlled using a hierarchical fixed-sequence procedure that followed the prespecified order listed above (1 to 6). Each hypothesis was tested 2 sided at $\alpha=.05$, and testing stopped at the first nonsignificant result (subsequent analyses were exploratory).

We modeled the primary outcome using multivariable linear regression, estimating the arm effect (blended vs traditional) while adjusting for a prespecified set of the following predictor variables. Precourse framing of a research question and prior research experience were included as proxies for baseline research preparedness. Full completion of another Moodle e-learning course was included as a proxy for familiarity with digital learning environments. Unblinding was included to account for potential expectancy or performance effects linked to perceived allocation. Teacher and course day were included to account for contextual instructional and organizational variability. Age and gender were included as standard demographic covariates. Model assumptions were checked (Multimedia Appendix 4). We report β coefficients with 95% CIs and adjusted R^2 . For power, a sample size of 144 (72 per arm) was planned to provide 85% power to detect a 2-point difference on a 20-point scale (SD 4; 2-sided $\alpha=.05$). The SD was conservatively inflated given limited prior data. Internal consistency per objective was estimated using the Cronbach α .

Ethical Considerations

This trial received ethics approval from the Comité d’Ethique de la Recherche du Collège Universitaire de Médecine Générale (CUMG-IRB 2024-04-30-02). Participants received oral and written information about the study and provided written informed consent. The data collected complied with the European Union’s General Data Protection Regulation and with the data management policy of Université Claude Bernard Lyon 1 as confirmed by the university’s data protection officer. No compensation was provided to participants.

Results

Overview

We included 163 family medicine residents randomized to the blended (n=84, 51.5%) and the traditional (n=79, 48.5%) course. A total of 155 participants completed the online form at the end of the day: 78 (50.3%) in the blended course, and 77 (49.7%) in the traditional course (Figure 1). Table 2 presents the participants’ characteristics. The mean age of the participants was 25.7 (SD 3.54) years, and most (112/155, 72.3%) were women. Most participants (114/155, 73.5%) had an idea for their thesis topic. A minority had framed their research question (24/155, 15.5%) or already had a thesis supervisor (22/155, 14.2%). In total, 29.7% (46/155) of the participants reported having research experience, and 3.9% (6/155) had completed a full master’s degree or a PhD in addition to their medical studies. Less than a quarter of the participants (32/155, 20.6%) had already completed a full e-learning course on the university’s digital learning platform (Moodle). Nearly a third of the participants (48/155, 31.0%) were unblinded at the end of the course. Five teachers offered the course, and 3 of them taught in both arms.

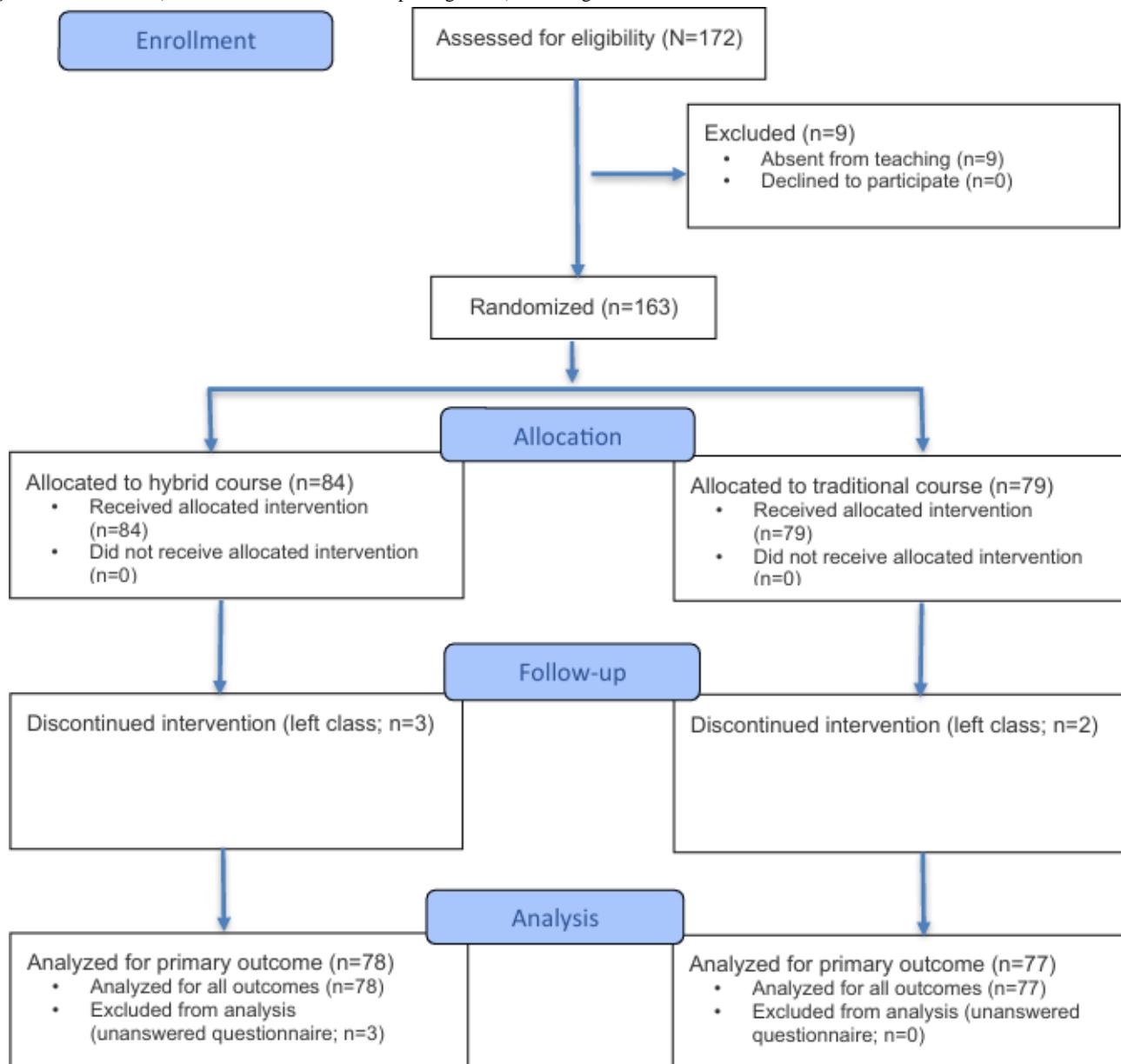
Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.

Table . Baseline participant demographics.

	Blended course (n=78)	Traditional course (n=77)
Age (y), mean (SD)	25.8 (4.1)	25.6 (2.8)
Gender, n (%)		
Men	19 (24.4)	22 (28.6)
Women	58 (74.4)	54 (70.1)
Nonbinary	1 (1.3)	1 (1.3)
Research experience (master's or PhD), n (%)		
None	45 (57.7)	57 (74.0)
Master's—first year	28 (35.9)	18 (23.4)
Master's—full	4 (5.1)	1 (1.3)
PhD	1 (1.3)	0 (0.0)
Achieved before the course, n (%)		
Idea for thesis subject	62 (79.5)	52 (67.5)
Research question formulated	10 (12.8)	14 (18.2)
Found a thesis supervisor	10 (12.8)	12 (15.6)
Other e-learning course completed on Moodle	21 (26.9)	11 (14.3)
Unblinded ^a	23 (29.5)	25 (32.5)

^aAware of group allocation.

The proportion of postrandomization exclusions was low (8/163, 4.9%) and similar in both group allocations (Multimedia Appendix 5). It was due to missing outcome data, mostly because of work-related early departure.

Primary Outcome

Knowledge and skill self-assessment did not differ significantly between the blended and traditional courses (mean difference

0.40 [maximum possible 20] points, 95% CI -0.21 to 1.02; $P=.19$; Cohen $d=0.21$; Table 3). This result remained not statistically significant after adjustment (Table 4). The most significant predictor of the primary outcome was having formulated a research question before the course, with a β coefficient of 1.87 (95% CI 1.06-2.68; $P<.001$). Unblinding during the day and teacher identity were not significantly associated with the primary outcome ($P=.13$).

Table . Education outcomes.

Outcome measure (KP ^a level)	Blended course (n=78)	Traditional course (n=77)	Mean difference (95% CI)	Cohen <i>d</i>	<i>P</i> value
Self-assessment of knowledge and skills (20 points; KP level: 2), mean (SD)	14.66 (1.92)	14.26 (1.94)	0.4 (-0.21 to 1.01)	0.21	.19 ^b
Overall satisfaction (20 points; KP level: 1), mean (SD)	14.93 (2.80)	14.77 (2.55)	0.16 (-0.68 to 1.01)	0.06	.70 ^{b,c}
Self-assessment of knowledge acquisition (10 points; KP level: 2), mean (SD)	7.49 (0.98)	7.09 (1.04)	0.4 (0.07 to 0.71)	0.38	.01 ^{b,c}
Self-assessment of skill acquisition (10 points; KP level: 2), mean (SD)	7.18 (1.20)	7.17 (1.17)	0.01 (-0.27 to 0.27)	0.009	.76 ^{c,d}
Satisfaction with knowledge-oriented educational objectives (10 points; KP level: 1), mean (SD)	7.54 (1.53)	7.42 (1.30)	0.12 (-0.00005 to 0.55)	0.08	.34 ^c
Satisfaction with skill-oriented educational objectives (10 points; KP level: 1), mean (SD)	7.40 (1.48)	7.36 (1.46)	0.04 (-0.00007 to 0.83)	0.03	.64 ^{c,e}
Submission of a thesis application form at 6 months (KP level: 3), n (%)	11 (14.1)	10 (13.0)	1.09 (0.39 to 3.1) ^f	— ^g	>.99 ^d

^aKP: Kirkpatrick [30].^bStudent *t* test (2 sided).^cOnce a hypothesis was tested and found not to be significantly different from the null hypothesis, all subsequent tests and results were considered exploratory. The secondary objectives described earlier were ranked according to this procedure.^dFisher exact test (2-sided).^eMann-Whitney *U* test.^fOdds ratio reported for this value.^gNot applicable.

Table . Effect of the study arm (blended or traditional course) on the primary outcome while adjusting for prespecified predictor variables.^a

Predictor variable	β^b (95% CI)	P value
Research question framed before the course (yes)	1.87 (1.06 to 2.68)	<.001
Research experience (yes)	-0.27 (-0.92 to 0.37)	.40
Unblinding (yes)	-0.52 (-1.20 to 0.16)	.13
Teacher 1	0.69 (-0.48 to 1.87)	.25
Teacher 2	0.91 (0.06 to 1.87)	.04
Teacher 3	-0.13 (-1.33 to 1.05)	.82
Teacher 4	-0.16 (-1.01 to 0.69)	.71
Another e-learning course completed on Moodle (yes)	0.09 (-0.63 to 0.82)	.79
Age	-0.08 (-0.16 to 0.01)	.07
Gender—man	0.26 (-0.40 to 0.93)	.43
Gender—nonbinary	1.41 (-1.20 to 4.03)	.28
Course day (day 2)	0.02 (-0.72 to 0.76)	.94
Arm (traditional course)	-0.23 (-0.99 to 0.53)	.55

^aAfter adjustment for prespecified predictor variables, there were no significant differences between the blended and traditional courses regarding the primary outcome ($P=.55$). Given the exploratory nature of the variable selection procedure, adjusted estimates should be interpreted cautiously. R^2 is the proportion of the variance explained by the model (multiple $R^2=0.21$; adjusted $R^2=0.13$; $P=.001$ for the model).

^bCoefficient representing the expected change in the primary outcome per 1-unit increase in the predictor variable.

Secondary Outcomes

As an exploratory result, the knowledge self-assessment was significantly higher in the blended course (mean difference 0.40 [maximum possible 10], 95% CI 0.07-0.71; $P=.01$; Cohen $d=0.38$).

The skill self-assessment by residents was not significantly different between the traditional and blended courses ($P=.76$). Their overall satisfaction or specific satisfaction regarding knowledge- or skill-oriented learning objectives did not differ between the traditional and blended courses. The submission of a thesis proposal form was also not found to be different between the 2 groups (11/78, 14.1% in the BL arm and 10/77, 13.0% in the traditional course arm).

Internal Consistency

Items assessing each educational objective showed acceptable Cronbach α values (>0.70) for all but one educational objective, which exhibited lower internal consistency (Cronbach $\alpha=0.47$; [Multimedia Appendix 6](#)).

Discussion

Principal Findings

Among French family medicine residents, self-assessed knowledge and skills did not differ between the blended and traditional courses. The between-arm mean difference was small, and the 95% CI was narrow, indicating low uncertainty around the estimated difference. At 6 months, academic achievement (thesis proposal submission) did not differ between arms. The only statistically significant finding—an exploratory, slightly

higher knowledge score in the blended arm—was not educationally meaningful.

Comparison With Other Studies

To our knowledge, no randomized controlled trial has evaluated whether partially replacing an already active, socioconstructivist face-to-face course with a blended format affects residents' self-reported knowledge and skills. A mixed methods randomized controlled trial compared a blended evidence-based medicine course with a lecture-based (behavioral) traditional course and found no significant difference in objectively measured knowledge and skills. However, the blended group reported higher self-efficacy, attitudes, and self-reported behaviors. Focus groups suggested a clear learner preference for the blended, integrated approach [31]. In health profession education, a meta-analysis comparing flipped classrooms with traditional, lecture-based (behavioral) courses found modest improvements in learning [32]. Flipped classrooms are a common BL configuration comparable to our study intervention combining preclass online preparation with active in-class work. Other observational studies have demonstrated the relevance of BL in medical education, most notably since the COVID-19 pandemic [33,34]. A recent meta-analysis of systematic reviews recommended further research to determine the relative benefits of BL in each individual context. It also confirmed our exploratory finding of improved knowledge acquisition with BL [35]. Moreover, a randomized controlled trial with undergraduate medical students demonstrated the noninferiority of BL vs traditional education in technical skill performance (life-saving trauma skills) and retention [36]. A randomized controlled trial of postgraduate medical students was conducted to compare BL to traditional education, and it showed significantly greater efficacy of BL for developing defibrillator

technical skills [37]. Our trial extends the evidence to academic skills related to developing a research project (thesis); it suggests that shortening face-to-face time through a structured e-learning component can yield comparable educational outcomes when the face-to-face component remains active and socioconstructivist.

More broadly, syntheses of digital education in the health professions consistently report large benefits vs no intervention [38] but little or no differences vs nondigital instruction [39].

Implications

This study may reassure teaching teams that BL can produce short-term educational outcomes comparable to those of traditional active and socioconstructivist education among residents, particularly in settings with constrained teaching capacity. Still, the implementation of BL does not address all the issues associated with the growing disparity between the number of teachers and residents. First, BL may not always be appropriate, either because of a misalignment between educational objectives aimed at developing complex competencies and the learning modality or because of the substantial resources required to develop online courses that effectively address such competencies [40]. For example, a course focused on professional communication competencies might be better suited to face-to-face relational simulation as the sole learning modality. Second, monitoring residents' acquisition of generic knowledge, skills, and competencies based on a national framework still necessitates frequent, individualized human contact. Third, strategies facilitating students' SRL may imply their supervision at the asynchronous e-learning phase, which would not allow for a significant reduction in teaching time.

The transition from traditional to blended education represents a substantial organizational shift. Teachers must allocate time to review asynchronous assessments and prepare collective feedback for the face-to-face session. It is important to estimate the time required for residents to complete all e-learning modules to avoid unintentionally increased workload. Strategies to promote SRL are also needed to optimize educational outcomes [41,42]. For example, providing dedicated time and space at the university to complete the e-learning module may help residents allocate time for it. Furthermore, working independently at home may increase feelings of stress and anxiety, as was observed during the COVID-19 pandemic [43].

Acknowledgments

The authors would like to thank Nathalie Gras (school administrator) for her help in organizing the courses (resident registration and scheduling).

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this paper.

Limitations and Strengths

One limitation of this study was the reliance on self-assessment rather than objective measures. Incorporating objective performance measures would enhance validity and mitigate self-report bias, for example, through standardized pretest-posttest knowledge tests and rubric-based scoring of residents' performance on key skills (framing a research question and assessing feasibility) rated by trained assessors. Still, formative self-assessment is integral to self-regulation, enabling feedback on progress relative to intended learning outcomes [44,45]. Moreover, the e-learning module was an initial version. A well-established, iteratively refined module might have produced higher scores among residents randomized to the BL arm. The e-learning module was completed immediately before the face-to-face session, which constrained residents' autonomy to plan their study and is partly misaligned with SRL. This design ensured full completion of the e-learning module, reduced the risk of early unblinding or contamination between groups before the face-to-face session, and standardized learning conditions (same setting and time on task). However, it did not capture real-life environmental variability, and the immediate sequencing in BL may also have facilitated knowledge activation for the subsequent face-to-face session. Although a high proportion of participants were unblinded by the end of the teaching day—a potential limitation—this variable was not associated with the primary outcome in multivariate analysis. The complete-case analysis may have introduced attrition bias; however, the proportion of missing outcome data remained low (8/163, 4.9%) and similar in both group allocations.

Regarding strengths, the narrow 95% CI suggests that an educationally meaningful difference of $\geq 10\%$ ($\geq 2/20$ points) between groups would have been detected even though the study was not designed to test equivalence or noninferiority. A formal noninferiority trial is still needed. Furthermore, the lack of significant teacher effects on the primary outcome indicates that the standardized, timed script effectively minimized instructor-specific variability.

Conclusions

Self-assessment of knowledge and skills by family practice residents did not differ between blended and traditional courses. This study can reassure educators that BL may produce short-term educational outcomes (knowledge and academic skills) comparable to those of traditional active and socioconstructivist education among residents, particularly in settings with constrained teaching capacity.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: PYM (lead), FZ (supporting), TC (supporting), RB (supporting), ML-B (lead)

Data curation: PYM

Formal analysis: PYM (equal), SS (equal)

Investigation: PYM (equal), FZ (equal), TC (equal), RB (equal), ML-B (equal)

Methodology: PYM (lead), SS (supporting), FZ (supporting), TC (supporting), RB (supporting), HM (supporting), ML-B (supporting)

Project administration: PYM (equal), ML-B (equal)

Resources: PYM

Supervision: ML-B (lead), SS (supporting), HM (supporting)

Validation: SS (equal), HM (equal), ML-B (equal)

Visualization: PYM

Writing—original draft: PYM

Writing—review and editing: PYM (equal), SS (equal), JM (equal), FZ (equal), TC (equal), RB (equal), ML-B (equal)

All authors meet all the International Committee of Medical Journal Editors criteria, approved the final version of the manuscript, and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshots of the e-learning module.

[\[DOCX File, 1677 KB - mededu_v12i1e86387_app1.docx\]](#)

Multimedia Appendix 2

Data collection form.

[\[DOCX File, 994 KB - mededu_v12i1e86387_app2.docx\]](#)

Multimedia Appendix 3

Questionnaire items for self-assessment and satisfaction (verbatim wording from intended learning outcomes).

[\[DOCX File, 21 KB - mededu_v12i1e86387_app3.docx\]](#)

Multimedia Appendix 4

Multiple linear regression model: checking assumptions.

[\[DOCX File, 86 KB - mededu_v12i1e86387_app4.docx\]](#)

Multimedia Appendix 5

Reasons for postrandomization exclusion in complete-case analysis.

[\[DOCX File, 14 KB - mededu_v12i1e86387_app5.docx\]](#)

Multimedia Appendix 6

Internal consistency of family medicine residents' self-assessment: Cronbach α values for each educational objective.

[\[DOCX File, 15 KB - mededu_v12i1e86387_app6.docx\]](#)

Checklist 1

CONSORT checklist.

[\[PDF File, 84 KB - mededu_v12i1e86387_app7.pdf\]](#)

References

1. Atun R. What are the advantages and disadvantages of restructuring a health care system to be more focused on primary care services? World Health Organization. 2004. URL: <https://iris.who.int/handle/10665/363801> [accessed 2024-08-20]

2. Russo G, Perelman J, Zapata T, Šantrić-Milićević M. The layered crisis of the primary care medical workforce in the European region: what evidence do we need to identify causes and solutions? *Hum Resour Health* 2023 Jul 14;21(1):55. [doi: [10.1186/s12960-023-00842-4](https://doi.org/10.1186/s12960-023-00842-4)] [Medline: [37443059](#)]
3. Majeed A. Shortage of general practitioners in the NHS. *BMJ* 2017 Jul 10;358:j3191. [doi: [10.1136/bmj.j3191](https://doi.org/10.1136/bmj.j3191)] [Medline: [28694250](#)]
4. Dall T, Reynolds R, Chakrabarti R, Ruttinger C, Zarek P, Parker O. The complexities of physician supply and demand: projections from 2021 to 2036. : Association of American Medical Colleges; 2024 URL: <https://www.aamc.org/media/75236/download?attachment> [accessed 2026-01-31]
5. Framework for action on the health and care workforce in the WHO European region 2023–2030. World Health Organization, European Region. 2023. URL: <https://iris.who.int/bitstream/handle/10665/372563/73wd08e-HealthCareWorkforce-230575.pdf?sequence=5> [accessed 2024-08-15]
6. Pettersson SM, Liaw WR, Tran C, Bazemore AW. Estimating the residency expansion required to avoid projected primary care physician shortages by 2035. *Ann Fam Med* 2015 Mar;13(2):107-114. [doi: [10.1370/afm.1760](https://doi.org/10.1370/afm.1760)] [Medline: [25755031](#)]
7. Chevillard G, Lucas-Gabrielli V, Mousques J. « Déserts médicaux » en France: état des lieux et perspectives de recherches. *L'Espace géographique* 2018 Dec 18;Tome 47(4):362-380. [doi: [10.3917/eg.474.0362](https://doi.org/10.3917/eg.474.0362)]
8. Mounthon L, Clavelou P. Réforme du troisième cycle des études médicales en France. *Presse Méd Form* 2024 Apr;5(2):97-107. [doi: [10.1016/j.ppmfor.2024.02.001](https://doi.org/10.1016/j.ppmfor.2024.02.001)]
9. Compagnon L, Bail P, Huez JF, et al. Définitions et descriptions des compétences en médecine générale. *Exercer* 2013;24(108):148-155 [FREE Full text]
10. Kulasegaram K, Mylopoulos M, Tonin P, et al. The alignment imperative in curriculum renewal. *Med Teach* 2018 May;40(5):443-448. [doi: [10.1080/0142159X.2018.1435858](https://doi.org/10.1080/0142159X.2018.1435858)] [Medline: [29490525](#)]
11. Bay E, Bagceci B, Cetin B. The effects of social constructivist approach on the learners' problem solving and metacognitive levels. *J Soc Sci* 2012;8(3):343-349. [doi: [10.3844/jssp.2012.343.349](https://doi.org/10.3844/jssp.2012.343.349)]
12. Vygotsky L, Cole M, John-Steiner V, Scribner S, Souberman E. *Mind in Society: The Development of Higher Psychological Processes*: Harvard University Press; 1978.
13. Heck AJ, Cross CE, Tatum VY, Chase AJ. Active learning among health professions' educators: perceptions, barriers, and use. *Med Sci Educ* 2023 May 2;33(3):719-727. [doi: [10.1007/s40670-023-01793-0](https://doi.org/10.1007/s40670-023-01793-0)] [Medline: [37501807](#)]
14. Tudor Car L, Poon S, Kyaw BM, et al. Digital education for health professionals: an evidence map, conceptual framework, and research agenda. *J Med Internet Res* 2022 Mar 17;24(3):e31977. [doi: [10.2196/31977](https://doi.org/10.2196/31977)] [Medline: [35297767](#)]
15. Bonk CJ, Graham CR. *The Handbook of Blended Learning: Global Perspectives, Local Designs*: John Wiley & Sons; 2012.
16. Saichai K. Blended, flipped, and hybrid learning: definitions, developments, and directions. *New Direct Teach Learn* 2020;2020(164):95-104. [doi: [10.1002/tl.20428](https://doi.org/10.1002/tl.20428)]
17. Hurtubise L, Hall E, Sheridan L, Han H. The flipped classroom in medical education: engaging students to build competency. *J Med Educ Curric Dev* 2015;2:JMECD.S23895. [doi: [10.4137/JMECD.S23895](https://doi.org/10.4137/JMECD.S23895)] [Medline: [35187252](#)]
18. Liu Q, Peng W, Zhang F, Hu R, Li Y, Yan W. The effectiveness of blended learning in health professions: systematic review and meta-analysis. *J Med Internet Res* 2016 Jan 4;18(1):e2. [doi: [10.2196/jmir.4807](https://doi.org/10.2196/jmir.4807)] [Medline: [26729058](#)]
19. Vallée A, Blacher J, Cariou A, Sorbets E. Blended learning compared to traditional learning in medical education: systematic review and meta-analysis. *J Med Internet Res* 2020 Aug 10;22(8):e16504. [doi: [10.2196/16504](https://doi.org/10.2196/16504)] [Medline: [32773378](#)]
20. Moskal PD. Evaluating the outcomes and impact of hybrid courses. *New Direct Teach Learn* 2017;2017(149):19-26. [doi: [10.1002/tl.20223](https://doi.org/10.1002/tl.20223)]
21. Mason GS, Shuman TR, Cook KE. Comparing the effectiveness of an inverted classroom to a traditional classroom in an upper-division engineering course. *IEEE Trans Educ* 2013 Nov;56(4):430-435. [doi: [10.1109/TE.2013.2249066](https://doi.org/10.1109/TE.2013.2249066)]
22. Lage MJ, Platt GJ, Treglia M. Inverting the classroom: a gateway to creating an inclusive learning environment. *J Econ Educ* 2000 Jan;31(1):30-43. [doi: [10.1080/00220480009596759](https://doi.org/10.1080/00220480009596759)]
23. AlSheikh MH, Iqbal MZ. Student perceptions regarding group learning activities in a hybrid medical curriculum. *Med Sci Educ* 2019 Sep 16;29(4):1221-1228. [doi: [10.1007/s40670-019-00817-y](https://doi.org/10.1007/s40670-019-00817-y)] [Medline: [34457606](#)]
24. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. *BMJ* 2025 Apr 14;389:e081123. [doi: [10.1136/bmj-2024-081123](https://doi.org/10.1136/bmj-2024-081123)] [Medline: [40228833](#)]
25. Create a blocked randomisation list. Sealed Envelope. 2024. URL: <https://www.sealedenvelope.com/simple-randomiser/v1/lists> [accessed 2024-05-14]
26. Schunk DH, Zimmerman BJ. *Self-Regulation of Learning and Performance: Issues and Educational Applications*: Routledge; 1994.
27. Zimmerman BJ. Becoming a self-regulated learner: an overview. *Theory Pract* 2002;41(2):64-70. [doi: [10.1207/s15430421tip4102_2](https://doi.org/10.1207/s15430421tip4102_2)]
28. Gavarkovs AG, Finan E, Jensen RD, Brydges R. When I say ... active learning. *Med Educ* 2024 Aug;58(8):896-897. [doi: [10.1111/medu.15383](https://doi.org/10.1111/medu.15383)] [Medline: [38503311](#)]
29. The data science code editor. Posit Software. URL: <http://www.posit.co/> [accessed 2026-01-24]
30. Kirkpatrick DL, Kirkpatrick JD. *Evaluating Training Programs: The Four Levels*: Berrett-Koehler Publishers; 2006.

31. Ilic D, Nordin RB, Glasziou P, Tilson JK, Villanueva E. A randomised controlled trial of a blended learning education intervention for teaching evidence-based medicine. *BMC Med Educ* 2015 Mar 10;15:39. [doi: [10.1186/s12909-015-0321-6](https://doi.org/10.1186/s12909-015-0321-6)] [Medline: [25884717](#)]
32. Hew KF, Lo CK. Flipped classroom improves student learning in health professions education: a meta-analysis. *BMC Med Educ* 2018 Mar 15;18(1):38. [doi: [10.1186/s12909-018-1144-z](https://doi.org/10.1186/s12909-018-1144-z)] [Medline: [29544495](#)]
33. Atwa H, Shehata MH, Al-Ansari A, et al. Online, face-to-face, or blended learning? Faculty and medical students' perceptions during the COVID-19 pandemic: a mixed-method study. *Front Med (Lausanne)* 2022;9:791352. [doi: [10.3389/fmed.2022.791352](https://doi.org/10.3389/fmed.2022.791352)] [Medline: [35186989](#)]
34. Abdull Mutualib 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](https://doi.org/10.1186/s12909-022-03579-1)] [Medline: [35786374](#)]
35. Lockey A, Bland A, Stephenson J, Bray J, Astin F. Blended learning in health care education: an overview and overarching meta-analysis of systematic reviews. *J Contin Educ Health Prof* 2022 Oct 1;42(4):256-264. [doi: [10.1097/CEH.0000000000000455](https://doi.org/10.1097/CEH.0000000000000455)] [Medline: [36070399](#)]
36. Bergmans E, Billington A, Thies KC. From tradition to innovation: a comparison of the traditional 4-step approach versus a blended learning modification for technical skills teaching. *Scand J Trauma Resusc Emerg Med* 2023 Nov 14;31(1):80. [doi: [10.1186/s13049-023-01127-4](https://doi.org/10.1186/s13049-023-01127-4)] [Medline: [37964361](#)]
37. Siebert JN, Gланетас A, Grange M, Haddad K, Courvoisier DS, Lacroix L. Impact of blended learning on manual defibrillator's use: a simulation-based randomized trial. *Nurs Crit Care* 2022 Jul;27(4):501-511. [doi: [10.1111/nicc.12713](https://doi.org/10.1111/nicc.12713)] [Medline: [34519140](#)]
38. 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](https://doi.org/10.1001/jama.300.10.1181)] [Medline: [18780847](#)]
39. Vaona A, Banzi R, Kwag KH, et al. E-learning for health professionals. *Cochrane Database Syst Rev* 2018 Jan 21;1(1):CD011736. [doi: [10.1002/14651858.CD011736.pub2](https://doi.org/10.1002/14651858.CD011736.pub2)] [Medline: [29355907](#)]
40. Wurth S, Maisonneuve H, Moussa MA, et al. Development of complex pedagogical competencies and reflexivity in clinical teachers via distance learning: a mixed methods study. *Med Educ Online* 2023 Dec;28(1):2265163. [doi: [10.1080/10872981.2023.2265163](https://doi.org/10.1080/10872981.2023.2265163)] [Medline: [37818594](#)]
41. Azevedo R. Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educ Psychol* 2005;40(4):199-209. [doi: [10.1207/s15326985ep4004_2](https://doi.org/10.1207/s15326985ep4004_2)]
42. Schumacher C, Ifenthaler D. Features students really expect from learning analytics. *Comput Human Behav* 2018 Jan;78:397-407. [doi: [10.1016/j.chb.2017.06.030](https://doi.org/10.1016/j.chb.2017.06.030)]
43. Patte KA, Battista K, Ferro MA, et al. School learning modes during the COVID-19 response and pre- to during pandemic mental health changes in a prospective cohort of Canadian adolescents. *Soc Psychiatry Psychiatr Epidemiol* 2024 Jan;59(1):137-150. [doi: [10.1007/s00127-023-02557-2](https://doi.org/10.1007/s00127-023-02557-2)] [Medline: [37668673](#)]
44. Brown GT, Harris LR. The future of self-assessment in classroom practice: reframing self-assessment as a core competency. *Frontline Learn Res* 2014 Apr;3:22-30. [doi: [10.14786/flr.v2i1.24](https://doi.org/10.14786/flr.v2i1.24)]
45. Andrade HL. Students as the definitive source of formative assessment: academic self-assessment and the self-regulation of learning. In: *Handbook of Formative Assessment*; Routledge; 2010.

Abbreviations

BL: blended learning

CONSORT: Consolidated Standards of Reporting Trials

SRL: self-regulated learning

Edited by LT Car; submitted 23.Oct.2025; peer-reviewed by J Wilkinson, M Arsenault; accepted 12.Jan.2026; published 04.Feb.2026.

Please cite as:

Meunier PY, Schlatter S, Macabrey J, Zorzi F, Colleony T, Boussageon R, Maisonneuve H, Lamort-Bouché M
Blended Learning Compared With Face-to-Face Learning Among Family Medicine Residents: Randomized Controlled Trial
JMIR Med Educ 2026;12:e86387

URL: <https://mededu.jmir.org/2026/1/e86387>

doi: [10.2196/86387](https://doi.org/10.2196/86387)

© Pierre-Yves Meunier, Sophie Schlatter, Juliette Macabrey, Frédéric Zorzi, Thomas Colleony, Rémy Boussageon, Hubert Maisonneuve, Marion Lamort-Bouché. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 4.Feb.2026. 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.

Fostering Inductive and Deductive Learning in Oral Microbiology and Immunology With a Dual-Role Duel Card Game: Explanatory Sequential Mixed Methods Study

Kawin Sipiyaruk¹, DDS, PhD; Ratchapin Laovanitch Srisatjaluk², PhD; Ponpentip Chaisiri³; Kanokwan Jantawatchai³; Jidapha Srihatai³; Chutikarn Sonyaem³; Thanaphat Traikorakot³; Suphasuta Pookboonmee⁴

¹Department of Orthodontics, Faculty of Dentistry, Mahidol University, 6 Yothi Road, Ratchathewi, Bangkok, Thailand

²Mahidol International Dental School, Faculty of Dentistry, Mahidol University, Bangkok, Thailand

³Doctor of Dental Surgery Program, Faculty of Dentistry, Mahidol University, Bangkok, Thailand

⁴Medical Education Technology Center, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

Corresponding Author:

Kawin Sipiyaruk, DDS, PhD

Department of Orthodontics, Faculty of Dentistry, Mahidol University, 6 Yothi Road, Ratchathewi, Bangkok, Thailand

Abstract

Background: Game-based learning has emerged as an effective learning strategy in health care education. However, no games have been specifically designed to support cognitive improvement for diverse learning styles in oral microbiology and immunology.

Objective: This study aimed to develop and evaluate an educational card game designed to support diverse learning styles in oral microbiology and immunology, using a duel-style format.

Methods: An explanatory sequential mixed methods study was conducted with 40 third-year dental students, where half of them were assigned to the first group, starting as the host, while those in the other group began as the microbe. Participants alternated between the microbe and host roles during gameplay. Active engagement through playing as the microbe facilitated knowledge acquisition through observation, supporting inductive learning. On the other hand, the host role aimed to promote the application of knowledge for decision-making, facilitating deductive learning. Quantitative data were collected using pre- and postknowledge assessments and satisfaction questionnaires. Qualitative insights were obtained through semistructured interviews exploring learning experiences when playing as the microbe compared to the host.

Results: Students demonstrated significant improvements in knowledge scores across the 3 assessments ($P<.001$), with no difference between groups ($P=.85$). They also perceived the game positively in all 3 aspects (usefulness, ease of use, and enjoyment). Qualitative findings revealed that role variation supported both inductive and deductive learning processes. Participants valued the combination of pedagogical and entertaining components, leading to the game's motivation and engagement. A conceptual framework demonstrated key emerging themes relevant to the game design and implementation, including learner profile, learning setting, game design, learning process, and learning outcomes.

Conclusions: The card game demonstrated its potential in enhancing knowledge acquisition and student engagement in oral microbiology and immunology. Role-switching between the host and microbe was perceived by participants to facilitate different learning experiences. Further research is recommended to investigate long-term retention and broader practicality.

(*JMIR Med Educ* 2026;12:e80048) doi:[10.2196/80048](https://doi.org/10.2196/80048)

KEYWORDS

active learning; card game; educational game; game-based learning; microbiology; immunology

Introduction

Oral microbiology provides essential insights into bacteria, viruses, and fungi in the oral cavity, while immunology focuses on the mechanisms by which the human body defends itself against these pathogens. This knowledge enables health care professionals, including dental practitioners, to understand pathogen transmission, immune interactions, and strategies for disease management and prevention [1-3]. The COVID-19

pandemic, for example, significantly impacted dental practices by introducing new protocols to reduce viral transmission through respiratory droplets [4,5]. Studying microbiology and immunology is therefore important for understanding infectious diseases, and it has proven essential for health care education.

Traditional teaching methods have served as the main method of education worldwide. While this approach can reduce anxiety by minimizing the need for active student participation, it can limit engagement due to the predominantly 1-way

communication [6,7]. Disengagement in traditional classrooms has been shown to negatively impact academic performance [8-10]. On the other hand, active learning strategies, which promote critical thinking and student interaction, are associated with better engagement [11,12]. In addition, not all students have the same learning preferences [13]. Some of them prefer direct participation and decision-making, while others may learn more effectively by observing patterns. Educational tools designed for both learning preferences may support diverse learning needs.

Game-based learning has been considered an innovative strategy to address the limitations of traditional teaching methods by incorporating game mechanics into educational settings. Game-based learning has been applied in medical and dental education [14-17], including microbiology and immunology [18,19]. This approach is widely recognized due to its ability to enhance student engagement and motivation in various disciplines of dental education, such as dental public health [20], human anatomy [21], oral diagnosis [22,23], orthodontics [24], endodontics [25], patient safety [26], clinical administration [27], and teledentistry [28]. These games can be implemented through various formats, ranging from digital simulations to physical formats such as board and card games.

Despite the growing number of educational games, most of them are not designed to accommodate a range of learning preferences or gameplay experience levels. In other words, they may not explicitly support both inductive (structured observation) and deductive learning (active problem-solving), which are 2 distinct approaches through which students engage with and make sense of content. This study aimed to develop and evaluate an educational card game designed to support diverse learning needs in oral microbiology and immunology, using a duel-style format in which players alternate between the roles of microbe and host to simulate infection and immune response dynamics. Specifically, the objectives were to assess the game's effectiveness in improving knowledge acquisition and to investigate how learners engaged differently when beginning as the microbe compared to the host.

Methods

Game Development

Game Concept

The “Invasion” game was designed to support teaching and learning in oral microbiology and immunology, where players alternate roles between microbes and hosts. This game served as a novel educational tool to engage dental students more deeply with the learning content. The game was designed to align with the expected learning outcomes in oral microbiology and immunology, requiring learners to apply their knowledge of microbial pathogenesis and human immune responses to

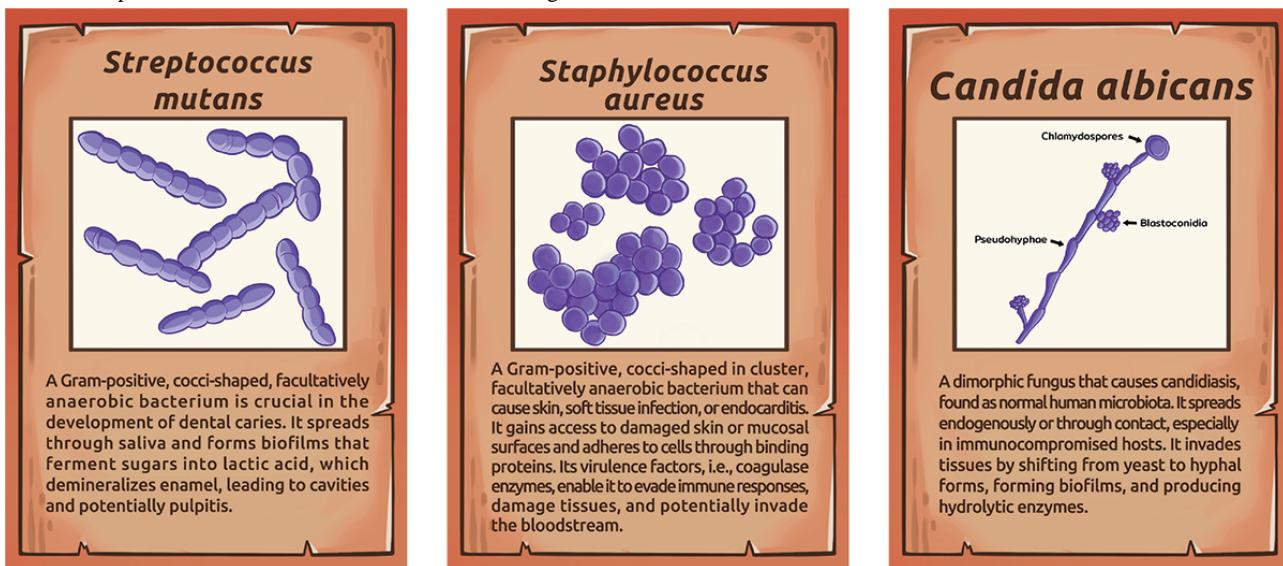
infection. The role alternation within the game provides different forms of engagement, encouraging learners to interact with the content either through structured observation as microbes (inductive learning) or through direct decision-making as hosts (deductive learning). A physical card format was chosen because it encourages face-to-face interaction, peer discussion, and active participation, which strengthens the learning experience when students switch between roles. In addition, compared with a digital format, this design helps students remain focused on the learning activity itself rather than being distracted by technological navigation.

Instructional Design and Learning Tasks

The pedagogical framework of this game is based on the constructivist learning theory, where learners are expected to actively construct new knowledge through problem-solving and critical reflection, based on their existing understanding. Engagement with the game elements can allow students to develop a deeper understanding of microbiological and immunological concepts. The dual-role format supports different types of cognitive engagement. Playing as the microbe emphasizes reflective observation and the construction of clues (inductive learning), while the host role requires active problem-solving and the application of knowledge to identify pathogens and apply appropriate responses (deductive learning). Another important component of the game is the feedback mechanism. Each action taken by the player provided an opportunity for formative assessment, allowing for real-time correction and learning from mistakes. The feedback mechanism can support learners to adjust their strategies and understandings during the gameplay. Through this process of testing, feedback, and adjustment, students can engage in a learning cycle that mirrors scientific reasoning and clinical decision-making.

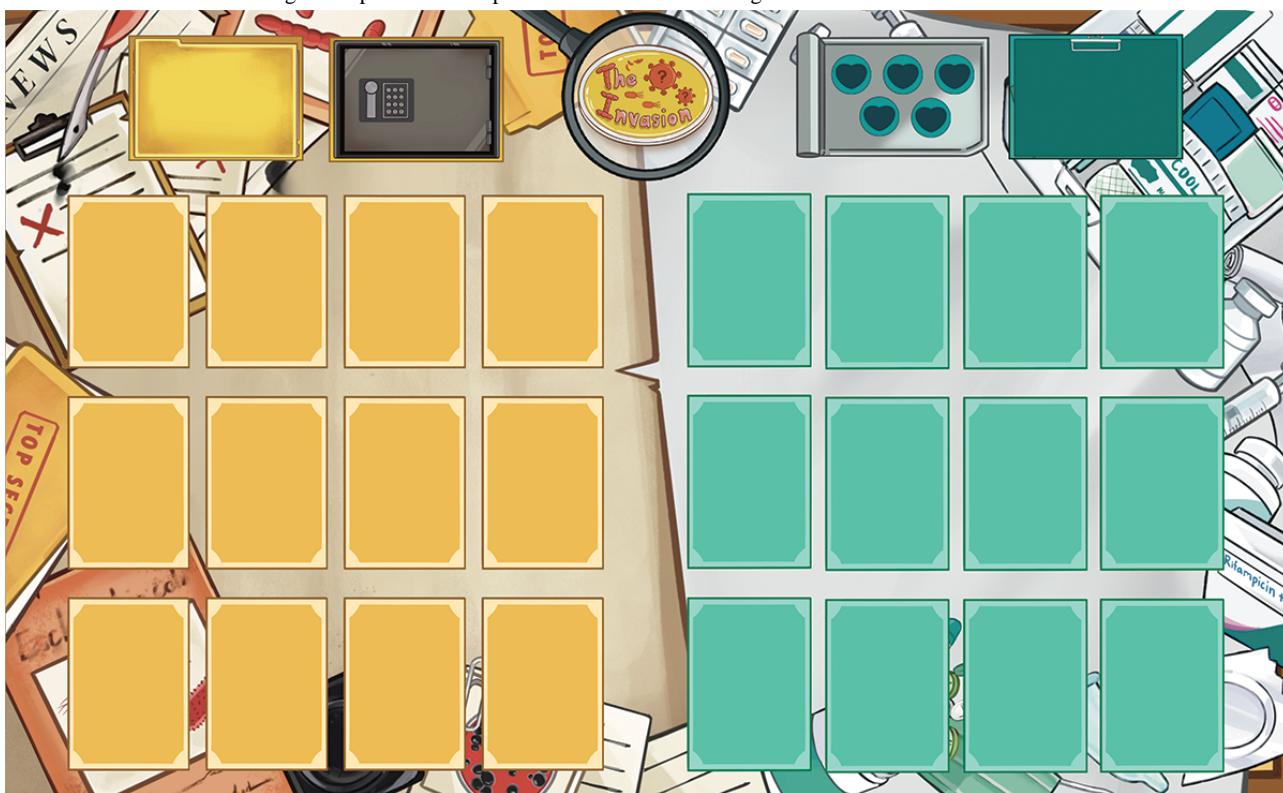
Game Interface and Interaction

According to the gameplay mechanics, the game begins with the microbe player selecting 3 “Microbe” cards, along with their corresponding “Key” cards and 7 “Hint” cards (Figure 1). The host, in turn, starts the game with 5 heart points and a set of 7 “Protection” cards. During gameplay, the microbe presents “Hint” cards, prompting the host to select 1 of 3 actions: play up to 3 “Protection” cards in response to the clues, discard 3 “Protection” cards in exchange for a “Key” card offering more detailed information, or use an “Attack” card to guess the identity of a microbe. A correct guess leads to the revelation of that microbe, while an incorrect guess results in a loss of heart points calculated as the number of “Hint” cards played minus the number of correctly placed “Protection” cards. Game completion occurs when the host successfully identifies 3 microbes, resulting in a host victory, or when all heart points or “Attack” cards are used, which means the microbe player wins.

Figure 1. Examples of “Microbe” cards used in the “Invasion” game.

To further support gameplay, a cardboard was designed as an organizing component (Figure 2). The cardboard provided designated spaces for different types of cards (eg, microbe, host, or support actions), guiding players in organizing card placement throughout the game. It was intended to reduce confusion during

gameplay, thereby enhancing usability and the overall learning experience. Its purpose was to minimize the cognitive load related to card management, allowing players to focus more on the learning tasks.

Figure 2. The cardboard with designated spaces for card placement in the “Invasion” game.

Effective gameplay requires the microbe player to present hints in a way that is strategically challenging to test the host’s interpretation skills. In turn, the host must analyze the clues, make careful decisions about resource use, and minimize incorrect guesses in order to preserve their heart points. These interactive dynamics create a learning environment that reflects the cognitive and analytical challenges involved in diagnosing and managing microbial infections in clinical practice. By

placing students in alternating roles with distinct cognitive demands, the game supports the development of microbiological understanding, analytical thinking, and adaptability, which are skills essential for clinical reasoning and for meeting the varied learning needs of students. This gameplay structure is designed to engage students in different types of thinking, depending on their assigned role. A summary of the game components is provided in Table 1.

Table . Description of card types used in the game.

Card type	Used by	Description
Microbe cards	Microbe	Represent various microbes selected for their clinical relevance in dentistry. These cards define the core challenges posed by the microbe player during gameplay.
Protection cards	Host	Provide the host with strategies to prevent or reduce damage from incorrect microbial identification. Each card contains information on treatments, immune responses, or behavioral changes.
Hint cards	Microbe	Contain clues related to the microbe's characteristics, such as virulence factors, clinical manifestations, and transmission routes. They support progressive disclosure of information to the host.
Key cards	Microbe	Offer more specific and detailed information about a microbe than "Hint" cards, enhancing the host's ability to make accurate identifications. Obtained through card exchange.
Attack cards	Host	Allow the host to make an informed guess about the identity of a microbe based on accumulated clues. Limited to 5 per game, requiring careful use.

Research Design

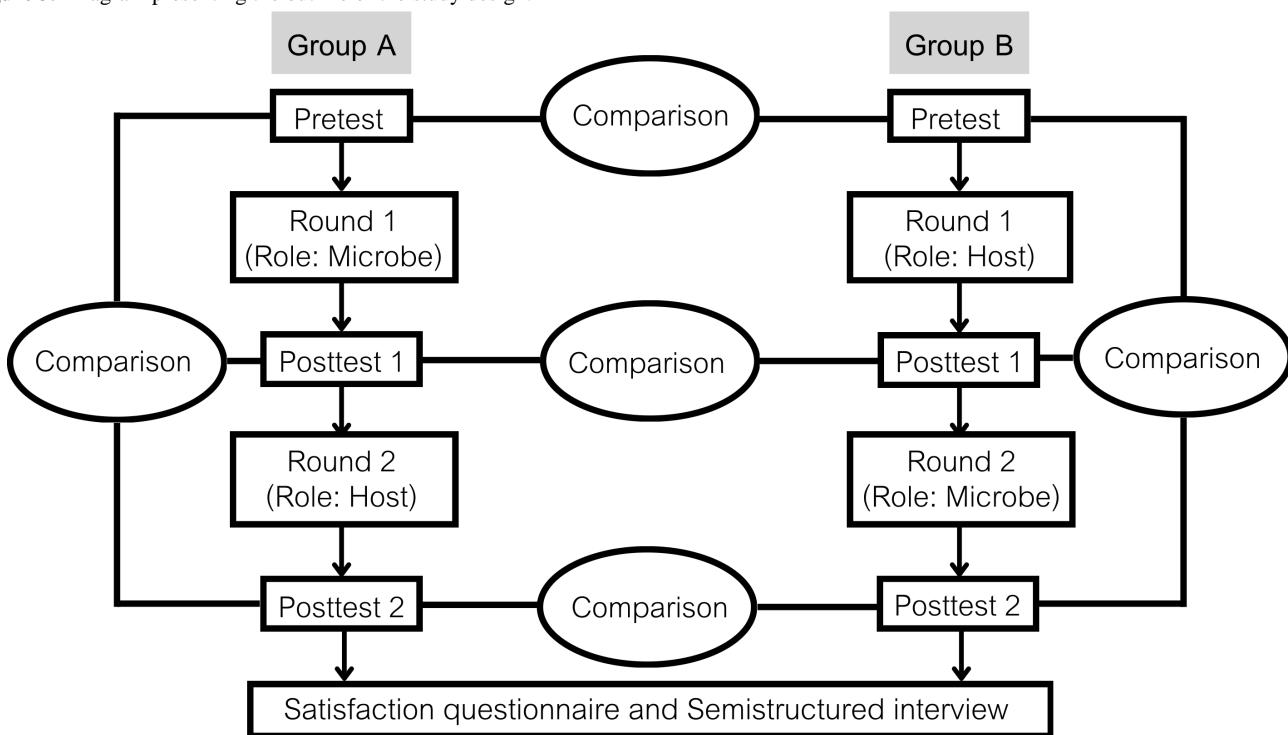
This study used an explanatory mixed methods design to evaluate the "Invasion" game and to examine differences in learning experiences and outcomes when students engaged in gameplay as either a microbe or a host.

The quantitative phase used an experimental design with partial randomization to assess knowledge acquisition and compare learning outcomes based on gameplay role order. A total of 40 participants were allowed to self-select their partners, forming pairs based on personal preference to reflect natural peer interactions, commonly found in educational settings [21]. Within each pair, roles were randomly assigned using a coin toss method to determine which player would begin as the host or the microbe. Prior to gameplay, participants completed a preknowledge test (Pretest), followed by 1 round of the "Invasion" game in their assigned role. A postknowledge test (Posttest 1) was administered immediately after the first round. In the second round, participants switched roles and completed another postknowledge test (Posttest 2). A satisfaction questionnaire was also completed at the end of the second round. All test items, game materials, and instructions were provided in English.

For the qualitative phase, all participants took part in short semistructured interviews to explore their experiences and

learning processes while engaging with both gameplay roles. This approach provided clarification on how the game's dual-role structure facilitated distinct forms of engagement, with students tending to adopt a more reflective and observational approach when playing as the microbe, and a more analytical and decision-driven approach when playing as the host. All 40 participants were interviewed to ensure full representation of perspectives across both gameplay roles and to enhance the integration of qualitative insights with quantitative findings. This decision was made because the approach was intended not only to achieve thematic saturation but also to ensure that the qualitative data contextualized the quantitative results, while the brief and focused nature of the interviews made it feasible to include all participants. The semistructured format allowed for a consistent set of guiding questions while providing flexibility to probe deeper into participant reflections. Interviews were conducted immediately after the quantitative phase to ensure relevance and recall.

All procedures were completed in a single session, where each game round (host or microbe) lasted approximately 25 - 30 minutes. All participants completed their activities under the same conditions, and knowledge assessments were completed independently within a fixed time frame. This standardized approach ensured consistency and comparability of learning experiences across groups. The full study protocol is illustrated in Figure 3.

Figure 3. Diagram presenting the outline of the study design.

Research Participants

The research population consisted of third-year undergraduate dental students enrolled in the 2024 academic year at the Faculty of Dentistry, Mahidol University. They were eligible if they had successfully completed 3 prerequisite courses: Fundamentals of Medical Microbiology for Dental Students, Essential Immunology for Dental Students, and Oral Microbiology and Immunology. Completion of these courses ensured a baseline understanding of core concepts relevant to the game content. While all game materials and knowledge assessments were designed in English, students with insufficient English proficiency to understand the game instructions were excluded from the study to ensure that language barriers did not affect gameplay engagement or performance. This exclusion was determined through an orientation session in which students were asked to demonstrate comprehension of the English instructions. Additionally, any participant who failed to complete a required task within the allocated time was withdrawn.

The required sample size was determined using G*Power software (version 3.1; Heinrich-Heine-Universität Düsseldorf). According to a previous study [21], the mean and SD of the pretest score were 9.37 and 2.40, respectively, while those of the posttest score were 11.37 and 1.96. With a statistical power of 95% and a significance level of $P=.85$, the required sample size per group was 18, resulting in a minimum total of 36 participants. To compensate for a potential 10% dropout rate, the sample size was set at 40 participants (20 per group).

Participant allocation followed a 2-step process. Students first formed pairs through voluntary matching. Within each pair, a coin toss was used to randomly assign 1 participant to begin as the host and the other as the microbe. This randomization step helped mitigate potential bias introduced by voluntary pairing. Allowing students to choose their own partners made the activity

feel more natural, while randomly assigning gameplay roles ensured a fair comparison between the host and microbe roles. By assigning roles randomly, the study preserved experimental control and enabled the qualitative phase to capture the experiences of participants who may have been placed in roles misaligned with their learning preferences. This role mismatch provided valuable insight into the cognitive and emotional challenges encountered by different learners, offering a more comprehensive understanding of the game's effectiveness and highlighting both the strengths and limitations of its design across diverse learner profiles.

Data Collection Tools

Knowledge Assessments

To evaluate knowledge acquisition, 3 structured assessments were administered at different stages of the study: Pretest, Posttest 1, and Posttest 2. Pretest aimed to assess students' baseline understanding and help identify individual learning gaps prior to the intervention. Posttest 1 and Posttest 2 were used to measure knowledge gains following participation in the learning activities. Each assessment consisted of the same 20 multiple-choice questions to ensure consistency in content and difficulty. However, to minimize the influence of recall bias, both the order of the questions and the arrangement of answer choices were randomized across the assessments.

Content validity was confirmed through expert review by 3 microbiologists who were not involved in the study. These experts evaluated each question for its relevance and alignment with the expected learning outcomes. The Index of Item-Objective Congruence (IOC) was used, with items scoring below 0.5 being revised based on expert feedback. Face validity was assessed by a group of fourth-year dental students who had completed the same microbiology and immunology courses required for inclusion in the study. Involving these senior

students helped ensure that the test items were appropriate and clearly worded, and this approach prevented contamination of the research participants.

Satisfaction Questionnaire

To investigate user perceptions of the learning experience, a self-administered questionnaire comprising 18 items was used. Each item was rated on a 5-point Likert scale, ranging from 1 (“Strongly disagree”) to 5 (“Strongly agree”), and adapted from previously validated instruments [22,27,28]. The questionnaire assessed 3 key dimensions commonly used in the evaluation of serious games: perceived usefulness, ease of use, and enjoyment. Content validity was established through expert review by 3 dental educators, who evaluated each item for relevance and alignment with the intended constructs. Items with an IOC below 0.5 were revised through iterative feedback until acceptable agreement was achieved.

Semistructured Interviews

To complement the quantitative findings on how different gameplay roles supported varied learning approaches, semistructured interviews were conducted using a topic guide developed by the research team. The topic guide comprised 5 main areas, with flexibility to probe further based on participants’ responses. These questions examined students’ reflections on the learning experience, the influence of assigned roles (microbe or host) on learning and engagement, and suggestions for improvement. Interview responses were audio-recorded with permission and transcribed using a verbatim technique. This process ensured that the learning experiences were captured with depth and fidelity.

Data Analysis

Quantitative data were analyzed using IBM SPSS Statistics for Windows, version 29.0. Descriptive statistics were used to summarize participant characteristics and assessment scores. A 2-way repeated-measures ANOVA was conducted to examine the effects of assigned role order (host and microbe) and time (Pretest, Posttest 1, and Posttest 2). Bonferroni-adjusted post hoc tests were used to explore pairwise differences across time and between groups. Assumptions of normality and sphericity were verified prior to analysis to ensure that the requirements for repeated-measures ANOVA were met. Data from the satisfaction questionnaire were analyzed using means and SDs. Statistical significance was set at $P < .05$.

Table . Knowledge assessments across time points by group.

Group	Assessment	Mean score (SD)	95% CI
A	Pretest	9.85 (3.72)	8.43 to 11.27
	Posttest 1	12.4 (3.65)	10.87 to 13.93
	Posttest 2	14.55 (3.38)	13.29 to 15.82
B	Pretest	9.55 (2.44)	8.13 to 10.97
	Posttest 1	11.8 (3.07)	10.27 to 13.33
	Posttest 2	15.9 (2.05)	14.64 to 17.17

Effects of Time and Group (A 2-Way Repeated-Measures ANOVA)

A 2-way repeated-measures ANOVA was conducted to examine improvements in knowledge scores over time of assessments and whether there were differences between the 2 groups (host-first vs microbe-first). According to Table 3, the analysis demonstrated a significant main effect of time ($F_{2,76}=54.02$; $P<.001$), indicating that there were improvements in knowledge

scores across the 3 assessments. However, the interaction between time and group was not statistically significant ($F_{2,76}=1.94$; $P=.15$), suggesting that both groups demonstrated a similar pattern of improvement over time. The between-subjects analysis showed no significant main effect of group on knowledge scores ($F_{1,38}=0.038$; $P=.85$), indicating that initial group assignment did not influence overall performance (Table 4).

Table . Results of 2-way repeated-measures ANOVA examining within-subjects effects.

Source	Type III sum of squares	Mean square	F test (df)	P value	η^2
Time	614.017	307.008	54.02 (2, 76)	<.001	.587
Time × group	22.050	11.025	1.94 (2, 76)	.15	.049
Error (time)	431.933	5.683	— ^a	—	—

^aNot applicable.

Table . Results of 2-way repeated-measures ANOVA examining between-subjects effects.

Source	Type III sum of squares	Mean square	F test (df)	P value	η^2
Group	0.675	0.675	0.04 (1, 38)	.85	.001
Error (group)	672.317	17.693	— ^a	—	—

^aNot applicable.

Pairwise Comparisons of Assessment Time Points

Given the absence of a significant interaction between time and group and no main effect of group, Bonferroni-adjusted pairwise comparisons were conducted by combining data from both groups. As shown in Table 5, knowledge scores increased significantly between each assessment point: from Pretest to

Posttest 1 ($P<.001$), from Posttest 1 to Posttest 2 ($P<.001$), and from Pretest to Posttest 2 ($P<.001$). These findings reflect a cumulative learning effect over time. To further explore learning progression within each role-assigned group, Bonferroni-adjusted pairwise comparisons were performed separately within each group (Table 6).

Table . Bonferroni-adjusted pairwise comparisons of knowledge scores across assessment time points.

Comparison		Mean difference (95% CI)	SE	P value
Pretest	Posttest 1	−2.4 (−3.696 to −1.104)	0.517	<.001
Pretest	Posttest 2	−5.525 (−6.889 to −4.161)	0.545	<.001
Posttest 1	Posttest 2	−3.125 (−4.47 to −1.78)	0.537	<.001

Table . Bonferroni-adjusted pairwise comparisons of knowledge scores across assessment time points within each group.

Group	Comparison		Mean difference (95% CI)	SE	P value
A	Pretest	Posttest 1	-2.55 (-4.382 to -0.718)	0.732	.004
	Pretest	Posttest 2	-4.70 (-6.629 to -2.771)	0.770	<.001
	Posttest 1	Posttest 2	-2.15 (-4.052 to -0.248)	0.759	.02
B	Pretest	Posttest 1	-2.25 (-4.082 to -0.418)	0.732	.01
	Pretest	Posttest 2	-6.35 (-8.279 to -4.421)	0.770	<.001
	Posttest 1	Posttest 2	-4.10 (-6.002 to -2.198)	0.759	<.001

User Satisfaction

Participants from both groups perceived the game as a positive experience in all aspects: perceived usefulness, ease of use, and

enjoyment (Table 7). Although Group B demonstrated slightly higher scores across all categories, there were no statistically significant differences between the groups ($P>.05$).

Table . Comparison of user satisfaction between the 2 groups.

Perceptions	Group A, mean (SD)	Group B, mean (SD)	Mean difference (95% CI)	P value
Perceived usefulness	4.51 (0.50)	4.53 (0.40)	0.02 (-0.31 to 0.27)	.91
Perceived ease of use	4.58 (0.51)	4.64 (0.34)	0.07 (-0.34 to 0.21)	.63
Perceived enjoyment	4.63 (0.55)	4.67 (0.55)	0.03 (-0.39 to 0.32)	.85

Learner Experiences When Engaging Within the Gameplay

Semistructured interviews were conducted to gain in-depth information about the gameplay experiences of participants, exploring how their initial roles (host-first or microbe-first) influenced their learning engagement and perceptions of the game. The following key themes emerged from the thematic analysis, highlighting factors that shaped learners' experiences during gameplay.

Theme 1: Learner Profile

Background Knowledge

Participants reported that their background knowledge significantly influenced their engagement and effectiveness during gameplay. Several students described challenges when their prior understanding was limited, particularly when they were required to make decisions. To manage these difficulties, participants strategically chose when to read detailed card descriptions, focusing more deeply on unfamiliar or difficult topics.

I read the description on the 'Microbe' and 'Hint' cards in detail only if I had limited knowledge about that card. If I was familiar with that microbe, I just read it briefly and sometimes just the heading. [Participant 15, Group A, Female]

When I was the host, it was challenging to identify the microbes, because my microbiology knowledge wasn't that strong. Imagining just one microbe out of

many possibilities felt difficult. [Participant 32, Group B, Male]

Personal Preferences

Personal preferences toward the game can be considered as having a key role in shaping their engagement and emotional responses. Preferences for specific roles were likely to be influenced by individual perceptions of stress, decision-making enjoyment, and familiarity with game-based learning. Those who preferred less cognitive demand tended to favor the microbe role, while others were more motivated by the host role's active decision-making component. Their preference for card games also contributed to positive attitudes, enhancing their overall willingness to engage and learn.

I prefer learning in a more relaxed setting, so playing as the microbe suited me, as it didn't require as much intense thinking. [Participant 12, Group A, Male]

I enjoy making decisions and being in control, so the host role felt more engaging and aligned with how I like to learn. [Participant 31, Group B, Female]

I've always loved card games, so learning microbiology and immunology through one just made sense to me. It didn't feel like studying, as I was having fun, but I could tell I was actually learning something useful at the same time. [Participant 32, Group B, Male]

Theme 2: Learning Process

Participants demonstrated differing patterns of engagement depending on whether they began the game as the microbe or the host, with each sequence influencing how they processed and applied microbiology and immunology content.

Host Before Microbe

Participants who began as the host described the experience as cognitively demanding but motivating. Without access to structured content at the outset, learners were required to draw on prior knowledge, make inferences, and formulate hypotheses, which align with higher-order thinking. This initial challenge appeared to increase engagement and stimulate curiosity.

I liked being the host first—it was challenging because I had no idea which microbes were involved. I had to gather information from both sides and think it through before attacking. [Participant 32, Group B, Male]

Guessing what the microbe was felt exciting. I was genuinely happy when I got it. [Participant 28, Group B, Female]

In subsequent rounds, these learners reported reflecting on their earlier performance, identifying knowledge gaps, and actively revisiting content to improve.

When I became the microbe, I went back and reviewed the Matching Key to see if I had used the right ‘Protection’ card before. I also wanted to know what other protections were possible because I had made mistakes in the last round. [Participant 36, Group B, Female]

Microbe Before Host

Students who started as the microbe described this role as a preparatory stage, providing them with a broad overview of key concepts. This role encouraged exposure to content in a lower-pressure context, which later supported more confident and efficient decision-making when switching to the host role. The structure of the game allowed them to gradually transition from recognition-based learning to recall and application.

Playing as the microbe first gave me an advantage. It helped me remember the content for each microbe. So, when I switched to the host role, I could recall those features more quickly. [Participant 15, Group A, Female]

Some participants also reported using short-term memorization techniques during the microbe phase, which helped build a basic conceptual scaffold for identifying pathogens in the host round. These findings suggest that beginning as the microbe may support initial content familiarization, enabling learners to make more informed and confident decisions when transitioning to the more cognitively demanding host role.

Since I started as the microbe, I tried to memorize the Matching Key. It helped me get a rough idea of which microbes stood out, so when I switched to host, I could answer based on what I remembered from the last round. [Participant 5, Group A, Female]

Theme 3: Game Design

The design of the game played an important role in shaping both the cognitive and emotional aspects of learning. Participants reflected on how the game’s structure, feedback mechanisms, visuals, and interactive elements supported their engagement and understanding.

Pedagogical Components

Learning Content

Participants reported they could learn about microbes and immunology from the information on the cards, whether by reviewing their own cards or observing those played by their opponents. They also noted that repeatedly experiencing the cards during gameplay deepened their understanding and helped them retain key concepts.

The cards had clear explanations. Sometimes I learned more from looking at my opponent’s cards. [Participant 14, Group A, Female]

Level of Difficulty

Participants generally perceived the level of difficulty as appropriate for their academic background. The content was challenging enough to stimulate interest, but it was not too complex to create frustration. This balance helped sustain motivation throughout gameplay. However, some noted that repeated play led to familiarity with answers, reducing difficulty over time.

The content was just the right level of difficulty; it wasn’t too hard or too easy. [Participant 8, Group A, Female]

The game was fun, but when I started to remember the answers, it became easier, so I’d love to see more variety in the microbes to keep it challenging. [Participant 36, Group B, Female]

Feedback Mechanism

The review phase where players checked their “Protection” cards against the correct answers was consistently mentioned as a key learning moment. This immediate feedback helped participants reflect on their decisions, correct misunderstandings, and reinforce correct associations between microbes and immune responses.

I learned a lot from checking whether the Host’s ‘Protection’ card matched the microbe. Even if it was wrong, I still got to learn from that mistake. [Participant 13, Group A, Male]

Entertaining Components

Visual Engagement

The visual design of the game, including colorful illustrations and a playful layout, added to its appeal and enhanced the learning atmosphere. Many participants described the cards as cute or attractive, which positively influenced their willingness to engage.

The game design was really well done. The pictures were cute and made the whole experience more inviting. [Participant 21, Group B, Female]

Active Decision-Making Elements

Gameplay was structured to allow active involvement from both players. While the host engaged in decision-making, the microbe observed and predicted responses. This dynamic encouraged deeper analysis and reflection on the game's logic and strategy.

Watching my friend play gave me time to see how the game worked overall. I was thinking about the clues they used and whether the microbe I had in mind matched what they played. [Participant 4, Group A, Female]

Social Interaction

Interactions between players, including discussion, guessing, and bluffing, contributed to enjoyment and emotional investment. Participants appreciated the playful tension between roles, which made learning more immersive and memorable.

I liked playing as the microbe because I could watch my friend guess and have fun trying to throw them off with my clues. I didn't enjoy guessing as much, but I loved being the one giving hints. [Participant 10, Group A, Female]

Theme 4: Learning Setting

The physical and contextual conditions under which gameplay occurred played a notable role in shaping students' comfort, focus, and willingness to engage with the learning material. Participants appreciated the autonomy offered by the flexible setting and timing, as well as the availability of technical support when needed.

Learning Atmosphere

Allowing participants to choose their preferred learning environment fostered a greater sense of comfort and autonomy. Many students reported feeling less anxious and more willing to take intellectual risks when playing outside a formal

Table . Frequency of reported learning outcomes achieved from the game.

Learning outcomes	Frequency
Knowledge recall	40
Knowledge acquisition	34
Critical thinking	30
Strategic thinking	26
Knowledge application	22

Knowledge Recall

As students alternated roles and revisited familiar content, their ability to retrieve specific microbial characteristics improved. The act of seeing clues repeatedly triggered memory recognition, reinforcing their confidence and retention.

Seeing the clues helped trigger my memory when I looked at the 'Key' card, the name of the microbe felt

classroom setting. The absence of an authoritative figure reduced pressure and created space for experimentation and open guessing.

I liked that we could choose where to play ... even if I wasn't sure about an answer, I felt okay taking a guess without worrying that someone would judge me for getting it wrong. [Participant 22, Group B, Male]

Flexibility in Learning Schedule

The flexibility to decide when to engage with the game contributed to a more focused and self-directed learning experience. Participants appreciated that they could play when they felt most ready, which made learning feel more like a voluntary activity than a scheduled obligation.

The game didn't force us to play at a specific time. I could just pick it up whenever I wanted, and that made it feel like something I chose to do ... not something I was made to do. [Participant 15, Group A, Female]

Real-Time Gameplay support

Access to facilitators or staff members during gameplay reassured participants and minimized concerns about game mechanics. This allowed them to devote more attention to the educational content rather than the rules. The ability to clarify doubts in real-time improved confidence and supported sustained engagement.

Having staff there to answer questions made me less worried about the rules, so I could actually focus on the content and not get stuck on how to play. [Participant 27, Group B, Female]

Theme 5: Learning Outcomes

Gameplay provided opportunities for cognitive engagement across multiple dimensions, from foundational knowledge acquisition to higher-order reasoning (Table 8). Participants described improvements in knowledge, critical and strategic thinking, and conceptual understanding, highlighting the game's educational value beyond passive content delivery.

familiar. And when it was revealed, it confirmed that I remembered it right. [Participant 22, Group B, Male]

Knowledge Acquisition

The game facilitated the acquisition of microbiology and immunology knowledge through its structured and interactive format. Repeated exposure to microbial features and host responses during gameplay helped build a stronger conceptual

foundation. Many participants reported gaining new information simply by observing their opponent's moves.

Reading the details on the cards played by my opponent gave me new information and helped me figure out the microbe. [Participant 8, Group A, Female]

Critical Thinking

Participants engaged in critical thinking by interpreting incomplete information and making evidence-based guesses. Particularly in the host role, students described the need to synthesize clues and evaluate possible answers, moving beyond memorization to diagnostic reasoning.

When I played as the host, I had no idea what the microbe was at first. I had to slowly piece things together using the information I had. [Participant 17, Group A, Male]

Strategic Thinking

The game structure encouraged participants to develop and adjust strategies based on observations of their opponent. This adaptive thinking extended beyond factual recall to prediction, planning, and real-time decision-making.

My opponent rarely played 'Protection' cards. He used a different strategy, like discarding cards to draw a keyword because it gave them a better advantage. [Participant 18, Group A, Female]

Knowledge Application

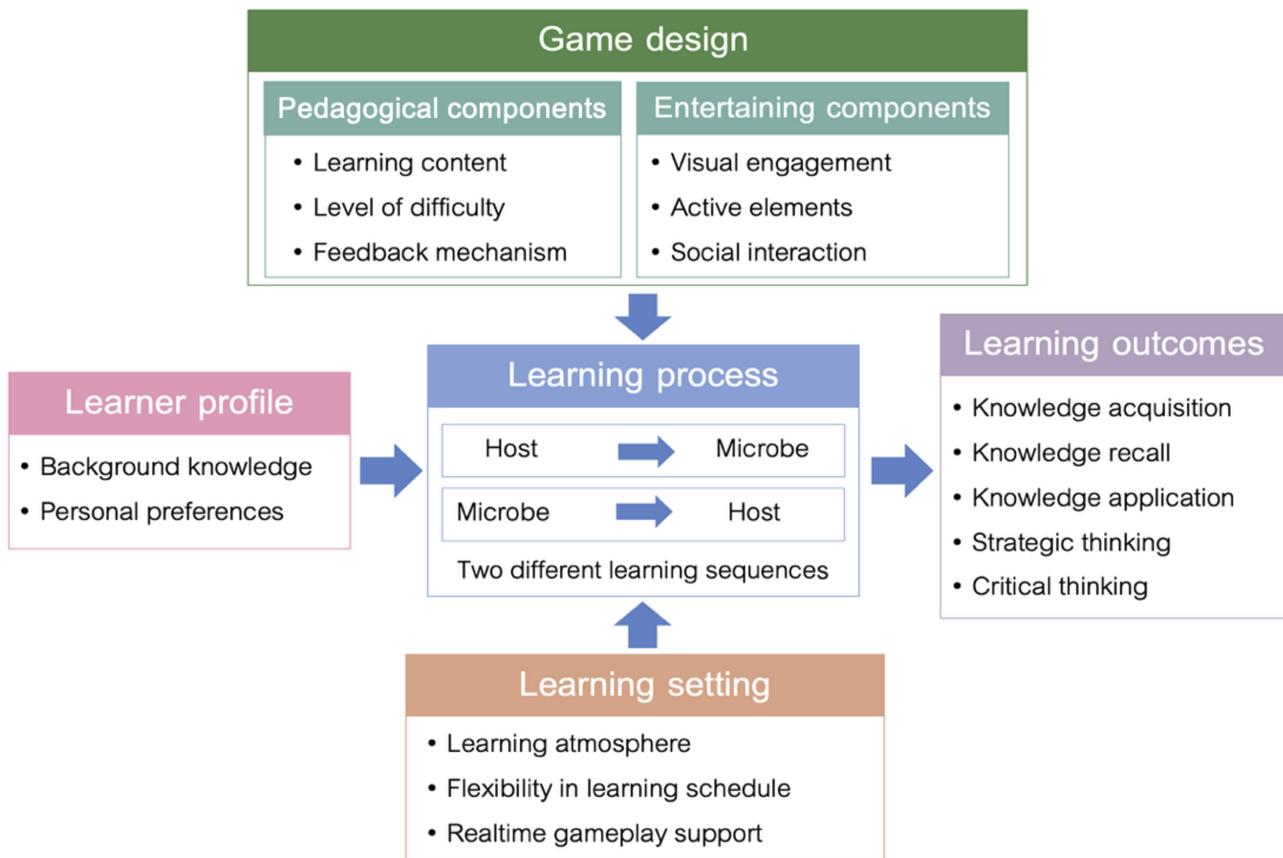
Participants not only recalled learning content but also applied their knowledge during gameplay. They demonstrated an ability to connect microbiological information to real-life

decision-making, such as selecting appropriate treatments or interventions.

I learned which medications go with which types of microbes. [Participant 9, Group A, Female]

Conceptual Framework for the Implementation of the Game in Dental Education

This conceptual framework explains how the emerging themes show the connections between the core elements of game design, learner characteristics, and the learning setting in supporting students to achieve expected learning outcomes (Figure 4). The game should be designed with a balance between pedagogical and entertaining components, supporting students in understanding key concepts while maintaining their motivation throughout the learning process. The learning process is at the center of the framework, where students move between playing as the host and the microbe. By switching roles, students encounter different kinds of challenges, which encourages them to think in new ways and gradually build their knowledge through the game. The game's rules allow for repeated practice and reflection, helping students remember what they learn and apply it in various situations. The framework highlights the importance of learners' background knowledge and personal preferences, as well as the influence of the learning settings. Factors such as the learning atmosphere, flexible scheduling, and support during gameplay all play a role in shaping the learning experience. These elements work together with the game design to affect how students participate, interact with each other, and respond to challenges in the game. When these factors are well aligned, the game not only helps students understand the subject content but also supports critical thinking and strategic decision-making skills.

Figure 4. Conceptual framework for the implementation of a dual-role duel card game.

Discussion

Effectiveness of Game-Based Learning on Knowledge Enhancement

The “Invasion” game demonstrated its potential as a supplementary learning tool for enhancing knowledge in oral microbiology and immunology. The progressive improvement in scores from Pretest to Posttest 1, and subsequently to Posttest 2, suggests that repeated gameplay contributed to knowledge enhancement. As students moved through the rounds, they were exposed to more diverse sets of microbes, with some organisms appearing more than once. This repetition likely reinforced recall, aligning with established principles of spaced learning and retrieval practice [29]. Participants reported varied learning experiences; some were able to recall prior coursework, others acquired new information, particularly related to antimicrobial treatments, and many developed a broader understanding of how microbiological knowledge applies within clinical contexts. The game was reported by participants to encourage strategic thinking and decision-making skills, which are important for developing clinical reasoning [30-32]. Participants perceived that the dual-role structure of the game, alternating between structured observation and active problem-solving, provided different learning experiences. Students learned not only through success but also through failure perspectives [33,34]. Lost heart points from incorrect guesses can be considered as immediate feedback, prompting them to reflect and adjust their approach. This process reflects the role of failure theory, where knowledge is built through doing, failing, and refining strategies. These findings align with previous research [21,23,28], showing that

game-based learning can enhance understanding through engagement, feedback, and active participation.

Impact of Role Sequence on Cognitive Processing

Although knowledge improvement between groups did not differ significantly, qualitative findings suggest that the sequence in which students engaged with gameplay roles experienced the learning process differently. Students who began as the microbe interacted with the material in a more observational manner, generating clues and observing how their partner responded. This role offered an opportunity to reflect on microbial characteristics without the immediate demand for decision-making. This process aligns with inductive learning, where knowledge is built from observed patterns and structured clues. In contrast, those who started as the host were required to analyze clues and make diagnostic judgments from the beginning, enhancing deeper cognitive processing [35,36]. This approach reflects deductive reasoning, in which learners begin with a hypothesis and test it through gameplay. The effectiveness of each sequence may depend on individual learner traits such as cognitive readiness, prior knowledge, or preferred learning approach [37,38]. While some students appeared to benefit from first observing how the content functioned in context, others were more engaged when challenged to act at an early stage. Across both pathways, students reported cognitive improvement through the interactive and engaging environment, highlighting the importance of role variation in supporting diverse learning styles.

Influence of Learner Diversity on Gameplay Experiences

In addition to role assignment, other factors also influenced how students experienced the game and developed cognitive skills through play. Background knowledge and prior experience of learners appear to be important factors in shaping their engagement with the game [28]. Those with a stronger background knowledge tended to participate more confidently throughout the activity [39]. It is possible that students who felt less familiar with the content were more likely to prefer the microbe role, where they could rely more on the structured clues provided by the cards. In addition, individual learning preferences can shape students' approach to gameplay [40-42]; some preferred roles that felt less cognitively demanding, while others were more engaged by strategic decision-making and competitive challenge. The learning context, including peer interaction and minimal technical barriers, can further support student engagement with the game tasks [20,21]. These findings highlight the importance of designing educational games that incorporate role variation and provide flexibility to support learners across a range of backgrounds and preferences.

Pedagogical Implications and Curriculum Integration

The "Invasion" game was developed to align with the pedagogical goals of dental education by accommodating diverse learning preferences and cognitive styles. Its dual-role format allows students to alternate between reflective observation as the microbe and active problem-solving as the host, supporting both inductive and deductive reasoning [43,44]. This structured variation enables learners to engage with microbiology and immunology content through multiple perspectives, promoting deeper understanding and flexible thinking. Within the dental curriculum, the "Invasion" game can be considered a supplemental learning tool alongside didactic lectures to reinforce basic biomedical knowledge, similar to other game-based strategies successfully applied in preclinical dental education [21]. This is consistent with the argument that game-based learning should not be implemented as a mandatory learning method for learners [45]. The adaptable structure of the "Invasion" game allows flexible integration into tutorials, self-directed study, or flipped classrooms, while its emphasis on peer interaction and strategic decision-making fosters motivation, collaboration, and active learning. Beyond dental education, the dual-role format could also be applied in other

health education contexts, such as medical, nursing, and pharmacy curricula, where collaborative learning and clinical decision-making are equally essential, although the present content is more directly aligned with oral microbiology and immunology.

Limitations and Recommendations for Future Research

While this mixed methods study provided a comprehensive integration of quantitative and qualitative insights into student learning, certain limitations should be acknowledged. The study was conducted at a single institution with a relatively small sample size, which may limit the generalizability of the findings to broader educational contexts. While the test items were randomized in sequence, the repeated use of the same items across all assessment points may have introduced a memory effect, enhancing posttest performance. Future studies could minimize this issue by using different but equivalent test items for pre- and postassessments, supported by analyses of item difficulty and parallel forms reliability. The short duration of the intervention did not allow for evaluation of long-term retention or application to real-world clinical tasks. Future research should incorporate diverse settings, larger cohorts, varied assessment formats, multiple gameplay sessions over time, and delayed postknowledge assessments to better evaluate the long-term educational impact of serious games.

Conclusions

This study demonstrated that a dual-role educational card game can effectively support diverse learning preferences in oral microbiology and immunology by offering both structured observation and active problem-solving. The game not only enhanced knowledge acquisition but was also perceived by participants to foster strategic thinking and learner engagement through its dynamic role-based format. By alternating between the roles of microbe and host, students engaged with content in different ways, enabling both inductive and deductive learning processes. These findings highlight the value of incorporating flexible, learner-centered tools into the dental curriculum to bridge theoretical science and clinical reasoning. Future research is recommended to investigate the adaptability of this game in broader health education contexts and to evaluate long-term knowledge retention through repeated gameplay sessions over time and follow-up assessments.

Acknowledgments

We would like to express our appreciation to Dr Sophita Suwutto, Chairman of the Medical Education Technology Center, Faculty of Medicine, Siriraj Hospital, Mahidol University, for her helpful guidance in the development of the game. Furthermore, we are thankful to Dr Khanathat Nuntachurat, Dr Alana Khorana, Dr Prarthana Ahuja, and Dr Apitchaya Saetung for their contribution to the pilot testing of the game concept and to the experts who offered their insightful suggestions during the validation process of the data collection tools. Finally, we acknowledge all research participants for their valuable contributions to this study.

The authors used ChatGPT-5 (OpenAI) to assist with grammar checking and language refinement during manuscript preparation, including minor linguistic edits of interview quotes to improve readability. All artificial intelligence-assisted edits were critically reviewed by the authors to ensure accuracy, preserve the original meanings of quotes, and maintain the integrity of the manuscript.

Funding

This research project was supported by Mahidol University.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to information that could compromise the privacy of research participants but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: KS, RLS

Data curation: KS, RLS, PC, KJ, JS, CS, TT

Formal analysis: KS, RLS, PC, KJ, JS, CS, TT

Investigation: KS, PC, KJ, JS, CS, TT

Methodology: KS, RLS

Project administration: KS

Resources: KS, RLS

Validation: KS, RLS

Visualization: KS, RLS, PC, KJ, JS, CS, TT, SP

Writing – original draft: KS, RLS

Writing – review & editing: KS, RLS, PC, KJ, JS, CS, TT, SP

Conflicts of Interest

The authors declare no conflicts of interest. Although the "Invasion" game has been designed and developed by the research team, its primary purpose is exclusively for academic purposes, with no commercial objectives associated with it.

References

- Peterson LR, Hamilton JD, Baron EJ, et al. Role of clinical microbiology laboratories in the management and control of infectious diseases and the delivery of health care. *Clin Infect Dis* 2001 Feb 15;32(4):605-611. [doi: [10.1086/318725](https://doi.org/10.1086/318725)] [Medline: [11181125](https://pubmed.ncbi.nlm.nih.gov/11181125/)]
- Cox JL, Simpson MD, Letts W, Cavanagh HM. Re-thinking microbiology/infection control education to enhance the practice-readiness of health professional students: More than just a curriculum issue. *J Learn Des* 2015 Mar;8(1):55-67. [doi: [10.5204/jld.v8i1.224](https://doi.org/10.5204/jld.v8i1.224)]
- Soni J, Sinha S, Pandey R. Understanding bacterial pathogenicity: a closer look at the journey of harmful microbes. *Front Microbiol* 2024;15:1370818. [doi: [10.3389/fmicb.2024.1370818](https://doi.org/10.3389/fmicb.2024.1370818)] [Medline: [38444801](https://pubmed.ncbi.nlm.nih.gov/38444801/)]
- Banakar M, Bagheri Lankarani K, Jafarpour D, Moayedi S, Banakar MH, MohammadSadeghi A. COVID-19 transmission risk and protective protocols in dentistry: a systematic review. *BMC Oral Health* 2020 Oct 8;20(1):275. [doi: [10.1186/s12903-020-01270-9](https://doi.org/10.1186/s12903-020-01270-9)] [Medline: [33032593](https://pubmed.ncbi.nlm.nih.gov/33032593/)]
- Rahman S, Montero MTV, Rowe K, Kirton R, Kunik F Jr. Epidemiology, pathogenesis, clinical presentations, diagnosis and treatment of COVID-19: a review of current evidence. *Expert Rev Clin Pharmacol* 2021 May;14(5):601-621. [doi: [10.1080/17512433.2021.1902303](https://doi.org/10.1080/17512433.2021.1902303)] [Medline: [33705239](https://pubmed.ncbi.nlm.nih.gov/33705239/)]
- Prince M. Does active learning work? A review of the research. *J Eng Edu* 2004 Jul;93(3):223-231. [doi: [10.1002/j.2168-9830.2004.tb00809.x](https://doi.org/10.1002/j.2168-9830.2004.tb00809.x)]
- Dong JJ, Hwang WY, Shadiev R, Chen GY. Pausing the classroom lecture: the use of clickers to facilitate student engagement. *Act Learn High Educ* 2017 Jul;18(2):157-172. [doi: [10.1177/1469787417707617](https://doi.org/10.1177/1469787417707617)]
- Schnitzler K, Holzberger D, Seidel T. All better than being disengaged: student engagement patterns and their relations to academic self-concept and achievement. *Eur J Psychol Educ* 2021 Sep;36(3):627-652. [doi: [10.1007/s10212-020-00500-6](https://doi.org/10.1007/s10212-020-00500-6)]
- Anokye Effah NA, Nkwantabisa AO. The influence of academic engagement on academic performance of university accounting students in Ghana. *S Afr J Account Res* 2022 May 4;36(2):105-122. [doi: [10.1080/10291954.2021.1988204](https://doi.org/10.1080/10291954.2021.1988204)]
- Acosta-Gonzaga E. The effects of self-esteem and academic engagement on university students' performance. *Behav Sci (Basel)* 2023 Apr 21;13(4):348. [doi: [10.3390/bs13040348](https://doi.org/10.3390/bs13040348)] [Medline: [37102862](https://pubmed.ncbi.nlm.nih.gov/37102862/)]
- Munna AS, Kalam MA. Impact of active learning strategy on the student engagement. *Gnosi: Interdiscip J Human Theol Praxis* 2021 Apr;4(2):96-114 [FREE Full text]
- Venton BJ, Pompano RR. Strategies for enhancing remote student engagement through active learning. *Anal Bioanal Chem* 2021 Mar;413(6):1507-1512. [doi: [10.1007/s00216-021-03159-0](https://doi.org/10.1007/s00216-021-03159-0)] [Medline: [33458779](https://pubmed.ncbi.nlm.nih.gov/33458779/)]
- Rohani N, Sowa S, Manataki A. Identifying learning preferences and strategies in health data science courses: systematic review. *JMIR Med Educ* 2024 Aug 12;10:e50667. [doi: [10.2196/50667](https://doi.org/10.2196/50667)] [Medline: [39133909](https://pubmed.ncbi.nlm.nih.gov/39133909/)]
- Gentry SV, Gauthier A, L'Estrade Ehrstrom B, et al. Serious gaming and gamification education in health professions: systematic review. *J Med Internet Res* 2019 Mar 28;21(3):e12994. [doi: [10.2196/12994](https://doi.org/10.2196/12994)] [Medline: [30920375](https://pubmed.ncbi.nlm.nih.gov/30920375/)]

15. Xu M, Luo Y, Zhang Y, Xia R, Qian H, Zou X. Game-based learning in medical education. *Front Public Health* 2023;11:1113682. [doi: [10.3389/fpubh.2023.1113682](https://doi.org/10.3389/fpubh.2023.1113682)] [Medline: [36935696](https://pubmed.ncbi.nlm.nih.gov/36935696/)]
16. Sipiyaruk K, Reynolds PA, Nakornnoi T, Santiwong P, Chintavalakorn R. Game-based learning in orthodontic education: a systematic review. *BDJ Open* 2024 Jul 4;10(1):56. [doi: [10.1038/s41405-024-00218-3](https://doi.org/10.1038/s41405-024-00218-3)] [Medline: [38965234](https://pubmed.ncbi.nlm.nih.gov/38965234/)]
17. Zaror C, Mariño R, Atala-Acevedo C. Current state of serious games in dentistry: a scoping review. *Games Health J* 2021 Apr;10(2):95-108. [doi: [10.1089/g4h.2020.0042](https://doi.org/10.1089/g4h.2020.0042)] [Medline: [33818135](https://pubmed.ncbi.nlm.nih.gov/33818135/)]
18. Efthimiou G, Tucker NP. Microbes Against Humanity, a workshop game for horrible students: using a creative card game in higher education microbiology teaching. *Access Microbiol* 2021 Feb;3(2):000186. [doi: [10.1099/acmi.0.000186](https://doi.org/10.1099/acmi.0.000186)] [Medline: [34151143](https://pubmed.ncbi.nlm.nih.gov/34151143/)]
19. Imwattana K, Methawirune A, Voralakkulchai I, Ngamskulrungroj P. BactoBattle: a game-based learning companion for medical bacteriology. *Access Microbiol* 2023;5(6):acmi000608.v3. [doi: [10.1099/acmi.0.000608.v3](https://doi.org/10.1099/acmi.0.000608.v3)] [Medline: [37424568](https://pubmed.ncbi.nlm.nih.gov/37424568/)]
20. Sipiyaruk K, Hatzipanagos S, Vichayanrat T, Reynolds PA, Gallagher JE. Evaluating a dental public-health game across two learning contexts. *Educ Sci* 2022;12(8):517. [doi: [10.3390/educsci12080517](https://doi.org/10.3390/educsci12080517)]
21. Arayapisit T, Pojmonpiti D, Dansirisomboon K, Jitverananrangsri K, Poosontipong D, Sipiyaruk K. An educational board game for learning orofacial spaces: an experimental study comparing collaborative and competitive approaches. *Anat Sci Educ* 2023;16(4):666-676. [doi: [10.1002/ase.2257](https://doi.org/10.1002/ase.2257)] [Medline: [36695535](https://pubmed.ncbi.nlm.nih.gov/36695535/)]
22. Buajeeb W, Chokpipatkul J, Achalanon N, Kriwattanawong N, Sipiyaruk K. The development of an online serious game for oral diagnosis and treatment planning: evaluation of knowledge acquisition and retention. *BMC Med Educ* 2023 Nov 3;23(1):830. [doi: [10.1186/s12909-023-04789-x](https://doi.org/10.1186/s12909-023-04789-x)] [Medline: [37924052](https://pubmed.ncbi.nlm.nih.gov/37924052/)]
23. Buajeeb W, Reynolds PA, Boontub H, Tangmanpuwadol Y, Sipiyaruk K. Comparison of the effectiveness of a serious game and pre-recorded lecture in diagnosis and treatment planning of oral lesions for dental students. *Sci Rep* 2024 Dec 27;14(1):30641. [doi: [10.1038/s41598-024-83433-0](https://doi.org/10.1038/s41598-024-83433-0)] [Medline: [39730901](https://pubmed.ncbi.nlm.nih.gov/39730901/)]
24. Chintavalakorn R, Chanwanichkulchai C, Buranasing N, Parivisutt N, Patchasri N, Sipiyaruk K. Development and evaluation of a case-based serious game for diagnosis and treatment planning in orthodontic education: quasi-experimental study. *JMIR Serious Games* 2025 Aug 27;13:e73956. [doi: [10.2196/73956](https://doi.org/10.2196/73956)] [Medline: [40864887](https://pubmed.ncbi.nlm.nih.gov/40864887/)]
25. Fernandez-Garcia I, Riaza P, Arias A. Game-based learning: a pilot educational strategy to foster the learning of endodontic terminology. *Int Endod J* 2025 Jul 18;n/a(n/a). [doi: [10.1111/iej.70001](https://doi.org/10.1111/iej.70001)] [Medline: [40682132](https://pubmed.ncbi.nlm.nih.gov/40682132/)]
26. Wu JH, Su PH, Wu HY, Hsin YM, Lin CH, Lee CY. Educational board game for training dental and dental hygiene students in patient safety issues. *BMC Med Educ* 2025 Apr 11;25(1):518. [doi: [10.1186/s12909-025-07115-9](https://doi.org/10.1186/s12909-025-07115-9)] [Medline: [40217221](https://pubmed.ncbi.nlm.nih.gov/40217221/)]
27. Peanchitlertkajorn S, Reynolds PA, Chaisawas M, Krungthong Y, Boonkaew B, Sipiyaruk K. The impact of an online gamified virtual tour on cognitive enhancement in dental practice management. *Sci Rep* 2024 Nov 6;14(1):26975. [doi: [10.1038/s41598-024-75128-3](https://doi.org/10.1038/s41598-024-75128-3)] [Medline: [39505891](https://pubmed.ncbi.nlm.nih.gov/39505891/)]
28. Teerawongpairoj C, Tantipoj C, Sipiyaruk K. The design and evaluation of gamified online role-play as a telehealth training strategy in dental education: an explanatory sequential mixed-methods study. *Sci Rep* 2024 Apr 22;14(1):9216. [doi: [10.1038/s41598-024-58425-9](https://doi.org/10.1038/s41598-024-58425-9)] [Medline: [38649451](https://pubmed.ncbi.nlm.nih.gov/38649451/)]
29. Wiklund-Hörnqvist C, Jonsson B, Nyberg L. Strengthening concept learning by repeated testing. *Scand J Psychol* 2014 Feb;55(1):10-16. [doi: [10.1111/sjop.12093](https://doi.org/10.1111/sjop.12093)] [Medline: [24313425](https://pubmed.ncbi.nlm.nih.gov/24313425/)]
30. Chon SH, Timmermann F, Dratsch T, et al. Serious games in surgical medical education: a virtual emergency department as a tool for teaching clinical reasoning to medical students. *JMIR Serious Games* 2019 Mar 5;7(1):e13028. [doi: [10.2196/13028](https://doi.org/10.2196/13028)] [Medline: [30835239](https://pubmed.ncbi.nlm.nih.gov/30835239/)]
31. Wong JYH, Ko J, Nam S, et al. Virtual ER, a serious game for interprofessional education to enhance teamwork in medical and nursing undergraduates: development and evaluation study. *JMIR Serious Games* 2022 Jul 14;10(3):e35269. [doi: [10.2196/35269](https://doi.org/10.2196/35269)] [Medline: [35834309](https://pubmed.ncbi.nlm.nih.gov/35834309/)]
32. Koelewijn G, Hennus MP, Kort HSM, Frenkel J, van Houwelingen T. Games to support teaching clinical reasoning in health professions education: a scoping review. *Med Educ Online* 2024 Dec 31;29(1):2316971. [doi: [10.1080/10872981.2024.2316971](https://doi.org/10.1080/10872981.2024.2316971)] [Medline: [38394053](https://pubmed.ncbi.nlm.nih.gov/38394053/)]
33. Anderson CG, Dalsen J, Kumar V, Berland M, Steinkuehler C. Failing up: how failure in a game environment promotes learning through discourse. *Think Skills Creat* 2018 Dec;30:135-144. [doi: [10.1016/j.tsc.2018.03.002](https://doi.org/10.1016/j.tsc.2018.03.002)]
34. Hefkaluk N, Linehan C, Trace A. Fail, fail again, fail better: how players who enjoy challenging games persist after failure in "Celeste". *Int J Hum Comput Stud* 2024 Mar;183:103199. [doi: [10.1016/j.ijhcs.2023.103199](https://doi.org/10.1016/j.ijhcs.2023.103199)]
35. Rossi IV, de Lima JD, Sabatke B, Nunes MAF, Ramirez GE, Ramirez MI. Active learning tools improve the learning outcomes, scientific attitude, and critical thinking in higher education: experiences in an online course during the COVID-19 pandemic. *Biochem Mol Biol Educ* 2021 Nov;49(6):888-903. [doi: [10.1002/bmb.21574](https://doi.org/10.1002/bmb.21574)] [Medline: [34652877](https://pubmed.ncbi.nlm.nih.gov/34652877/)]
36. Ghezzi J, Higa EFR, Lemes MA, Marin MJS. Strategies of active learning methodologies in nursing education: an integrative literature review. *Rev Bras Enferm* 2021;74(1):e20200130. [doi: [10.1590/0034-7167-2020-0130](https://doi.org/10.1590/0034-7167-2020-0130)] [Medline: [33787786](https://pubmed.ncbi.nlm.nih.gov/33787786/)]
37. Gu X, Tan Y, Wu X, et al. Active versus passive strategy in online creativity training: how to best promote creativity of students with different cognitive styles? *Think Skills Creat* 2022 Jun;44:101021. [doi: [10.1016/j.tsc.2022.101021](https://doi.org/10.1016/j.tsc.2022.101021)]
38. Aga K. Comfort in active learning spaces – students' perceptions and preferences. *Eur J Engi Educ* 2024 Jul 3;49(4):785-806. [doi: [10.1080/03043797.2024.2341756](https://doi.org/10.1080/03043797.2024.2341756)]

39. Shukla J, Singh RM. Knowledge exploration among students: role of feedback, feeling of confidence, and academic motivation. *Metacogn Learn* 2024 Dec;19(3):1137-1159. [doi: [10.1007/s11409-024-09404-y](https://doi.org/10.1007/s11409-024-09404-y)]
40. Khenissi MA, Essalmi F, Jemni M, Graf S, Chen NS. Relationship between learning styles and genres of games. *Comput Educ* 2016 Oct;101:1-14. [doi: [10.1016/j.compedu.2016.05.005](https://doi.org/10.1016/j.compedu.2016.05.005)]
41. Bontchev BP, Terzieva V, Paunova-Hubenova E. Personalization of serious games for learning. *Interact Technol Smart Educ* 2021 May 19;18(1):50-68. [doi: [10.1108/ITSE-05-2020-0069](https://doi.org/10.1108/ITSE-05-2020-0069)]
42. Hwang GJ, Sung HY, Hung CM, Huang I, Tsai CC. Development of a personalized educational computer game based on students' learning styles. *Education Tech Research Dev* 2012 Aug;60(4):623-638. [doi: [10.1007/s11423-012-9241-x](https://doi.org/10.1007/s11423-012-9241-x)]
43. Prince MJ, Felder RM. Inductive teaching and learning methods: definitions, comparisons, and research bases. *J Eng Educ* 2006 Apr;95(2):123-138 [FREE Full text] [doi: [10.1002/j.2168-9830.2006.tb00884.x](https://doi.org/10.1002/j.2168-9830.2006.tb00884.x)]
44. Pashler H, McDaniel M, Rohrer D, Bjork R. Learning styles: concepts and evidence. *Psychol Sci Publ Int* 2008 Dec;9(3):105-119. [doi: [10.1111/j.1539-6053.2009.01038.x](https://doi.org/10.1111/j.1539-6053.2009.01038.x)] [Medline: [26162104](https://pubmed.ncbi.nlm.nih.gov/19033000/)]
45. Van Gaalen AEJ, Jaarsma ADC, Georgiadis JR. Medical students' perceptions of play and learning: qualitative study with focus groups and thematic analysis. *JMIR Serious Games* 2021 Jul 28;9(3):e25637. [doi: [10.2196/25637](https://doi.org/10.2196/25637)] [Medline: [34319237](https://pubmed.ncbi.nlm.nih.gov/34319237/)]

Abbreviations

IOC: The Index of Item-Objective Congruence

Edited by J Gentges; submitted 03.Jul.2025; peer-reviewed by A Allam, B Bunt, S Mukhida; accepted 09.Oct.2025; published 21.Jan.2026.

Please cite as:

Sipiyaruk K, Srisatjaluk RL, Chaisiri P, Jantawatchai K, Srihatai J, Sonyaem C, Traikorakot T, Pookboonmee S. *Fostering Inductive and Deductive Learning in Oral Microbiology and Immunology With a Dual-Role Duel Card Game: Explanatory Sequential Mixed Methods Study*. *JMIR Med Educ* 2026;12:e80048

URL: <https://mededu.jmir.org/2026/1/e80048>

doi:[10.2196/80048](https://doi.org/10.2196/80048)

© Kawin Sipiyaruk, Ratchapin Laovanitch Srisatjaluk, Ponpentip Chaisiri, Kanokwan Jantawatchai, Jidapha Srihatai, Chutikarn Sonyaem, Thanaphat Traikorakot, Suphasuta Pookboonmee. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 21.Jan.2026. 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.

Data Science Education for Residents, Researchers, and Students in Psychiatry and Psychology: Program Development and Evaluation Study

Hayoung K Donnelly^{1,2}, PhD; David Mandell¹, SCD; Sy Hwang², MS; Emily Schriver^{2,3}, MS; Ugurcan Vurgun², PhD; Graydon Neill⁴, BS; Esha Patel¹, BS; Megan E Reilly¹, MPH; Michael Steinberg¹, MA; Amber Calloway¹, PhD; Robert Gallop¹, PhD; Maria A Oquendo¹, MD, PhD; Gregory K Brown¹, PhD; Danielle L Mowery^{2,5}, PhD

¹Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

²Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

³University of Pennsylvania Health System, University of Pennsylvania, Philadelphia, PA, United States

⁴PennDNA, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

⁵Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, 3700 Hamilton Walk, Philadelphia, PA, United States

Corresponding Author:

Danielle L Mowery, PhD

Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Abstract

Background: The use of artificial intelligence (AI) to analyze health care data has become common in behavioral health sciences. However, the lack of training opportunities for mental health professionals limits clinicians' ability to adopt AI in clinical settings. AI education is essential for trainees, equipping them with the literacy needed to implement AI tools in practice, collaborate effectively with data scientists, and develop skills as interdisciplinary researchers with computing skills.

Objective: As part of the Penn Innovation in Suicide Prevention Implementation Research Center, we developed, implemented, and evaluated a virtual workshop to educate psychiatry and psychology trainees on using AI for suicide prevention research.

Methods: The workshop introduced trainees to natural language processing (NLP) concepts and Python coding skills using Jupyter notebooks within a secure Microsoft Azure Databricks cloud computing and analytics environment. We designed a 3-hour workshop that covered 4 key NLP topics: data characterization, data standardization, concept extraction, and statistical analysis. To demonstrate real-world applications, we processed chief complaints from electronic health records to compare the prevalence of suicide-related encounters across populations by race, ethnicity, and age. Training materials were developed based on standard NLP techniques and domain-specific tasks, such as preprocessing psychiatry-related acronyms. Two researchers drafted and demonstrated the code, incorporating feedback from the Methods Core of the Innovation in Suicide Prevention Implementation Research to refine the materials. To evaluate the effectiveness of the workshop, we used the Kirkpatrick program evaluation model, focusing on participants' reactions (level 1) and learning outcomes (level 2). Confidence changes in knowledge and skills before and after the workshop were assessed using paired *t* tests, and open-ended questions were included to gather feedback for future improvements.

Results: A total of 10 trainees participated in the workshop virtually, including residents, postdoctoral researchers, and graduate students from the psychiatry and psychology departments. The participants found the workshop helpful (mean 3.17 on a scale of 1 - 4, SD 0.41). Their overall confidence in NLP knowledge significantly increased ($P=.002$) from 1.35 (SD 0.47) to 2.79 (SD 0.46). Confidence in coding abilities also improved significantly ($P=.01$), increasing from 1.33 (SD 0.60) to 2.25 (SD 0.42). Open-ended feedback suggested incorporating thematic analysis and exploring additional datasets for future workshops.

Conclusions: This study illustrates the effectiveness of a tailored data science workshop for trainees in psychiatry and psychology, focusing on applying NLP techniques for suicide prevention research. The workshop significantly enhanced participants' confidence in conducting data science research. Future workshops will cover additional topics of interest, such as working with large language models, thematic analysis, diverse datasets, and multifaceted outcomes. This includes examining how participants' learning impacts their practice and research, as well as assessing knowledge and skills beyond self-reported confidence through methods such as case studies for deeper insights.

(*JMIR Med Educ* 2026;12:e75125) doi:10.2196/75125

KEYWORDS

data science; AI literacy education; artificial intelligence; psychiatry education; suicide research; natural language processing; cloud computing

Introduction

Artificial intelligence (AI) applications in behavioral medicine now span a wide range, including detecting at-risk populations, assisting clinicians in decision-making, and providing feedback to improve the quality of interventions [1]. Exposure to data science and AI education is crucial for medical trainees, as it prepares them to implement AI tools in their practice with the necessary literacy, collaborate effectively with data scientists, and explore careers as interdisciplinary researchers [2,3]. Medical trainees recognize the importance of AI and support its inclusion in their curriculum [4,5]; however, they also report low levels of knowledge and skills in addition to anxiety about using AI [5-7]. As a result, AI literacy among clinicians remains low, hindering the adoption of AI in clinical settings [8,9].

Self-learning methods, such as online articles and media resources, have been found to be the most common way for students to learn AI, with more than 70% relying on these approaches [2,10]. However, only 36% of students have exposure to formal AI training [2]. Previous studies have highlighted the need for guided and structured AI education tailored to health trainees, such as formal curricula and workshops [3,8]. Despite this need, barriers to implementing formal AI education persist, primarily due to limited time within an already demanding medical curriculum. A trainee-centered approach to structured learning opportunities may help reduce their burden. This includes integrating AI training into the existing didactic schedule [11] and adopting a human-centered design from recruitment to workshop facilitation.

Trainees in psychiatry and psychology may face even more limited opportunities for AI training tailored to their specialty and interests. Most existing AI education programs focus on general AI concepts and broad medical applications rather than addressing field-specific needs [3]. While some programs have been tailored for a specific domain such as radiology, a systematic review showed a lack of AI education programs specifically designed for psychiatry and psychology trainees [3]. Clinicians' learning of data science should be directly applicable to their area of expertise, helping them generate ideas for their research applications. For instance, AI education programs for psychiatry and psychology trainees should be designed around mental health-related questions, data sources, and AI applications to ensure meaningful translation into their research and clinical practice.

One area to consider in AI education for psychiatry and psychology trainees is the application of AI in suicide prevention. Researchers are leveraging AI to identify patients at risk for suicidal thoughts and behaviors in clinical settings and the wider population [12]. Research in this space can scan electronic health records (EHRs) to identify elevated suicide risk, improve ED suicide screening, and decrease racial disparities in suicide [12]. This underscores the importance of becoming familiar with these technologies as future clinicians.

In particular, analyzing text data and developing natural language processing (NLP)-based AI can deepen the understanding of the complexity of suicidal ideation and behaviors, allowing for the design of contextually sensitive suicide prevention strategies. For instance, clinical notes from EHR and NLP have demonstrated their strengths in enhancing suicide prevention strategies [13,14]. Developing AI education focused on suicide risk would offer trainees valuable insights into leveraging AI for both their research and practice.

Finally, developing and implementing AI education for medical trainees requires incorporating a structured evaluation process to assess its effectiveness. However, among 41 existing AI education development studies in medicine, only 5 included formal evaluation methods, such as structured surveys and assessments, whereas the remaining 36 studies solely described the education programs without measuring their impact [8]. Without rigorous evaluation, it is difficult to determine whether these programs effectively enhance knowledge and skills among trainees. One study specifically highlighted the importance of a competency-based curriculum in enhancing AI knowledge and skills, ultimately increasing the adoption of AI in clinical practice [8].

To address these gaps, we developed and implemented a 3-hour workshop to introduce psychiatry and psychology trainees to the applications of AI in suicide research. Specifically, we incorporated NLP to highlight the potential of leveraging clinical text data for psychiatry research and adopted a program evaluation framework to systematically assess the program's efficacy.

Methods

Overview

We designed, disseminated, and evaluated a 3-hour training workshop to introduce psychiatry and psychology trainees to data science and AI. The workshop was conducted using Python Jupyter notebooks, Microsoft Azure Databricks, and Zoom (Zoom Communications). We began with a brief introduction and presurvey (5 min), followed by a Databricks setup session originally planned for 15 minutes. Although participants' accounts and workspace access had been configured in advance, the setup process took approximately 30 minutes, which was 15 minutes longer than expected. A site reliability engineer was present to help manage the situation effectively; however, we needed to cancel the welcome speech by a psychiatry faculty member. The introductory remarks were intended to highlight the need for NLP research in suicide prevention. Once all participants' environments were set up, we began the first phase of the hands-on coding practice, which focused on data characterization and preprocessing (1 h 20 min). We took a 10-minute break and then continued with the second phase, which covered concept extraction and statistical analysis (1 h 20 min). The session concluded with a wrap-up that included a postsurvey and office hour sign-up (10 min).

Ethical Considerations

It was determined by the University of Pennsylvania Institutional Review Board that this project is a quality improvement initiative that does not meet the definition of human subjects' research. Hence, further institutional review board review is not required. Before starting the survey, participants were informed that the data would be used to develop or enhance future workshops and that survey participation was voluntary. No identifiable information was collected from the survey, and participants did not receive any compensation.

Computing Environment

To optimize our instruction time during the workshop, we required an environment preconfigured with the appropriate security for storing and analyzing deidentified clinical data, access to workshop materials, and the necessary software packages for processing data. We leveraged a secure cloud computing environment within Microsoft Azure and created a workspace designed to meet Health Insurance Portability and Accountability Act security requirements and provide controlled access to the study materials. Lessons were constructed in Jupyter notebooks using the Microsoft Azure Databricks analytics platform and written in Python that explained rationale and demonstrated the processing steps in context.

Workshop Materials

The main workshop materials included practice data and prewritten code in Jupyter notebooks. The coding practice primarily focused on running prewritten code rather than writing new code, as the registration survey indicated that most participants had little to no experience with Python coding and NLP. However, we also offered participants the flexibility to modify the code during or after the workshop if they wished. Most importantly, we prioritized creating an interactive learning environment rather than delivering 1-directional instruction. For example, we regularly checked in with participants to assist with any issues they encountered during the coding practice, and 3 data scientists, in addition to the instructors, were available to provide support.

Specifically, the objective of the workshop materials was to demonstrate the application of data science, AI, and NLP to answer 3 biomedical research questions (RQs) using a dataset of chief complaints from the EHR:

- RQ1: How many chief complaints are related to suicidality?
- RQ2: Is the prevalence rate of suicidality statistically different between White and Black or African American populations?
- RQ3: Does the prevalence rate of suicidality increase with age?

To answer these questions, we organized the workshop materials to address 4 subtasks in data science and AI: data characterization, data standardization, concept extraction, and statistical analysis. For data characterization, we demonstrated how to generate frequency distributions, identify missing data, and handle outliers within the dataset. For data standardization, we applied general standardization techniques with Python function `str`, including addressing lexical variation with case

reduction (eg, replacing "SUICIDE" with "suicide"). We used the WordNetLemmatizer for lemmatization to address term variation within the same concept. For instance, "suicide," "suicidal," and "suicidality" were standardized to "suicide." We detected negation using `negex` (eg, "no suicide thought" was classified as "negated"). We also applied domain- and context-specific standardization. For example, in medical settings within the state of Pennsylvania, "302" refers to involuntary commitment for emergency evaluation and treatment, whereas "201" refers to voluntary commitment. These context-specific terms were replaced with descriptive and general terms (eg, "302" was replaced with "involuntary commitment").

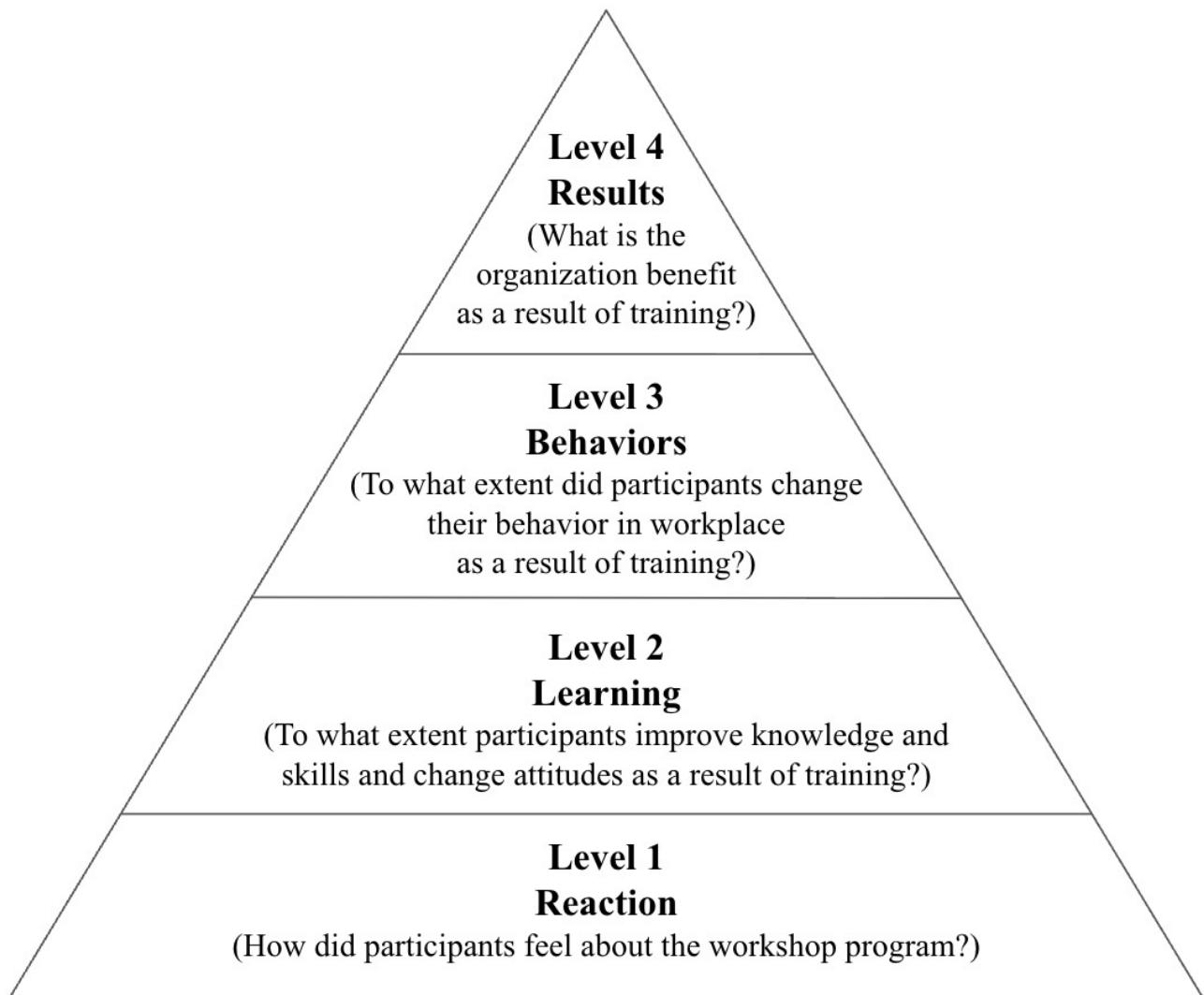
For concept extraction, we developed a dictionary containing suicide-related terms (eg, "suicide ideation" and "fatal self-harm behavior") and applied them using regular expressions. Finally, for statistical analysis, we demonstrated how to calculate the prevalence rate of suicidality between demographic groups and apply a *t* test to determine if there is a significant difference in prevalence rates between the groups, namely White and Black or African American populations. We also demonstrated how to examine the predictive effect of age (independent variable) on suicidality (dependent variable) using logistic regression. Two researchers drafted and demonstrated the code, and feedback from the Methods Core of the Penn Innovation in Suicide Prevention Implementation Research Center was integrated to improve the materials. The Methods Core's mission is to support the Penn Innovation in Suicide Prevention Implementation Research Center's suicide prevention research by applying computational, implementation science, and human-centered design approaches. Their role in the development and dissemination of this workshop was to provide feedback from both data science and clinical perspectives, ensuring that the workshop met the needs of psychiatry and psychology trainees and was developed in a user-friendly format. This included reorganizing NLP tasks to distinguish between content that target audiences may already know (eg, mental health-related terms) and content they may not know (eg, NLP-specific language). Feedback also recommended introducing types of clinical data (eg, chief complaints), referencing existing studies related to the research problem (eg, health disparities), and designing psychological research questions (eg, predicting suicide risk by race, ethnicity, and age).

Trainee-Centered Recruitment

Human-centered design was used to design the recruitment flyer, which prioritized clarity and appeal through the use of a logical visual hierarchy and accessible language and by highlighting speaker credentials to build trust and credibility. A QR code, which directed users to the Research Electronic Data Capture registration survey, was included in the flyer to simplify the registration process and minimize barriers to action in an effort to maximize participation. A digital flyer was shared via the Department of Psychiatry's email listserv at the University of Pennsylvania to recruit workshop participants. We did not restrict the number of workshop participants. Participation was open to any trainees (ie, students, postdoctoral researchers, and residents) in the Departments of Psychology

and Psychiatry. With the aim of reducing the burden on trainees participating in the data science workshop alongside their regular training requirements, we collaborated with the resident training office. To enhance dissemination efforts, we aligned the workshop schedule with the annual resident training calendar, leveraged prescheduled didactics time for trainees, and remained flexible toward other obligations. The 3-hour training workshop took place in August 2024 during a prescheduled weekday didactic session for trainees.

Figure 1. Kirkpatrick model for program evaluation [15].



At level 1, participants' overall satisfaction was assessed using 2 questions: (1) "Overall, did you find the workshop helpful?" rated on a 4-point Likert scale, ranging from 1 (not helpful) to 4 (very helpful); and (2) "How likely are you to recommend this workshop to others?" rated on a 4-point Likert scale, ranging from 1 (very unlikely) to 4 (very likely). Participants were also asked to identify specific components of the training they found helpful, with options including the introduction to overall NLP concepts, coding demonstrations, and materials that were provided. Each component was rated as 1 (helpful) or 0 (not helpful).

At level 2, the confidence level of NLP knowledge and Python coding skills was measured with 8 questions (eg, "How confident are you in the following concepts or skills?") with a 4-point Likert scale, ranging from 1 (not confident) to 4 (very confident). Specifically, confidence in NLP knowledge was assessed using 4 questions, each evaluating participants' confidence in data characterization, data standardization, data transformation from unstructured to structured data, and concept extraction. Participant confidence in their Python coding skills was measured with 4 questions assessing participants' confidence in performing tasks such as data standardization, data transformation, statistical analysis, and visualization using

Python. We evaluated changes in confidence level before and after the workshop using paired *t* tests for the 6 participants with pre- and post-training data. Additionally, open-ended questions such as “Are there other topics you wish future workshops to cover?” were included to gain insights into ways to revise the workshop.

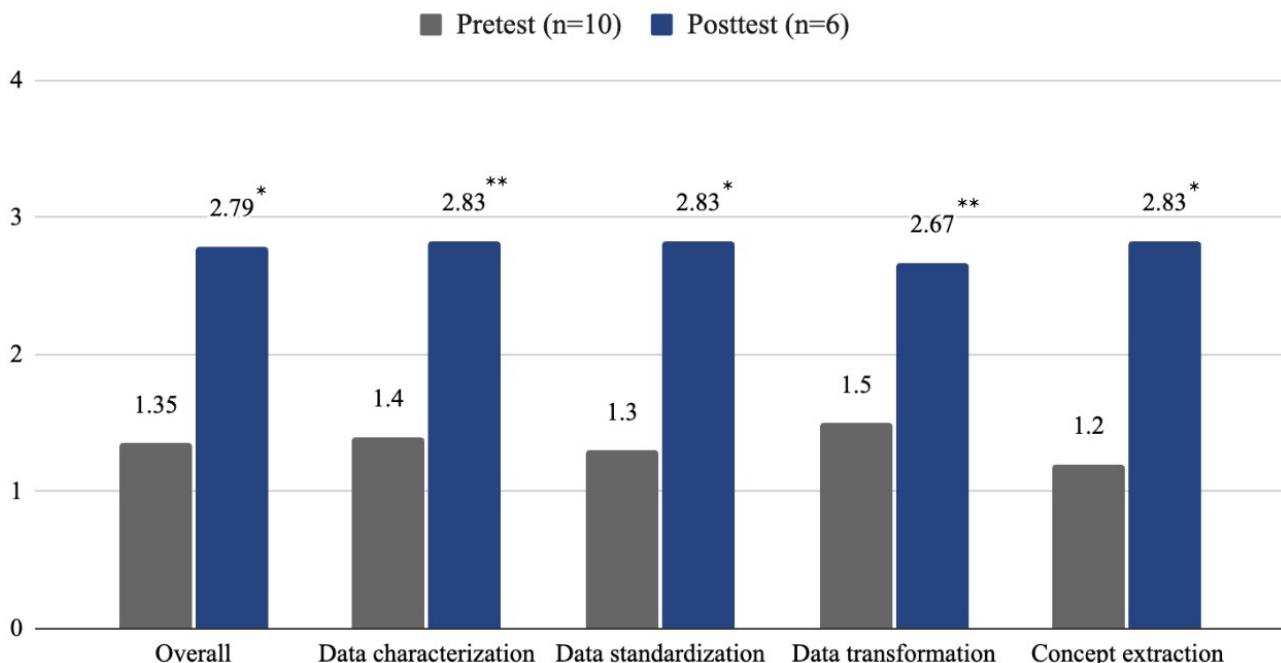
Results

Level 1: Reaction Evaluation

A total of 10 trainees participated in the workshop virtually. They included residents, postdoctoral researchers, and graduate students from the psychiatry and psychology departments. All participants completed the presurvey, whereas 6 (60%) participants completed the postsurvey. Only 2 (20%) of the 10 participants had experience with Python or NLP training before this workshop.

Figure 2. Confidence levels in natural language processing (NLP) knowledge before and after the workshop. * $P<.05$, ** $P<.01$.

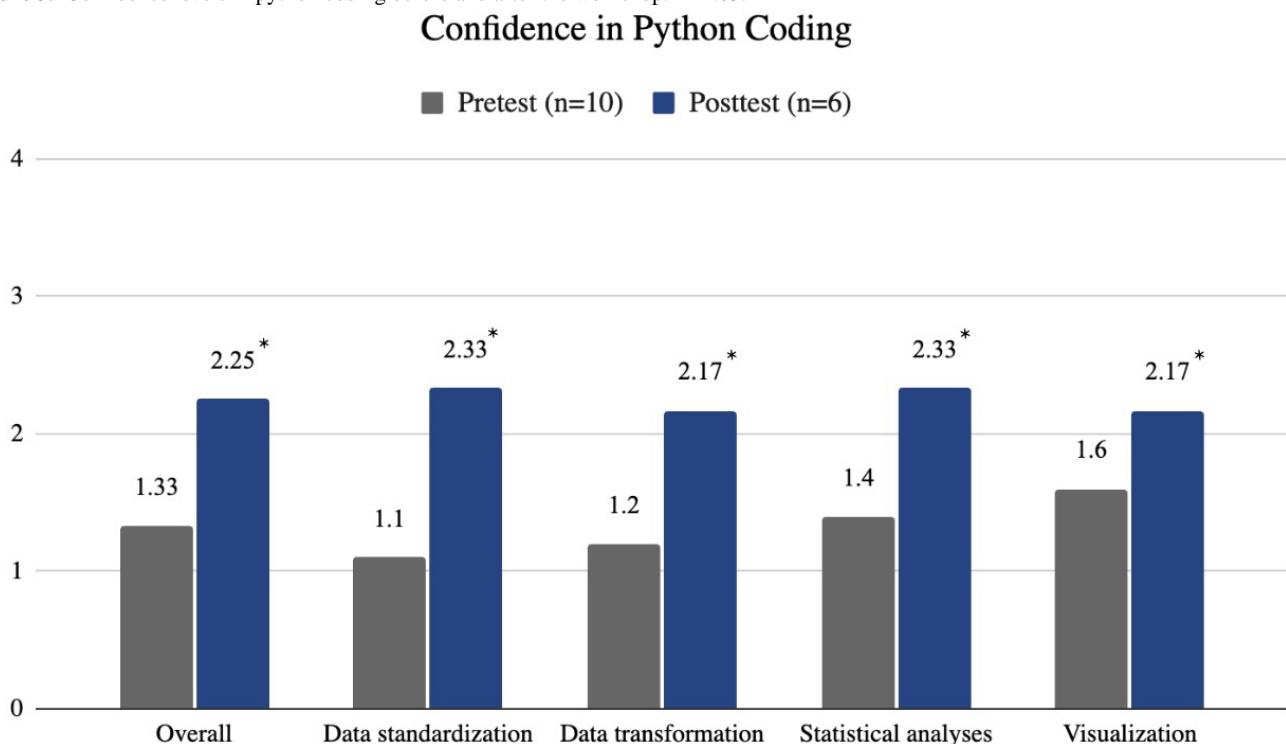
Confidence in NLP Knowledge



Level 2: Learning Evaluation—Python Coding Skills

Participants’ confidence in their overall coding abilities also significantly improved ($P=.01$), increasing from 1.33 (SD 0.60) to 2.25 (SD 0.42) as demonstrated in Figure 3. Participants reported significant improvements in confidence across all 4 individual Python coding skills. Confidence in standardizing text data increased ($P=.01$) from 1.10 (SD 0.75) to 2.33 (SD

0.52). Similarly, confidence in transforming unstructured data into structured data improved ($P=.01$) from 1.20 (SD 0.52) to 2.17 (SD 0.41). Confidence in conducting statistical analysis (ie, descriptive statistics) using Python increased significantly ($P=.01$) from 1.40 (SD 0.75) to 2.33 (SD 0.52). Confidence in visualizing analysis results increased significantly ($P=.04$) from 1.60 (SD 0.75) to 2.17 (SD 0.41).

Figure 3. Confidence levels in python coding before and after the workshop. * $P<.05$.

In addition to the quantitative findings, qualitative feedback included the positive experiences of participants with the workshop. Overall, the feedback from the open-ended questions by trainees was positive and insightful (eg, “I’m glad you did this workshop, I hope there are more to come with some more hands-on experiences, but I think this served as a good introduction to the process and rationale behind the code.”). Feedback also included suggestions for incorporating thematic analysis and exploring additional datasets for future workshop development, which are discussed further in the following section.

Discussion

Principal Findings

This study illustrates the effectiveness of a tailored data science workshop for trainees in psychiatry and psychology, focusing on applying NLP techniques and clinical notes for suicide prevention research. The workshop significantly enhanced participants’ skills and confidence in conducting data science research, especially for those with substantial clinical experience but limited AI research training.

We received constructive comments from trainees through open-ended questions to guide future training efforts ranging from data-driven use cases, Python coding, and advanced NLP skills. In terms of data-driven use cases, trainees thought the workshop could be enhanced by “...work[ing] independently with data sets like national available data sets” and “starting our own projects working with real data.” We could incorporate more experiential training by providing more diverse datasets, including national datasets such as the National Longitudinal Study of Adolescent to Adult Health dataset or recommending that users bring their own dataset to apply these approaches. One trainee expressed an interest in “how to learn basics of

coding - like any resources that would be helpful”; links to Python coding support could be provided or a Python coding bootcamp could be offered for trainees with less coding skills before the workshop. Another trainee recommended incorporating thematic analysis, for example, incorporating clustering techniques for deriving latent patterns in data. Additionally, future workshops could cover advanced NLP methods, including large language models (eg, GPT-4) to complete NLP subtasks. We could also incorporate exploring multifaceted outcomes, including clinical pathways, readmissions, and mortality, to demonstrate how to identify actionable insights from data.

Limitations

Our study has several limitations. First, we implemented this study among trainees from a single institution. Feedback might vary if trainees had more or less experience with NLP and Python coding skills at another institution. Second, in terms of assessing learning, we measured confidence rather than actual knowledge or skill in level 2. We could also assess actual levels of knowledge and skills beyond self-reported confidence using methods such as case studies for deeper insights. Third, in the future, we intend to examine whether trainees acquired level 3 (behaviors) and level 4 (results) learning. We could assess how participants’ learning impacts their practice and research in the long term. This includes whether participants applied the gained knowledge and skills to their clinical practice, research engagement, and health care improvements in the long term [16]. For instance, it would be valuable to assess whether participants are more likely to adopt NLP-based clinical support tools after the workshop (level 3) and whether increased use of these tools enhances the efficacy of clinical workflows in health care systems at a systemic level (level 4). The knowledge and skills gained may also encourage participants to initiate

independent research or engage in grant writing (level 3) to improve suicide prevention services over time (level 4). Fourth, our pilot study included a relatively small sample size. We aim to increase the size of the course by making it accessible to other trainee cohorts, including informatics, nursing, and engineering students with interest in health and AI. Finally, this study was an open trial without a controlled experimental approach. Adopting controlled methodologies, such as randomized

controlled trials, will be beneficial for validating these findings in future studies.

Conclusions

We successfully developed and implemented a 3-hour virtual workshop to educate psychiatry and psychology trainees on using data science and AI for suicide prevention research. Our preliminary results suggest the workshop enhanced participants' knowledge and coding abilities, establishing a strong foundation for future advanced curriculum and learning objectives.

Acknowledgments

No generative artificial intelligence was used in any content of this manuscript, including text and graphics.

Funding

This research was funded by the National Institute of Mental Health (P50MH127511).

Data Availability

The data that support the findings of this study are available on request from the corresponding author and with approval from the institutional review board.

Authors' Contributions

Conceptualization: HKD, DM, ES, UV, MER, MS, AC, RG, DLM

Data curation: HKD, DLM

Formal analysis: HKD

Investigation: HKD

Methodology: HKD, DM, SH, ES, UV, MER, AC, RG, GKB, DLM

Project administration: HKD, ES, UV, MAO, GKB, DLM

Resources: HKD, DM, SH, GN, EP, MAO, GKB, DLM

Software: HKD, SH, GN, DLM

Supervision: DM, MAO, GKB, DLM

Validation: HKD, SH, MS

Visualization: HKD

Writing – original draft: HKD, DLM

Writing – review & editing: HKD, DM, SH, EP, MS, MAO, GKB, DLM

Conflicts of Interest

MAO receives royalties from the Research Foundation for Mental Hygiene for commercial use of the Columbia Suicide Severity Rating Scale. She volunteers as a scientific advisor to Mind Medicine and reviews grants for Alkermes and is a Trustee of Tufts University. She also advises St. George's University and Fundacion Jimenez Diaz. GKB receives royalties from the Research Foundation for Mental Hygiene for the commercial use of the Columbia Severity Rating Scale.

References

1. Dehbozorgi R, Zangeneh S, Khooshab E, et al. The application of artificial intelligence in the field of mental health: a systematic review. *BMC Psychiatry* 2025 Feb 14;25(1):132. [doi: [10.1186/s12888-025-06483-2](https://doi.org/10.1186/s12888-025-06483-2)] [Medline: [39953464](https://pubmed.ncbi.nlm.nih.gov/39953464/)]
2. Subaveerapandian A, Mvula D, Ahmad N, Taj A, Ahmed MG. Assessing AI literacy and attitudes among medical students: implications for integration into healthcare practice. *JHOM* 2024 Oct 28. [doi: [10.1108/JHOM-04-2024-0154](https://doi.org/10.1108/JHOM-04-2024-0154)]
3. Kimiafar K, Sarbaz M, Tabatabaei SM, et al. Artificial intelligence literacy among healthcare professionals and students: a systematic review. *Front Health Inform* 2023 Nov 11;12:168. [doi: [10.30699/fhi.v12i0.524](https://doi.org/10.30699/fhi.v12i0.524)]
4. 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](https://doi.org/10.1007/s00330-018-5601-1)] [Medline: [29980928](https://pubmed.ncbi.nlm.nih.gov/29980928/)]
5. Mousavi Baigi SF, Sarbaz M, Ghaddaripour K, Ghaddaripour M, Mousavi AS, Kimiafar K. Attitudes, knowledge, and skills towards artificial intelligence among healthcare students: a systematic review. *Health Sci Rep* 2023 Mar;6(3):e1138. [doi: [10.1002/hsr2.1138](https://doi.org/10.1002/hsr2.1138)] [Medline: [36923372](https://pubmed.ncbi.nlm.nih.gov/36923372/)]
6. Güdük Ö, Vural A, Disiäçik G. Investigating the effect of perceived empowerment on artificial intelligence anxiety levels in healthcare workers. *Çalışma ve Toplum* 2025 Jan 1;1(84):285-310. [doi: [10.54752/ct.1624015](https://doi.org/10.54752/ct.1624015)]

7. Theros J, Soetikno A, Liebovitz D. Foundations of AI for future physicians: a practical, accessible curriculum. *Med Teach* 2025 Oct;47(10):1601-1603. [doi: [10.1080/0142159X.2025.2463492](https://doi.org/10.1080/0142159X.2025.2463492)] [Medline: [39940108](#)]
8. 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](https://doi.org/10.2196/31043)] [Medline: [34898458](#)]
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](https://doi.org/10.1097/ACM.0000000000004291)] [Medline: [34348374](#)]
10. Wood EA, Ange BL, Miller DD. Are we ready to integrate artificial intelligence literacy into medical school curriculum: students and faculty survey. *J Med Educ Curric Dev* 2021;8:23821205211024078. [doi: [10.1177/23821205211024078](https://doi.org/10.1177/23821205211024078)] [Medline: [34250242](#)]
11. Alelyani M, Alamri S, Alqahtani MS, et al. Radiology community attitude in Saudi Arabia about the applications of artificial intelligence in radiology. *Healthcare (Basel)* 2021 Jul 1;9(7):834. [doi: [10.3390/healthcare9070834](https://doi.org/10.3390/healthcare9070834)] [Medline: [34356212](#)]
12. Arowosegbe A, Oyelade T. Application of natural language processing (NLP) in detecting and preventing suicide ideation: a systematic review. *Int J Environ Res Public Health* 2023 Jan 13;20(2):1514. [doi: [10.3390/ijerph20021514](https://doi.org/10.3390/ijerph20021514)] [Medline: [36674270](#)]
13. Tsui FR, Shi L, Ruiz V, et al. Natural language processing and machine learning of electronic health records for prediction of first-time suicide attempts. *JAMIA Open* 2021 Jan;4(1):ooab011. [doi: [10.1093/jamiaopen/ooab011](https://doi.org/10.1093/jamiaopen/ooab011)] [Medline: [33758800](#)]
14. Korda A, Heide M, Nag A, et al. Suicide prediction with natural language processing of electronic health records. *medRxiv*. Preprint posted online on Sep 29, 2023. [doi: [10.1101/2023.09.28.23296268](https://doi.org/10.1101/2023.09.28.23296268)]
15. Kirkpatrick D, Kirkpatrick J. *Evaluating Training Programs: The Four Levels*; Berrett-Koehler Publishers; 2006.
16. Parker ND, Michaels M, Fisher CL, et al. A web-based training intervention for primary care providers on preparing patients for cancer treatment decisions and conversations about clinical trials: evaluation of a pilot study using mixed methods and follow-up. *JMIR Med Educ* 2025 Jul 17;11:e66892. [doi: [10.2196/66892](https://doi.org/10.2196/66892)] [Medline: [40674725](#)]

Abbreviations

AI: artificial intelligence

EHR: electronic health record

NLP: natural language processing

RQ: research question

Edited by T Leung; submitted 03.Apr.2025; peer-reviewed by B Senst, NC Favour; revised version received 07.Sep.2025; accepted 16.Sep.2025; published 16.Jan.2026.

Please cite as:

Donnelly HK, Mandell D, Hwang S, Schriver E, Vurgun U, Neill G, Patel E, Reilly ME, Steinberg M, Calloway A, Gallop R, Oquendo MA, Brown GK, Mowery DL

Data Science Education for Residents, Researchers, and Students in Psychiatry and Psychology: Program Development and Evaluation Study

JMIR Med Educ 2026;12:e75125

URL: <https://mededu.jmir.org/2026/1/e75125>

doi: [10.2196/75125](https://doi.org/10.2196/75125)

© Hayoung K Donnelly, David Mandell, Sy Hwang, Emily Schriver, Ugurcan Vurgun, Graydon Neill, Esha Patel, Megan E Reilly, Michael Steinberg, Amber Calloway, Robert Gallop, Maria A Oquendo, Gregory K Brown, Danielle L Mowery. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 16.Jan.2026. 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.

Original Paper

Effectiveness of Informed AI Use on Clinical Competence of General Practitioners and Internists: Pre-Post Intervention Study

Eyad A Qunaibi¹; Ayman M Al-Qaaneh²; Baraa F Ismail³; Hussam I Muhidat⁴; Farhia S Rageh⁵; Najwa A Musallam⁶; Alaa K Fawzy⁷

¹Department of International Research, MedOne Academy, Dover, DE, United States

²Faculty of Allied Medical Sciences, Al-Balqa Applied University, Al-Salt, Jordan

³Faculty of Public Health, Jinan University, Tripoli, Lebanon

⁴PriceWaterhouseCoopers Middle East, Doha, Qatar

⁵Fakeeh University Hospital, Dubai, United Arab Emirates

⁶King Hussein Cancer Center, Amman, Jordan

⁷Medical Biochemistry and Molecular Biology Department, Faculty of Medicine, Zagazig University, Zagazig, Egypt

Corresponding Author:

Ayman M Al-Qaaneh

Faculty of Allied Medical Sciences

Al-Balqa Applied University

19117

Al-Salt

Jordan

Phone: 962 779596290

Email: ayman.alqenah@bau.edu.jo

Abstract

Background: Artificial intelligence (AI) shows promise in clinical diagnosis, treatment support, and health care efficiency. However, its adoption in real-world practice remains limited due to insufficient clinical validation and an unclear impact on practitioners' competence. Addressing these gaps is essential for effective, confident, and ethical integration of AI into modern health care settings.

Objective: This study aimed to evaluate the effectiveness of informed AI use, following a tailored AI training course, on the performance of general practitioners (GPs) and internists in test-based clinical competence assessments and their attitudes toward clinical AI applications.

Methods: A pre-post intervention study was conducted with 326 physicians from 39 countries. Participants completed a baseline test of clinical decision-making skills, covering diagnosis, treatment planning, and patient counseling; attended a 1.5-hour online training on effective AI use; and then took a similar postcourse test with AI assistance permitted (GPT-4.0). Test performance and time per question were compared before and after the training. Participants also rated AI accuracy, efficiency, perceived need for structured AI training, and their willingness to use AI in clinical practice before and after the course.

Results: The average test scores improved from 56.9% (SD 15.7%) to 77.6% (SD 12.7%; $P<.001$), and the pass rate increased from 6.4% (21/326) to 58.6% (191/326), with larger gains observed among GPs and younger physicians. All skill domains (diagnosis, treatment planning, and patient counseling) improved significantly (all $P<.001$), while time taken to complete the test increased slightly from before to after the course (mean 40.25, SD 16.14 min vs 42.29, SD 14.02 min; $P=.03$). By the end of the intervention, physicians viewed AI more favorably, reporting increased confidence in its accuracy and time efficiency, greater appreciation for the need for structured AI training, and increased confidence and willingness to integrate AI into patient care.

Conclusions: Informed use of AI, based on tailored training, was associated with higher performance in test-based clinical decision-making assessments and greater confidence in using AI among GPs and internists. Building on previous research that often lacked structured training, focused primarily on model performance, or was limited in clinical scope, this study provides empirical evidence of both competence and perceptual improvement following informed AI use in a large, multinational cohort, enhancing the generalizability. These findings support the integration of structured AI training into medical education and continuing professional development to improve clinical performance and promote competent use of AI in clinical practice.

KEYWORDS

artificial intelligence; AI; clinical competence; artificial intelligence training; AI training; clinical decision support; medical education; digital health tools

Introduction

Artificial intelligence (AI) has demonstrated significant potential in disease diagnosis, treatment recommendations, patient engagement, and medical writing and education [1-3]. Extensive research has explored and compared the accuracy of AI models across these domains [4-7]. In addition, AI has revolutionized diagnostic radiology by enhancing image analysis and interpretation [8]. However, despite these promising applications, most AI models and software remain within the realm of research rather than real-world settings, reflecting a notable gap in implementation [9,10].

Before widespread adoption, AI applications in health care require not only algorithm validation but also clinical validation, comparing AI-based interventions with standard treatments and existing clinical practices [11]. A limited number of studies have evaluated the effect of AI assistance in clinical settings and demonstrated improvements in the quality, efficiency, and effectiveness of health care services [12,13]. In contrast, there is an abundance of studies that have evaluated the impact of AI assistance in controlled experimental settings rather than real-world clinical scenarios. Feigerlova et al [14] highlighted the limitations of these studies, including small sample sizes, single-center designs, inadequate control for confounding factors, and the absence of well-defined clinical competence frameworks.

In addition, existing studies often do not include training on appropriate AI use. To enable AI integration into clinical practice, tailored AI training for health care practitioners is required [15]. The situation is even more critical for primary care physicians, as Liaw et al [16] have highlighted the lack of attention given to training primary care physicians in the use of AI-based tools and emphasized the necessity of such training to maximize benefits and minimize potential harms. Ideally, this training should be followed by objective assessments of improvements in clinical competence after AI use. However, evidence on the objective improvements in clinical competence with AI assistance after structured AI training remains scarce.

Beyond technical accuracy and the need for training, the successful adoption of AI in clinical settings is heavily influenced by health care professionals' perceptions and attitudes. Acceptance is an indispensable prerequisite for the widespread implementation of AI [15]. Although many studies have assessed health care professionals' general attitudes toward AI integration into medical practice [15], to the best of our knowledge, no study has specifically examined how tailored AI training impacts practitioners' perceptions of AI and their willingness to incorporate it into clinical workflows. A recent systematic review identified key factors that influence health care workers' trust in AI tools for making informed clinical decisions [17]. However, none of the 27 included studies

implemented interventions with pre-post repeated-measurement designs; instead, their assessment methods were predominantly semistructured qualitative interviews, focus groups, or Likert-scale cross-sectional surveys. Collectively, many previous studies were limited by small sample sizes, narrow clinical scope, a lack of structured AI training, evaluation in simulated rather than real-world clinical contexts, and a lack of pre-post intervention assessments of perceptual change. These accumulated limitations and gaps in knowledge highlight the significance of our study.

Unlike previous research, our study approximates real-world clinical practice by integrating AI as a decision-support tool for general practitioners (GPs) and internists who were tested using real clinical case scenarios. Through a structured, tailored training course with systematic pre-post course assessments, we objectively evaluate test-based clinical competence of GPs and internists following informed AI use. We assess improvements in the skills of diagnosis, management, and patient counseling. We also uniquely assess shifts in GPs' and internists' perceptions of AI diagnostic accuracy, treatment recommendation efficacy, time efficiency, confidence, and willingness to integrate AI into clinical practice. In addition to these novelties, our study is multinational, large-scale, and outcome-driven, addressing key limitations of previous research.

By providing a replicable model for evaluating AI-assisted clinical decision-making, this study establishes a framework for optimizing AI use in health care settings, helping to bridge the gap between research-based AI performance and real-world clinical effectiveness. The aim of this study was to evaluate the test-based clinical competence and perceptions of GPs and internists, following informed AI use after a tailored AI training course.

Methods

Study Design and Procedures

This quasi-experimental study used a pre-post test design to evaluate the impact of informed AI use on clinical competence among GPs and internists. Conducted entirely online, the course and assessments were asynchronous, allowing participants to complete the training and tests at their preferred time. Each participant completed a precourse test before accessing the recorded modules, followed by a postcourse test, in which AI assistance was permitted (ChatGPT [version 4.0; OpenAI]), to measure changes in clinical competence. In this study, *informed AI use* refers to physicians using AI tools after receiving a tailored AI training course on how to effectively and responsibly integrate these tools into their practice. Participants' self-reported previous use of AI in clinical practice was recorded and categorized for analysis (Multimedia Appendix 1).

Ethical Considerations

This study was conducted in accordance with ethical guidelines and approved by the institutional review board of MedOne Academy (MO-IA-24/25-EDU-1). All participants were informed about the study objectives and the voluntary nature of their participation, with the right to withdraw at any time. Electronic informed consent was obtained, and participants were assured of data confidentiality. Participants' names and email addresses were collected solely for administrative purposes to assign and link precourse and postcourse assessments. These identifiers were not used for analysis and were not shared with any third party. All study data were analyzed in coded and anonymized form. Access to identifiable information was strictly limited to the first and corresponding authors. All data were stored on secure, password-protected systems accessible only to authorized members of the research team. Participants did not receive any financial or material compensation for participation in this study.

Participants

Participants included GPs, encompassing family medicine specialists and internists, including those pursuing subspecialty training after internal medicine. These classifications apply throughout the study. Details of participant recruitment, eligibility verification, and enrollment timeline are provided in [Multimedia Appendix 2](#).

Assessment Tool

Overview

This study used a structured assessment tool comprising multiple-choice questions (MCQs) designed to evaluate clinical competence across three key areas: (1) diagnosis and patient assessment, (2) treatment planning and personalized medicine, and (3) patient counseling. Each test set (A and B) included items covering these same domains to ensure content consistency across assessments. The MCQs were developed by 2 of the study authors (AKF and BFI), who are physicians, and reviewed and revised by 2 additional expert physicians (Kamel Hatahet and Suleiman Al Ashi), to ensure their relevance to participants and the adequacy in assessing the 3 studied competencies. The test sets underwent expert review for face and content validity, and their clarity and timing were further verified in the pilot phase as described subsequently. [Multimedia Appendix 3](#) includes the source references from which the clinical cases were derived, along with selected adapted sample cases and their corresponding MCQs with answer keys.

Crossover Design and Validation

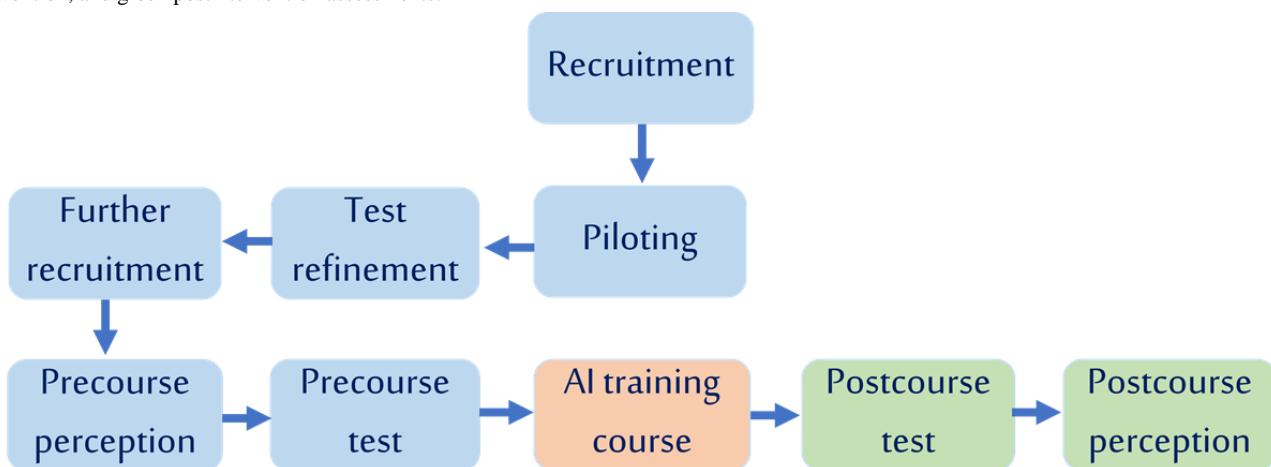
A crossover design was implemented, with participants allocated to 2 groups for the precourse test: group A received 25 MCQs, and group B received 23 MCQs of comparable difficulty. Participants who registered through the MedOne Academy platform were alternately allocated to group A or group B to maintain balanced group sizes and comparable baseline characteristics. No significant difference was observed in the GP-to-internist distribution between the 2 test-order groups ($P=.11$), supporting the comparability of participant composition. After the course, participants switched question sets. This crossover design was selected because it allowed each participant to serve as their own control, thereby reducing variability due to individual differences and maintaining statistical power with a feasible sample size while ensuring that all participants received the educational intervention. Conducting a conventional randomized controlled design would have required a nonintervention control group, which was impractical and would likely have compromised motivation and completion rates, as the course itself served as the primary educational and engagement incentive.

The difficulty level was preassessed by the 4 contributing experts (AKF, BFI, Kamel Hatahet, and Suleiman Al Ashi) to ensure equivalence, and the comparability of participant performance across both test sets A and B was statistically confirmed, as detailed in the Results section and Table S1 in [Multimedia Appendix 4](#).

Assessment Administration and Scoring Criteria

Each test was allotted a maximum duration of 1 hour. A passing threshold of 80% was applied, consistent with established benchmarks in medical certification ([Multimedia Appendix 5](#)). Clear and standardized instructions on AI use were provided to participants to ensure uniform conditions across assessments. These instructions, along with the technical setup implemented to maintain the reliability of experimental conditions, are detailed in [Multimedia Appendix 6](#). Additionally, participants completed the same set of perception questions before and after the course to assess how the course and study experience influenced their attitudes toward AI in clinical practice. These questions assessed participants' perceptions of AI accuracy in diagnosis and treatment planning, its time-saving potential in clinical settings, the perceived need for structured AI training, and their willingness and confidence to incorporate AI into clinical practice ([Multimedia Appendix 7](#)). Each participant was given up to 2 weeks to complete the entire study process after enrollment. A summary of the study flow is illustrated in [Figure 1](#).

Figure 1. Study flow diagram. Participant recruitment and piloting (including test refinement), followed by precourse perception survey and competence test, artificial intelligence training course, and postcourse competence test and perception survey. Blue indicates preparatory phases, orange the intervention, and green postintervention assessments.



GPT-4.0 was selected as the designated large language model (LLM) for participants to use in the postcourse test. Its selection was based on published evidence demonstrating the superior performance of GPT models on medical examination-style questions compared with other contemporary LLMs [7].

Intervention (Tailored AI Course)

The training program, titled AI Skills in Medicine, was designed as an asynchronous 1.5-hour online workshop prepared and presented by the main author (EAQ). It introduced AI applications in diagnosis, management, predictive analytics, and triage and emphasized ethical and responsible implementation. The course also included brief segments on privacy, account customization, and optimal AI use in clinical reasoning. Participants were instructed to evaluate AI outputs and their underlying rationale carefully rather than accepting them as is. This approach aligns with the viewpoint of Izquierdo-Condoy et al [18], who argued that, when thoughtfully embedded within educational frameworks, generative AI tools can enhance cognitive abilities, supporting rather than replacing clinical reasoning. By addressing these various modules, the program covered the competencies suggested by previous research [16] to maximize potential benefits and minimize harms associated with AI incorporation in clinical practice. A detailed outline of the 4 modules and learning objectives is provided in [Multimedia Appendix 8](#).

Pilot Study

Before the main study, a pilot group of 15 GPs and internists was recruited to follow the same study procedure, including attending the course and completing the tests. This phase allowed researchers to evaluate the clarity and timing of the tests as well as the applicability of the test platform. On the basis of participant feedback, the assessment process and training content were refined. For instance, open-ended questions were removed after observing excessive time consumption and noting that poor scores often reflected time constraints rather than knowledge gaps. The 15 pilot participants were excluded from the final analysis.

Studied Parameters

Primary Outcome: Clinical Competence Assessment

This study assessed multiple parameters to evaluate the impact of informed AI use comprehensively. Primary outcomes were the clinical competence scores obtained in the precourse and postcourse tests, the difference between postcourse and precourse scores, the pass or fail status based on an 80% threshold in both tests, and the total test completion time (min) before and after the course.

Secondary Outcomes: Skill-Specific Performance Measures

Secondary outcomes focused on skill-specific performance and included the number of correct answers per skill domain (diagnosis, treatment planning, and patient counseling) and the average time per correct answer per skill. The latter was calculated by summing the time spent on correct answers across all relevant case scenarios within each skill domain and dividing by the number of correct responses.

Perception Outcomes and Assessment

Perceptual changes were also evaluated, covering participants' perceptions of AI accuracy in diagnosis and treatment planning, its time-saving potential, the perceived benefit of structured AI training, and their willingness to incorporate AI into clinical practice. The perception shift was defined as the change in responses to these perception items before and after the intervention.

All outcomes were analyzed to determine improvements in clinical competence and shifts in attitudes following the tailored AI training and informed AI use.

Statistical Analysis

Categorical variables were presented as frequencies and percentages, and continuous variables were reported as means with SDs or medians with IQRs, depending on data distribution assessed using the Kolmogorov-Smirnov test. Baseline differences in categorical variables were analyzed using the chi-square goodness-of-fit test, whereas the Wilcoxon

signed-rank test (Mann-Whitney *U* test) or the independent samples *t* test was applied for continuous variables, as appropriate. The validity of the crossover design was assessed using the chi-square test, Mann-Whitney *U* test, or independent sample *t* test, as appropriate. The McNemar test or paired-samples *t* test was used to evaluate differences in the number of participants who passed or failed, test scores (%), and completion time (min) before and after AI use, as appropriate. An independent samples *t* test or 1-way ANOVA determined which group benefited most from AI assistance. Spearman rank correlation assessed the association between participants' age and score differences. The Wilcoxon signed-rank test with effect size analysis was used to evaluate changes in scores, time, and perceptions before and after AI use, stratified by skill domains (diagnosis, treatment planning, and patient counseling). Perception differences between groups were analyzed using the Mann-Whitney *U* or Kruskal-Wallis test, as appropriate. All statistical tests were 2 tailed, and

statistical significance was set at $P < .05$. No imputations were made for missing data points. All data used in the study were analyzed using SPSS (version 25.0; IBM Corp).

Results

Recruitment and Baseline Characteristics

A total of 2336 individuals completed the recruitment form, of whom 1665 (71.3%) met the eligibility criteria by belonging to the target specialties and providing valid medical licensure. The remaining 671 (28.7%) individuals were excluded for lacking licensure or for being medical students, specialists, or professionals outside the core medical field. Of the eligible 1665 participants, 326 (19.6%) completed the precourse test, attended the course, and took the postcourse test. Of the 326 participants, 250 (76.7%) also completed the precourse and postcourse perception assessments. All 326 participants were included in the final analysis (Figure 2).

Figure 2. Recruitment, eligibility screening, and participant inclusion flowchart. Of 2336 individuals who expressed interest, 671 (28.7%) were ineligible and 1665 (71.3%) were eligible. Among eligible individuals, 1339 (80.4%) did not enroll or did not complete the postcourse test, and 326 (19.6%) completed both tests and attended the artificial intelligence training course. Of these, 76 (23.3%) did not complete both perception assessments and 250 (76.7%) completed all study components. Red boxes indicate attrition or exclusion at each stage.

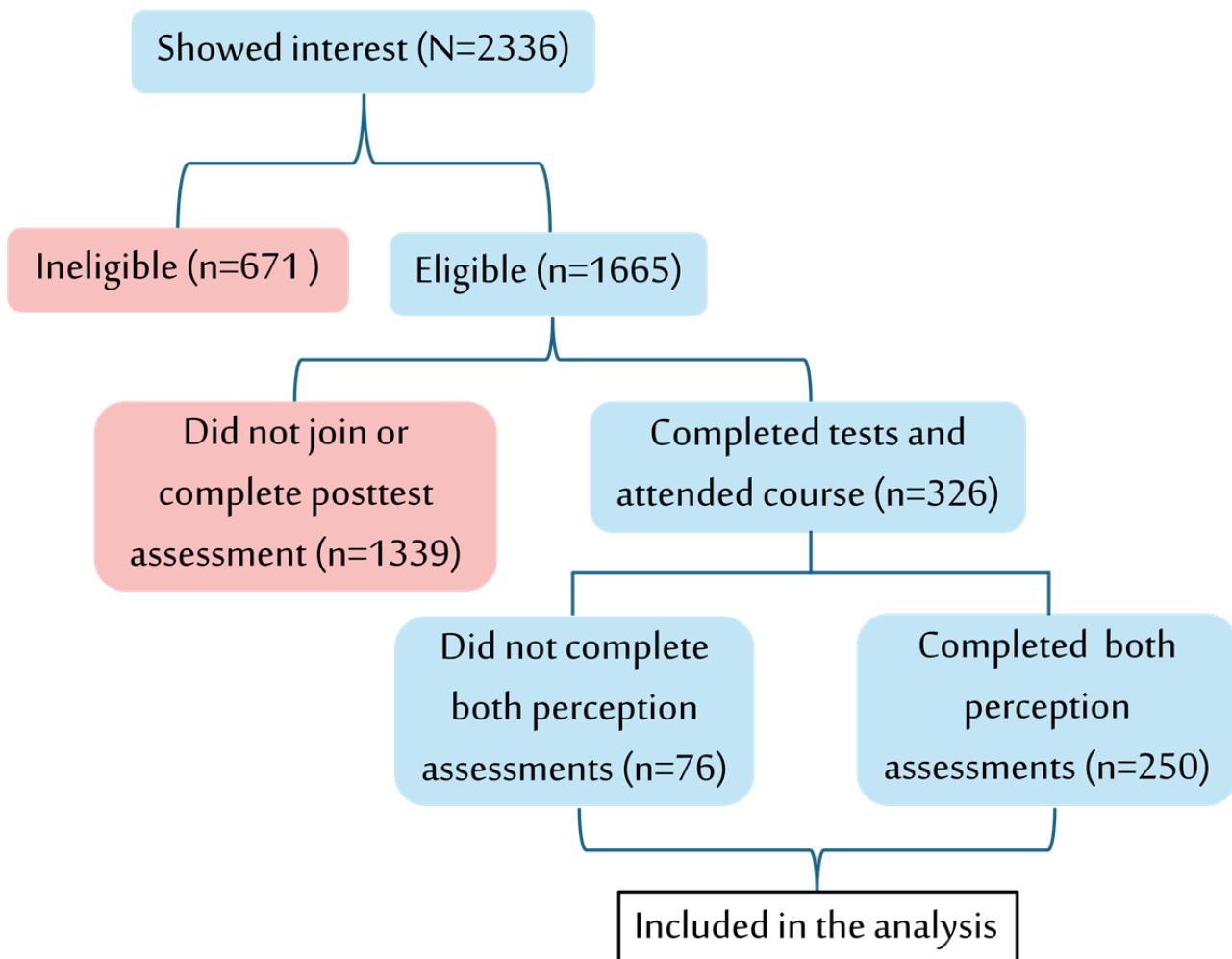


Table 1 summarizes the demographics of the 326 participants (male: n=193, 59.2%; median age 31, IQR 27.75-38 y). Most (n=229, 70.2%) were GPs. Most participants (n=208, 63.8%) reported little to no previous use of AI in clinical practice, and

308 (94.5%) had no previous AI training. Nearly half (n=173, 53.1%) served both inpatients and outpatients. The baseline median score on the precourse test was 57% (47.9%-69.6%), with a median completion time of 41.0 (27.0-56.0) minutes.

The participants were distributed across 39 countries, with the highest participation from Saudi Arabia, Syria, Egypt, Algeria, and Jordan.

Table 1. Participant demographics and baseline characteristics (n=326)^a.

Variable	Values
Sex, n (%)	
Male	193 (59.2)
Female	133 (40.8)
Specialty, n (%)	
GP ^b	229 (70.2)
Internist	97 (29.8)
Use of AI^c in clinical practice, n (%)	
Not at all	139 (42.6)
Rarely	69 (21.2)
Sometimes	87 (26.7)
Often	18 (5.5)
Almost every day	13 (4)
Familiarity with AI, n (%)	
GP (n=229)	84 (36.7)
Internist (n=97)	34 (35.1)
Received previous training in AI use, n (%)	
Yes	18 (5.5)
No	308 (94.5)
Type of patients served, n (%)	
Outpatients	96 (29.4)
Inpatients	57 (17.5)
Both	173 (53.1)
Test results before the AI training course, n (%)	
Fail	305 (93.6)
Pass	21 (6.4)
Age (y), median (IQR)	
GPs	31.00 (27.75-38.00)
Internists	30.00 (27.00-37.00)
35.00 (29.00-40.00)	
Precourse score (%), median (IQR)	
Precourse score of GPs	57.00 (47.85-69.60)
Precourse score of internists	56.00 (44.00-64.00)
68.00 (52.00-74.00)	
Time taken for precourse test (min), median (IQR)	41.00 (27.00-56.00)

^bGP: general practitioner.

^cAI: artificial intelligence.

^aData are presented as counts (percentages) for categorical variables and as medians (IQRs) for continuous variables.

Validation of the Crossover Design

To validate the crossover design, precourse scores ($P=.65$), postcourse scores ($P=.65$), and score differences ($P=.15$) were compared between groups A and B, along with pass and fail

comparisons between test sets A and B at both time points (pretest: $P=.31$; posttest: $P=.19$; Table S1 in [Multimedia Appendix 4](#)). None of these comparisons yielded a statistically significant difference, confirming the validity of the study design.

Effect of the Informed AI Use on Participants' Competence

Table 2 illustrates the impact of informed AI use on participants' competence. The proportion of participants who passed the test after taking the course and being allowed to use AI increased

significantly ($P<.001$). The mean score improved from 56.88% (SD 15.65%) to 77.56% (SD 12.71%), with a mean difference of 20.68% ($P<.001$). The mean time taken to complete the test increased slightly (by approximately 2 minutes) from before to after the course (40.25, SD 16.14 vs 42.29, SD 14.02 minutes; $P=.03$).

Table 2. Effect of informed artificial intelligence use on the participants' competencies (n=326). Comparison of participants' performance before and after the artificial intelligence training course.

Variable	Before course	After course	Mean difference	Test statistic	P value
Test results, n (%)				$\chi^2_1=153.55^a$	<.001
Fail	305 (93.6)	135 (41.4)	— ^b		
Pass	21 (6.4)	191 (58.6)	—		
Score (%), mean (SD)	56.88 (15.65)	77.56 (12.71)	20.69	$t_{325}=19.58^c$	<.001
Time taken (min), mean (SD)	40.25 (16.14)	42.29 (14.02)	2.04	$t_{325}=2.17^c$.03

^aMcNemar χ^2 .

^bNot applicable.

^cPaired samples t test.

Score improvements were analyzed across participant characteristics (Table S2 in [Multimedia Appendix 4](#)), revealing no significant differences except between GPs and internists, with the former showing significantly greater improvement than the latter (23.7%, SD 19.1% vs 13.7%, SD 17.3%; $P<.001$). Spearman correlation analysis revealed a statistically significant weak negative correlation between participants' age and score difference ($\rho=-0.143$; $P=.01$).

Skill-Specific Performance Outcomes

To determine whether the observed improvement was broad across all competencies or specific to certain skill areas, individual skill-specific performance was analyzed. Significant score increases were observed across skills of diagnosis, treatment planning, and patient counseling ([Table 3](#); [Figure 3](#); $P<.001$ for all). However, the average time per correct answer per skill increased significantly after informed AI use (Table S3 in [Multimedia Appendix 4](#)).

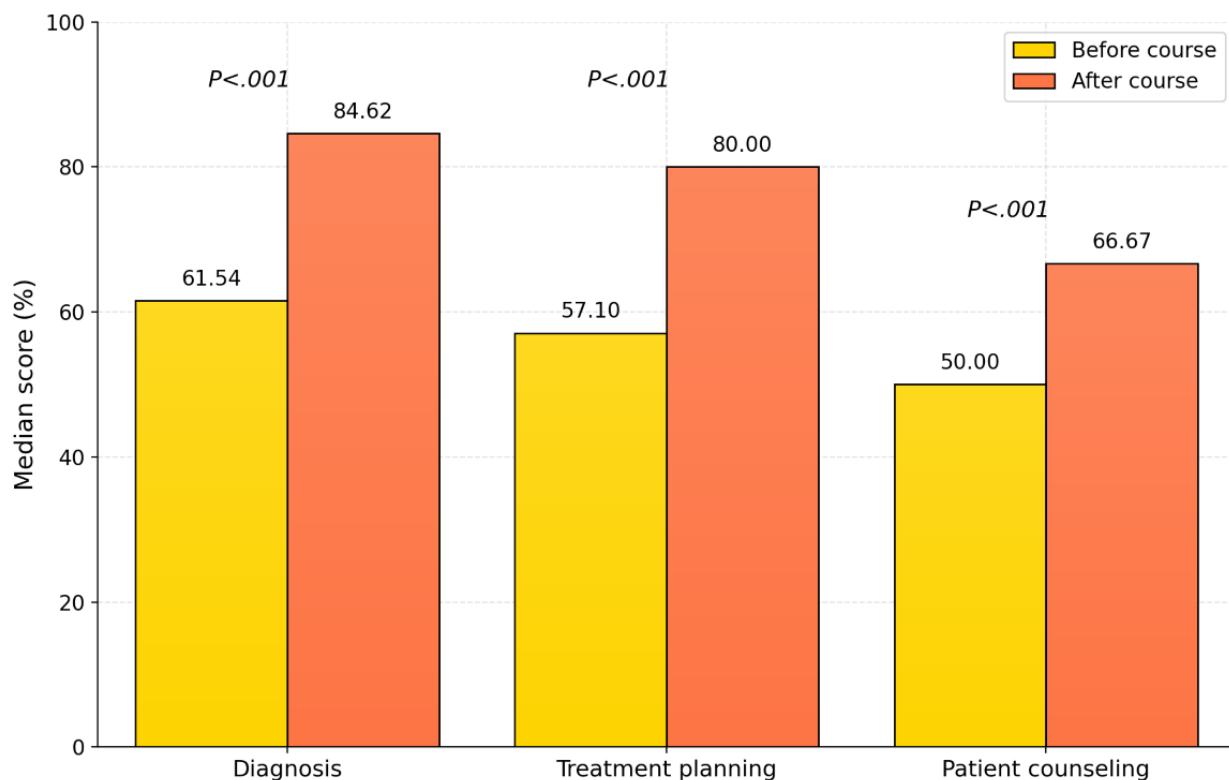
Table 3. Score comparison for each skill (n=326)^a.

Skill	Precourse score (%), median (IQR)	Postcourse score (%), median (IQR)	z score	P value ^b	Effect size, r
Skill 1: diagnosis	61.54 (46.15-69.23)	84.62 (76.90-92.30)	-12.57	<.001	0.738
Skill 2: treatment planning	57.10 (42.14-70.00)	80.00 (70.00-85.71)	-12.39	<.001	0.686
Skill 3: patient counseling	50.00 (33.33-100.00)	66.67 (50.00-100.00)	-7.75	<.001	0.420

^aComparison of participants' precourse and postcourse scores across 3 assessed competencies: diagnosis, treatment planning, and patient counseling. Data are presented as median (IQR).

^bWilcoxon signed-rank test.

Figure 3. Skill-specific median scores before and after informed artificial intelligence (AI) use (N=326). Median scores (%) for diagnosis, treatment planning, and patient counseling before and after the AI training course. Wilcoxon signed-rank tests showed statistically significant differences for all skills ($P<.001$).

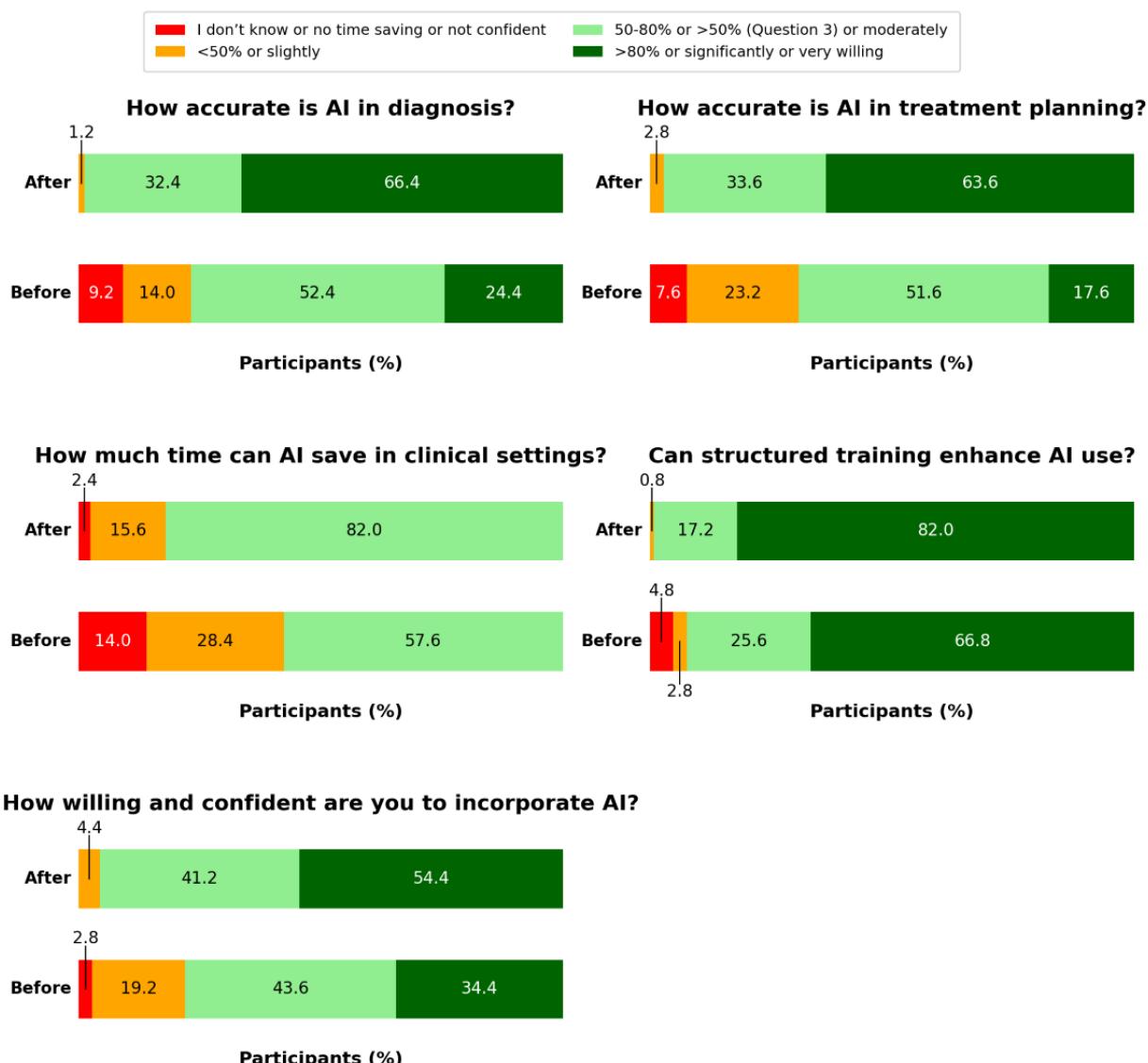


Effect of the AI Course and Informed AI Use on Participants' Perceptions

As shown in [Figure 4](#) and Table S4 in [Multimedia Appendix 4](#), participants' perceptions of AI accuracy, time efficiency, and integration into clinical practice improved markedly in the postcourse perception assessment, reflecting changes observed after both the course and informed AI use. All 5 perception changes (questions 1-5) were statistically significant (Table S5 in [Multimedia Appendix 4](#)). The proportion of participants who believed AI to be highly accurate in diagnosis (question 1) and

treatment planning (question 2) (>80% of the time) increased substantially. More participants also recognized AI's time-saving potential (question 3), with a significant shift toward believing that AI can reduce health care professionals' workload by 50% or more. Additionally, structured AI training (question 4) was perceived as highly beneficial, with a notable rise in participants rating it as significantly effective. Confidence in and willingness to incorporate AI into clinical practice (question 5) also increased, with a greater proportion of participants feeling very willing and confident after the course and informed AI use.

Figure 4. Changes in participants' perceptions of artificial intelligence (AI) before and after the course. Stacked bar charts show participants' responses to 5 perception questions administered before and after the AI training course. Each bar represents the percentage of participants selecting each response category. Dark green indicates very positive perceptions (>80%, significantly, or very willing), light green indicates intermediately positive perceptions (50%-80% or moderately; >50% for question 3), orange indicates conservative perceptions (<50% or slightly), and red represents uncertain or negative responses ("I don't know," "no time saving," or "not confident").



No significant difference in perception shift was observed based on participants' sex, specialty, type of patients served, or previous AI training. However, participants who were unfamiliar with AI before the course exhibited a greater shift in their perception of AI diagnostic accuracy ($P=.01$) and greater willingness and confidence to incorporate AI tools into clinical practice ($P<.001$; Table S6 in [Multimedia Appendix 4](#)).

Discussion

Principal Findings

This study aimed to examine the effect of informed AI use, following a tailored course, on test-based clinical decision-making and the perceptions of GPs and internists. We have shown that informed AI use was associated with higher

clinical decision-making scores using clinical scenarios that simulated daily practice. Furthermore, participants reported higher perceived AI accuracy as a diagnostic and therapeutic assistant, increased perceived time-saving efficiency, a stronger belief in the need for structured AI training programs, and greater confidence in incorporating AI into clinical practice.

Our study has several strengths and novel elements. It included structured, tailored AI training; assessed both objective changes in competence and subjective shifts in perceptions; and specifically targeted GPs and internists, a group previously underrepresented in AI research [19]. Additionally, it uniquely bridged education and objective assessments through structured precourse and postcourse tests. Moreover, this study was multinational and large scale, and it covered a comprehensive

range of clinical competencies rather than a single isolated skill, addressing key limitations cited in previous research [14]. However, because participation was voluntary and completion required finishing both assessments, the analyzed sample likely represents a more AI-motivated subset of eligible registrants, which may limit the generalizability of the observed effects.

Furthermore, this study directly addresses a widely recognized training gap in medical education. Oftring et al [20] emphasized that the growing number and impact of medical AI applications necessitate more AI-focused curricula and research on their educational impact, as most current practitioners and trainees remain underprepared for AI integration into clinical practice. Similarly, Ichikawa et al [21] found that US Colleges of Osteopathic Medicine largely lack AI policy guidance or training for students and faculty, highlighting the urgent need to implement need-driven training in their programs.

The observed improvement in competence scores (57%-78%) is substantial. Whether AI assistance alone, without structured training, can reliably enhance physicians' performance remains debated. AI has been shown to improve diagnostic accuracy in specific tasks such as detecting breast lesions on ultrasound [22], increasing fracture detection sensitivity [23], and interpreting pediatric radiographs [24]. However, Goh et al [25], in a study closest to ours in terms of participant specialties, found that unguided access to LLMs did not improve overall diagnostic reasoning among family medicine, internal medicine, or emergency medicine physicians.

Overall, the substantial improvement observed in our study is unlikely to be attributable to AI use alone, separately from the tailored course. Our training course focused on informed use of AI as a decision-support tool, helping participants double-check their reasoning, consider broader differentials, interact with AI-generated answers, and retrieve relevant literature and clinical guidelines, ultimately enhancing their responses to clinical scenarios. In addition, the substantial improvement in scores cannot be attributed to increased test-taking time, as completion time increased only slightly after the course, indicating that GPs and internists became more accurate with minimal impact on efficiency.

In contrast, a large-scale study by Yu et al [26] involving 140 radiologists and including training on AI use reported heterogeneous effects of AI assistance across 15 chest X-ray diagnostic tasks. Although the study specified that physicians received onboarding training on the AI system, it did not provide details on its content or structure. It is possible that a more clearly structured and tailored training approach, such as the one implemented in our study, might have produced more consistent improvements.

An interesting finding was that GPs demonstrated significantly greater improvement in test scores than internists (23.7% vs 13.7%, respectively). Given that baseline familiarity with AI was comparable between both groups, this discrepancy may be partly explained by the significantly lower baseline scores among GPs, which allowed a wider margin for improvement. GPs were significantly younger than internists, and younger generations are digital learners, engage more frequently in online communities [27], and use technology more fluently than older

cohorts [28]. This interpretation also aligns with our findings of greater score improvement among younger participants and reflects the relatively young age of our study cohort (median 31.00 y).

The observation of no significant difference in participant score improvement based on familiarity with AI use in clinical practice aligns with findings reported by Yu et al [26] among radiologists, in which familiarity with AI tools failed to reliably predict the impact of AI assistance.

One surprising finding was the increased time per correct answer across all 3 competencies (diagnosis, treatment planning, and patient counseling) after AI use. This may reflect more deliberate and reflective reasoning, as participants were instructed to verify AI-generated suggestions and references against their own clinical judgment rather than accept outputs uncritically. Additionally, many participants were new to structured AI use; most (308/326, 94.5%) had no previous AI training, and approximately two-thirds (208/326, 63.8%) reported little to no previous AI use. Therefore, a short adaptation period was anticipated during their first guided application of AI in case-based testing.

The crossover design minimized familiarity bias and was validated by demonstrating no significant difference in precourse scores, postcourse scores, score improvements, and proportions of pass versus fail between the 2 test-order groups. This approach enhanced internal validity by enabling each participant to serve as their own control.

Beyond objective competence improvements, the observed changes in participants' perceptions of AI's accuracy, time efficiency, and potential integration into clinical practice emphasize the importance of structured training and guided AI use in realistic clinical scenarios. Following the intervention, the percentage of participants who perceived high accuracy (>80%) of optimally used AI platforms in diagnosis and treatment planning domains nearly tripled. This finding aligns with Abbas et al [7], who reported comparable accuracy of GPT models on National Board of Medical Examiners clinical subject examination questions.

The notable postcourse increase in physicians' willingness and confidence to incorporate AI into clinical practice reflects one of the most impactful outcomes of this study, suggesting the potential value of structured AI training. The systematic review by Tun et al [17] identified training and familiarity as perceived facilitators of clinician trust in AI tools in making informed clinical decisions. Our data empirically demonstrate this effect. If health professionals lack sufficient trust in AI tools, they may disregard their recommendations, limiting the potential to enhance patient outcomes and optimize clinical workflows [21]. Through its tailored AI course, this study aimed to help participants understand AI capabilities and limitations, enabling them to integrate its assistance beneficially neither erroneously ignoring its outputs nor following them uncritically.

A novel aspect of our study is its pre-post assessment of attitude changes following the intervention, which, to our knowledge, has not been previously applied among physicians in the context of AI integration. Our study uniquely provides evidence linking

structured training and competence improvement with enhanced perception of AI, highlighting the potential of targeted educational interventions to promote informed AI acceptance in clinical practice.

Implications and Future Directions

Our study highlights the potential of informed AI use after a tailored AI training course to enhance GPs' and internists' diagnostic and management skills. Despite the Federation of State Medical Boards advocating for AI competence in medical education, structured AI training remains largely absent [19]. Similarly, 82% of medical professionals in a survey by Tezpal et al [29] recognized the need for AI education. To address this gap, AI training should be integrated into medical education, residency programs, and continuing medical education to ensure responsible and effective AI use. Given the rapid evolution of AI technologies, ongoing updates and refresher courses are essential.

Future research should investigate whether these test-based competence gains are sustained over time and whether they translate into measurable improvements in patient outcomes and clinical decision quality. Finally, further research should explore customized AI training for specialties such as emergency and critical care medicine, given the crucial role of timely and accurate decisions in these specialties.

Limitations

Our study design does not allow isolation of the effect of the tailored AI training course from AI use without training, as it

did not include a nonintervention control arm. Adding a third arm comparing informed versus noninformed AI use was impractical due to expected low motivation and completion rates among participants asked to repeat time-intensive tests without the educational incentive. A similar approach was taken in previous research assessing AI assistance among 140 radiologists, in which onboarding training was not isolated from AI impact [26].

Our participants were already inclined toward AI adoption, as evidenced by their willingness to enroll in our time-intensive study. This suggests that their baseline openness to AI was likely higher than that of physicians more broadly. Thus, the observed shift may underestimate the potential impact of informed AI use based on a tailored AI training course.

Conclusions

This study provides evidence that informed AI use, guided by a tailored AI course, is associated with higher clinical decision-making competence among GPs and internists in test-based clinical scenarios. Beyond higher scores across diagnosis, treatment planning, and patient counseling domains, participants reported higher perceived AI diagnostic and therapeutic accuracy, greater perceived time efficiency, and increased willingness and confidence to integrate AI as a decision-support tool into clinical practice. These findings support the integration of structured AI training into medical education and continuing professional development to facilitate informed and responsible AI use, with plausible potential to improve patient care.

Acknowledgments

This research was conducted as part of the Research Mentorship Program of MedOne Academy. The authors thank Ms Sara M Chenini from MedOne Academy for her continuous support. The senior author thanks Kamel Hatahet, nephrology consultant, North Kansas City Hospital, and Suleiman Al Ashi, MD, endocrinology, diabetes, and metabolism fellow, for their revision of the test questions.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this paper.

Data Availability

All data generated or analyzed during this study are included in this published paper and its multimedia appendices.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Classification of previous artificial intelligence use.

[[DOCX File , 16 KB - mededu_v12i1e75534_app1.docx](#)]

Multimedia Appendix 2

Recruitment of participants.

[[DOCX File , 16 KB - mededu_v12i1e75534_app2.docx](#)]

Multimedia Appendix 3

Clinical case scenarios, assessment questions, and source references.

[[DOCX File , 28 KB - mededu_v12i1e75534_app3.docx](#)]

Multimedia Appendix 4

Additional statistical tables and supporting results.

[[DOCX File , 30 KB - mededu_v12i1e75534_app4.docx](#)]

Multimedia Appendix 5

Passing score rationale.

[[DOCX File , 16 KB - mededu_v12i1e75534_app5.docx](#)]

Multimedia Appendix 6

Artificial intelligence use instructions and technical setup.

[[DOCX File , 16 KB - mededu_v12i1e75534_app6.docx](#)]

Multimedia Appendix 7

Evaluation of participants' perceptions of artificial intelligence accuracy, time efficiency, training needs, and clinical adoption.

[[DOCX File , 23 KB - mededu_v12i1e75534_app7.docx](#)]

Multimedia Appendix 8

Course structure and content.

[[DOCX File , 24 KB - mededu_v12i1e75534_app8.docx](#)]

References

1. Alowais SA, Alghamdi SS, Alsuhbany N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ* 2023 Sep 22;23(1):689 [[FREE Full text](#)] [doi: [10.1186/s12909-023-04698-2](#)] [Medline: [37740191](#)]
2. Meng X, Yan X, Zhang K, Liu D, Cui X, Yang Y, et al. The application of large language models in medicine: a scoping review. *iScience* 2024 May 17;27(5):109713 [[FREE Full text](#)] [doi: [10.1016/j.isci.2024.109713](#)] [Medline: [38746668](#)]
3. 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 [[FREE Full text](#)] [doi: [10.1111/medu.15402](#)] [Medline: [38639098](#)]
4. McGinity C, Clarke EL, Jennings C, Matthews G, Cartlidge C, Freduah-Agyemang H, et al. Artificial intelligence in digital pathology: a systematic review and meta-analysis of diagnostic test accuracy. *NPJ Digit Med* 2024 May 04;7(1):114 [[FREE Full text](#)] [doi: [10.1038/s41746-024-01106-8](#)] [Medline: [38704465](#)]
5. Rutledge GW. Diagnostic accuracy of GPT-4 on common clinical scenarios and challenging cases. *Learn Health Syst* 2024 Jul 25;8(3):e10438 [[FREE Full text](#)] [doi: [10.1002/lrh2.10438](#)] [Medline: [39036534](#)]
6. Penny P, Bane R, Riddle V. Advancements in AI medical education: assessing ChatGPT's performance on USMLE-style questions across topics and difficulty levels. *Cureus* 2024 Dec;16(12):e76309. [doi: [10.7759/cureus.76309](#)] [Medline: [39850172](#)]
7. 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 [[FREE Full text](#)] [doi: [10.7759/cureus.55991](#)] [Medline: [38606229](#)]
8. Nakaura T, Ito R, Ueda D, Nozaki T, Fushimi Y, Matsui Y, et al. The impact of large language models on radiology: a guide for radiologists on the latest innovations in AI. *Jpn J Radiol* 2024 Jul 29;42(7):685-696. [doi: [10.1007/s11604-024-01552-0](#)] [Medline: [38551772](#)]
9. van de Sande D, Chung EF, Oosterhoff J, van Bommel J, Gommers D, van Genderen ME. To warrant clinical adoption AI models require a multi-faceted implementation evaluation. *NPJ Digit Med* 2024 Mar 06;7(1):58 [[FREE Full text](#)] [doi: [10.1038/s41746-024-01064-1](#)] [Medline: [38448743](#)]
10. Brereton TA, Malik MM, Lifson M, Greenwood JD, Peterson KJ, Overgaard SM. The role of artificial intelligence model documentation in translational science: scoping review. *Interact J Med Res* 2023 Jul 14;12:e45903 [[FREE Full text](#)] [doi: [10.2196/45903](#)] [Medline: [37450330](#)]
11. Maleki Varnosfaderani S, Forouzanfar M. The role of AI in hospitals and clinics: transforming healthcare in the 21st century. *Bioengineering (Basel)* 2024 Mar 29;11(4):337 [[FREE Full text](#)] [doi: [10.3390/bioengineering11040337](#)] [Medline: [38671759](#)]

12. Herter WE, Khuc J, Cinà G, Knottnerus BJ, Numans ME, Wiewel MA, et al. Impact of a machine learning-based decision support system for urinary tract infections: prospective observational study in 36 primary care practices. *JMIR Med Inform* 2022 May 04;10(5):e27795 [FREE Full text] [doi: [10.2196/27795](https://doi.org/10.2196/27795)] [Medline: [35507396](https://pubmed.ncbi.nlm.nih.gov/35507396/)]
13. Khosravi M, Zare Z, Mojtabaeian SM, Izadi R. Artificial intelligence and decision-making in healthcare: a thematic analysis of a systematic review of reviews. *Health Serv Res Manag Epidemiol* 2024;11:23333928241234863 [FREE Full text] [doi: [10.1177/23333928241234863](https://doi.org/10.1177/23333928241234863)] [Medline: [38449840](https://pubmed.ncbi.nlm.nih.gov/38449840/)]
14. Feigerlova E, Hani H, Hothersall-Davies E. A systematic review of the impact of artificial intelligence on educational outcomes in health professions education. *BMC Med Educ* 2025 Jan 27;25(1):129 [FREE Full text] [doi: [10.1186/s12909-025-06719-5](https://doi.org/10.1186/s12909-025-06719-5)] [Medline: [39871336](https://pubmed.ncbi.nlm.nih.gov/39871336/)]
15. Lambert SI, Madi M, Sopka S, Lenes A, Stange H, Buszello CP, et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. *NPJ Digit Med* 2023 Jun 10;6(1):111 [FREE Full text] [doi: [10.1038/s41746-023-00852-5](https://doi.org/10.1038/s41746-023-00852-5)] [Medline: [37301946](https://pubmed.ncbi.nlm.nih.gov/37301946/)]
16. 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](https://doi.org/10.1370/afm.2887)] [Medline: [36443071](https://pubmed.ncbi.nlm.nih.gov/36443071/)]
17. Tun HM, Rahman HA, Naing L, Malik OA. Trust in artificial intelligence-based clinical decision support systems among health care workers: systematic review. *J Med Internet Res* 2025 Jul 29;27:e69678 [FREE Full text] [doi: [10.2196/69678](https://doi.org/10.2196/69678)] [Medline: [40772775](https://pubmed.ncbi.nlm.nih.gov/40772775/)]
18. Izquierdo-Condoy JS, Arias-Intriago M, Tello-De-la-Torre A, Busch F, Ortiz-Prado E. Generative artificial intelligence in medical education: enhancing critical thinking or undermining cognitive autonomy? *J Med Internet Res* 2025 Nov 03;27:e76340 [FREE Full text] [doi: [10.2196/76340](https://doi.org/10.2196/76340)] [Medline: [41183320](https://pubmed.ncbi.nlm.nih.gov/41183320/)]
19. Liaw W, Hischier B, James CA, Kakadiaris I, Kueper JK, Rahimzadeh V. Bridging the gap: transforming primary care through the artificial intelligence and machine learning for primary care (aim-pc) curriculum. *Ann Fam Med* 2024;22(6):570-571 [FREE Full text] [doi: [10.1370/afm.240537](https://doi.org/10.1370/afm.240537)] [Medline: [39586707](https://pubmed.ncbi.nlm.nih.gov/39586707/)]
20. 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 May 26;11:e65220 [FREE Full text] [doi: [10.2196/65220](https://doi.org/10.2196/65220)] [Medline: [40418795](https://pubmed.ncbi.nlm.nih.gov/40418795/)]
21. Ichikawa T, Olsen E, Vinod A, Glenn N, Hanna K, Lund GC, et al. Generative artificial intelligence in medical education-policies and training at US osteopathic medical schools: descriptive cross-sectional survey. *JMIR Med Educ* 2025 Feb 11;11:e58766 [FREE Full text] [doi: [10.2196/58766](https://doi.org/10.2196/58766)] [Medline: [39934984](https://pubmed.ncbi.nlm.nih.gov/39934984/)]
22. Lai YC, Chen HH, Hsu JF, Hong YJ, Chiu TT, Chiou HJ. Evaluation of physician performance using a concurrent-read artificial intelligence system to support breast ultrasound interpretation. *Breast* 2022 Oct;65:124-135 [FREE Full text] [doi: [10.1016/j.breast.2022.07.009](https://doi.org/10.1016/j.breast.2022.07.009)] [Medline: [35944352](https://pubmed.ncbi.nlm.nih.gov/35944352/)]
23. Meetschen M, Salhöfer L, Beck N, Kroll L, Ziegenfuß CD, Schaarschmidt BM, et al. AI-assisted X-ray fracture detection in residency training: evaluation in pediatric and adult trauma patients. *Diagnostics (Basel)* 2024 Mar 11;14(6):33 [FREE Full text] [doi: [10.3390/diagnostics14060596](https://doi.org/10.3390/diagnostics14060596)] [Medline: [38535017](https://pubmed.ncbi.nlm.nih.gov/38535017/)]
24. Zech JR, Ezuma CO, Patel S, Edwards CR, Posner R, Hannon E, et al. Artificial intelligence improves resident detection of pediatric and young adult upper extremity fractures. *Skeletal Radiol* 2024 Dec;53(12):2643-2651. [doi: [10.1007/s00256-024-04698-0](https://doi.org/10.1007/s00256-024-04698-0)] [Medline: [38695875](https://pubmed.ncbi.nlm.nih.gov/38695875/)]
25. Goh E, Gallo R, Hom J, Strong E, Weng Y, Kerman H, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. *JAMA Netw Open* 2024 Oct 01;7(10):e2440969 [FREE Full text] [doi: [10.1001/jamanetworkopen.2024.40969](https://doi.org/10.1001/jamanetworkopen.2024.40969)] [Medline: [39466245](https://pubmed.ncbi.nlm.nih.gov/39466245/)]
26. Yu F, Moehring A, Banerjee O, Salz T, Agarwal N, Rajpurkar P. Heterogeneity and predictors of the effects of AI assistance on radiologists. *Nat Med* 2024 Mar 19;30(3):837-849 [FREE Full text] [doi: [10.1038/s41591-024-02850-w](https://doi.org/10.1038/s41591-024-02850-w)] [Medline: [38504016](https://pubmed.ncbi.nlm.nih.gov/38504016/)]
27. 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 Dec 11;19(1):460 [FREE Full text] [doi: [10.1186/s12909-019-1891-5](https://doi.org/10.1186/s12909-019-1891-5)] [Medline: [31829208](https://pubmed.ncbi.nlm.nih.gov/31829208/)]
28. Nakagawa K, Yellowlees P. Inter-generational effects of technology: why millennial physicians may be less at risk for burnout than baby boomers. *Curr Psychiatry Rep* 2020 Jul 13;22(9):45 [FREE Full text] [doi: [10.1007/s11920-020-01171-2](https://doi.org/10.1007/s11920-020-01171-2)] [Medline: [32661817](https://pubmed.ncbi.nlm.nih.gov/32661817/)]
29. Tezpal M, Ghosh S, Lalwani R, Yadav J, Yadav AK. Artificial intelligence in health care: a study on perceptions of and readiness for artificial intelligence in health-care professionals. *J Mar Med Soc* 2024 Dec;26(3):376-381. [doi: [10.4103/jmms.jmms_161_23](https://doi.org/10.4103/jmms.jmms_161_23)]

Abbreviations

AI: artificial intelligence

GP: general practitioner

LLM: large language model

MCQ: multiple-choice question

Edited by AH Sapci; submitted 05.Apr.2025; peer-reviewed by GK Gupta, TK Tran, M Vejdanihemmat; comments to author 16.Aug.2025; revised version received 19.Jan.2026; accepted 20.Jan.2026; published 05.Feb.2026.

Please cite as:

Qunaibi EA, Al-Qaaneh AM, Ismail BF, Muhidat HI, Rageh FS, Musallam NA, Fawzy AK

Effectiveness of Informed AI Use on Clinical Competence of General Practitioners and Internists: Pre-Post Intervention Study

JMIR Med Educ 2026;12:e75534

URL: <https://mededu.jmir.org/2026/1/e75534>

doi:10.2196/75534

PMID:

©Eyad A Qunaibi, Ayman M Al-Qaaneh, Baraa F Ismail, Hussam I Muhidat, Farhia S Rageh, Najwa A Musallam, Alaa K Fawzy. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 05.Feb.2026. 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.

Comparing AI-Assisted Problem-Solving Ability With Internet Search Engine and e-Books in Medical Students With Variable Prior Subject Knowledge: Cross-Sectional Study

Ajiith Xavier*, MBBS, MD; Syed Shariq Naeem*, MBBS, MD, DM; Waseem Rizwi, MBBS, MD, PhD; Hiramani Rabha, MBBS, MD

Department of Pharmacology, Jawaharlal Nehru Medical College and Hospital, Aligarh Muslim University, Medical Road, Aligarh, Uttar Pradesh, India

*these authors contributed equally

Corresponding Author:

Syed Shariq Naeem, MBBS, MD, DM

Department of Pharmacology, Jawaharlal Nehru Medical College and Hospital, Aligarh Muslim University, Medical Road, Aligarh, Uttar Pradesh, India

Abstract

Background: Artificial intelligence (AI), particularly large language models (LLMs) such as ChatGPT (OpenAI), is rapidly influencing medical education. Its effectiveness for students with varying levels of prior knowledge remains underexplored.

Objective: This study aimed to evaluate the performance of medical students with and without formal pharmacology knowledge when using AI-LLM GPTs, internet search engines, e-books, or self-knowledge to solve multiple-choice questions (MCQs).

Methods: A cross-sectional study was conducted at a tertiary care teaching hospital with 100 medical students, divided into a “naive” group (n=50; no pharmacology training) and a “learned” group (n=50; completed pharmacology training). The study was started after approval from the Institutional Ethics Committee of Jawaharlal Nehru Medical College Hospital, Aligarh Muslim University (1018/IEC/23/8/23). Each participant answered 4 sets of 20 MCQs using self-knowledge, e-books, Google, or ChatGPT-4o. Scores were compared using analysis of covariance with self-knowledge scores as a covariate.

Results: Learned students significantly outperformed naive students across all methods ($P<.001$), with the largest effect size in the AI-LLM GPT set (partial $\eta^2=0.328$). For both groups, the performance hierarchy was AI-LLM GPT > internet search engine > self-knowledge ≈ e-books. Notably, the naive students who used AI scored higher (mean 13.24, SD 3.31) than the learned students who used Google (mean 12.14, SD 2.01; $P=.01$) or e-books (mean 10.22, SD 3.12; $P<.001$).

Conclusions: AI-LLM GPTs can significantly enhance problem-solving performance in MCQ-based assessments, particularly for students with limited prior knowledge, even allowing them to outperform knowledgeable peers using traditional digital resources. This underscores the potential of AI to transform learning support in medical education, although its impact on deep learning and critical thinking requires further investigation.

(*JMIR Med Educ* 2026;12:e81264) doi:[10.2196/81264](https://doi.org/10.2196/81264)

KEYWORDS

artificial intelligence; AI; large language models; LLM; medical education; ChatGPT; cognitive performance; subject-naive learners

Introduction

Artificial intelligence (AI) is the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings [1]. This includes the ability to reason, discover meaning, generalize, and learn from past experiences [2]. Since the development of digital computers in the 1940s, it has been demonstrated that computers can be programmed to perform extremely complex tasks, such as proving mathematical theorems or playing chess, with great proficiency [3]. AI involves the science and engineering of creating systems capable of performing tasks that require

humanlike intelligence, including learning, judgment, and decision-making [4]. AI has successfully solved complex problems in various domains, including education. The application of AI in natural language processing has led to the creation of intelligent chatbots and virtual assistants capable of understanding and producing human language [5].

Popenici and Kerr [6] investigated the impact of AI systems on learning and teaching, highlighting potential conflicts between students and instructors, such as privacy concerns, changes in power structures, and excessive control. These studies have called for further research into the impact of AI systems on

learner-instructor interactions to identify gaps, issues, or barriers that prevent AI systems from achieving their intended potential [7].

In the modern era of education, technology has become an integral part of the learning process. The advent of advanced language models such as ChatGPT (OpenAI), along with the vast availability of information on platforms such as Google and e-books, enables educators to enhance students' educational experiences [7].

Medical education demands not only the acquisition of factual knowledge but also the development of critical thinking and decision-making skills [8]. However, the effectiveness of AI-large language model (LLM) GPT-driven tools in enhancing these competencies remains underexplored, particularly concerning the prior knowledge of learners [9]. Students with strong foundational knowledge can use AI tools to deepen their understanding and sharpen their reasoning. In contrast, those with less prior knowledge may use them to fill gaps in their learning or face some difficulties in comprehension [10].

This study used ChatGPT-4o as the representative AI-LLM GPT tool for 2 reasons. First, ChatGPT was widely accessible to students at the time of study and required minimal technical expertise. Second, it represents the current generation of general-purpose LLMs that students are likely to encounter in real-world educational settings. The most notable advantage is its breadth of training data, which expanded from 175 billion parameters in GPT-3.5 to 1 trillion in the GPT-4.0 model [11]. Bing's current AI model has 175 billion parameters, and Microsoft's Bard has 540 billion in comparison [12,13]. The domain-specific LLMs, enriched with pharmacology databases or specialized medical AI agents, may offer better performance characteristics, but widespread use of such tools is still limited due to paywall and validation issues.

The aim of this study is to compare the effectiveness of AI-LLM GPT tools (ChatGPT-4o), internet search engines (Google), e-books, and self-knowledge for answering pharmacology multiple-choice questions (MCQs) among medical students with different levels of prior pharmacology knowledge (naive vs learned groups). The study explores the assistance provided by AI-LLM GPT tools in the ability of problem solving using a cross-sectional design.

Methods

Ethical Considerations

This was a cross-sectional study conducted at a tertiary care teaching hospital. The study was started after approval from the Institutional Ethics Committee of Jawaharlal Nehru Medical College Hospital, Aligarh Muslim University (1018/IEC/23/8/23). All participants were adult MBBS students and provided written informed consent after receiving information about the study purpose and procedures. Data were collected anonymously using coded identifiers. No personal or

identifiable information was recorded or published. Participants received no monetary or academic compensation.

Setting and Participants

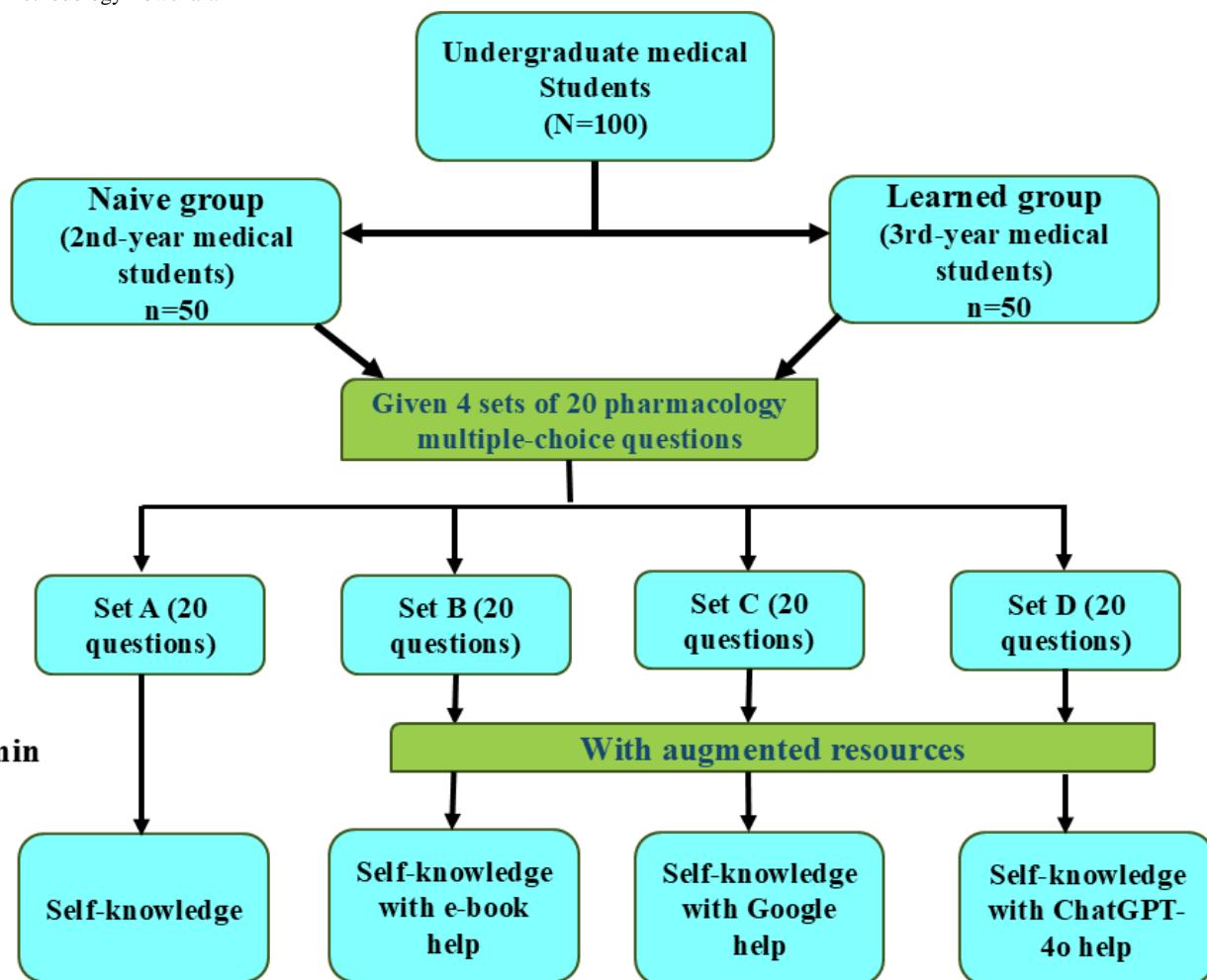
Participants included 50 second-year medical students who had just joined the second year and 50 third-year students who had recently passed their second year (ie, N=100). The second-year students, designated as the "naive group," had no prior formal exposure to pharmacology coursework. The third-year students, designated as the "learned group," had completed a comprehensive 1-year pharmacology curriculum during their second year. This curriculum consisted of 80 hours of didactic lectures, 150 hours of interactive teaching, including practical laboratory sessions, case-based learning modules, self-directed learning modules, short group teaching, and regular assessments, including sessional and professional examinations. Students had completed this coursework approximately 2 - 3 months prior to participation in this study. The curriculum covered topics such as general pharmacology, autonomic pharmacology, cardiovascular drugs, and antimicrobials, aligned with the medical council syllabus for undergraduate medical education. All learned group participants had passed their second-year pharmacology examination with a minimum score of 50%.

Study Development

Prior to the assessment, all participants received a 15-minute standardized orientation session on using ChatGPT and were provided with an eBook of Pharmacology. This training covered basic prompt formulation and how to ask follow-up questions. Students accessed ChatGPT through computer workstations in a controlled examination hall environment to ensure standardized conditions. For the Google search set, students used desktop computers with standard Google Chrome or Mozilla Firefox browsers. For the e-book set, students were provided with PDF versions of a pharmacology textbook accessible on the same computers. The training session was conducted using a standardized presentation to maintain homogeneity of assistance tool exposure. The self-knowledge set required no external tools. All sessions were proctored to ensure compliance with assigned methods for each question set.

The 80 MCQs were divided into four parallel sets (A, B, C, and D), each containing 20 questions. To ensure equivalence across sets, questions were matched for content area coverage (eg, autonomic, cardiovascular, and chemotherapy) [1], question type (theory, calculation, or image-based) [2], and estimated difficulty level based on faculty assessment [3].

Students had 15 minutes to respond to each set of questions. They were instructed to respond to the questions in a manner that self-knowledge was used to answer questions in set A; e-books were used to answer questions in set B; internet search engine (Google) was used to answer questions in set C; and AI-LLM GPT (ChatGPT) was used to answer questions in set D. Each correct answer was given 1 mark, and no negative marking was given. The methodology flowchart of this study is shown in [Figure 1](#).

Figure 1. Methodology flowchart.

Statistical Analysis

Data analysis was done using SPSS (version 23; IBM Corp) and R software package (R Foundation for Statistical Computing). The mean (the average number of correctly answered questions out of the 20 MCQs included in each augmentation set such as self-knowledge, e-book, internet search engine, and GPT-4o) between different groups was assessed using the independent test. Intergroup and intragroup comparisons were conducted using analysis of covariance (ANCOVA), with self-knowledge group (set A) being a covariate. A *P* value *<.05* was considered to indicate statistical significance.

Results

The intergroup comparison between the learned and naive student groups was analyzed using ANCOVA. The estimated marginal means, partial η^2 values, and statistical significance for each group are presented in Table 1. The self-knowledge set was used as a covariate in the analysis to control for baseline differences in knowledge across the sets. This approach allowed for a more accurate assessment of the effects of the augmented resources methods (ChatGPT-4o, Google, and e-books) on student performance.

Table . Intergroup comparison between the learned and naive using analysis of covariance (N=100).

Augmentation resource and group	Score, mean (SD)	Estimated marginal mean	Partial η^2	P value
AI ^a -LLM ^b GPT (ChatGPT-4o)			0.328	<.001
Naive	13.24 (3.31)	13.18		
Learned	15.42 (1.89)	15.48		
Internet search engine (Google)			0.241	<.001
Naive	10.82 (2.13)	10.76		
Learned	12.14 (2.01)	12.20		
e-Books			0.195	<.001
Naive	5.82 (2.77)	5.78		
Learned	10.22 (3.12)	10.26		

^aAI: artificial intelligence.

^bLLM: large language model.

The analysis of estimated marginal means revealed that the learned students consistently outperformed the naive students across all sets. In the AI-LLM GPT (ChatGPT-4o) set, the learned students achieved an estimated marginal mean of 15.48 compared to 13.18 for the naive students, with a large effect size (partial $\eta^2=0.328$, $P<.001$). In the internet search engine (Google) set, the estimated marginal mean for the learned students was 12.20, whereas the naive students scored 10.76, with a moderate effect size (partial $\eta^2=0.241$; $P<.001$). The learned students in the e-book set had an estimated marginal mean of 10.26, whereas the naive students had 5.78, a moderate effect size (partial $\eta^2=0.195$; $P<.001$).

The post hoc analysis for the naive group shows that the AI-LLM GPT-assisted method resulted in the highest performance when compared to other methods like internet search engine-based learning, self-knowledge, and e-books methods. A significant difference was found in all pairwise comparisons ($P<.001$) except for e-books versus self-knowledge, which was not statistically significant (mean difference=−0.30; $P=.91$). The performance hierarchy observed was AI-LLM GPT versus self-knowledge ($P<.001$) > internet search engine versus self-knowledge ($P<.001$) > self-knowledge ≈ e-books ($P=.91$), showing the superior effectiveness of AI tools for naive learners.

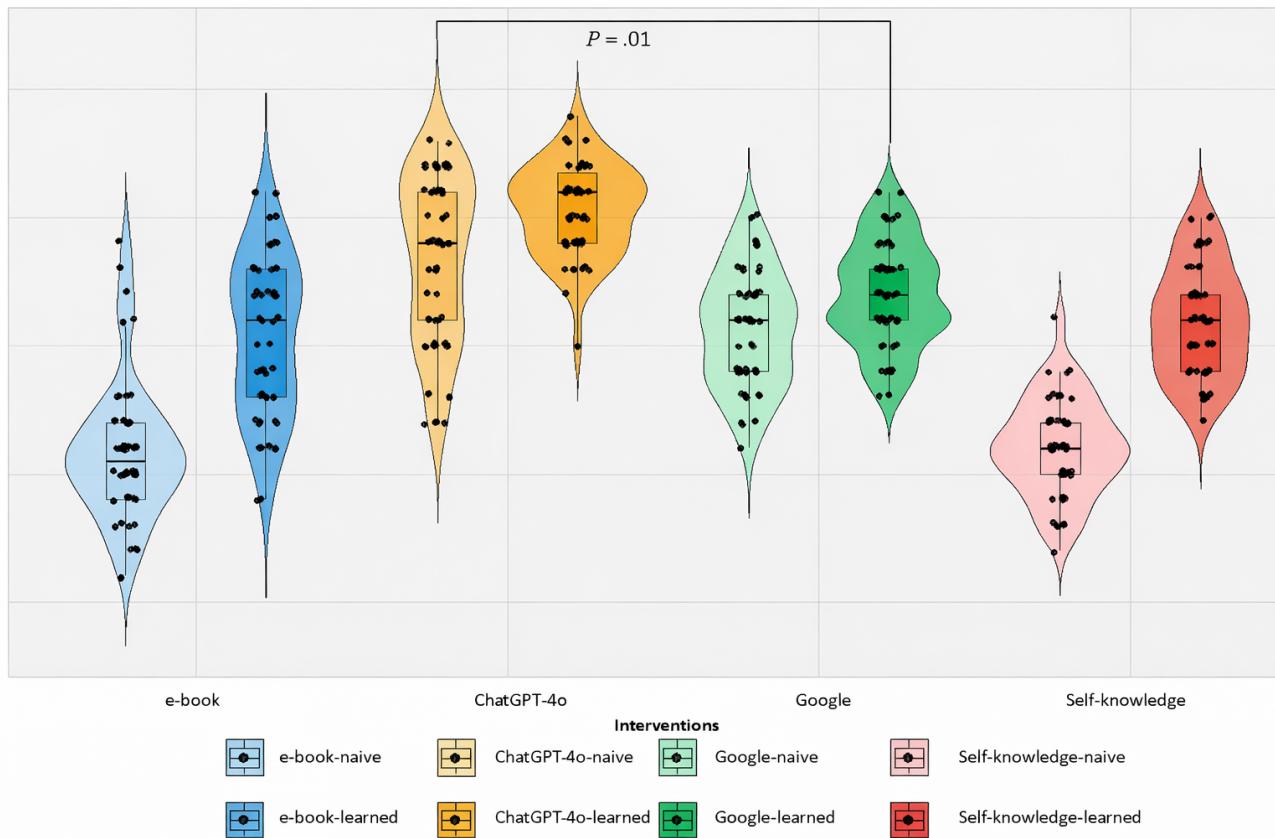
The post hoc analysis for the learned group shows that the AI-LLM GPT-assisted learning method resulted in the highest performance when compared to other methods, whereas self-knowledge and e-books methods showed the lowest and statistically similar outcomes (mean difference=−0.72, $P=.33$). Significant differences were observed between Google and self-knowledge sets ($P=.04$), ChatGPT-4o and self-knowledge sets ($P<.001$), Google and e-books sets ($P<.001$), ChatGPT-4o and e-books ($P<.001$), and ChatGPT-4o and Google ($P<.001$),

indicating a superior performance across these comparisons. The overall performance hierarchy observed was: AI-LLM GPT > internet search engine > self-knowledge ≈ e-books, indicating the superior effectiveness of AI-LLM GPT in enhancing learning outcomes, even for participants with prior knowledge (Multimedia Appendix 1).

The mean comparison between the sets, as presented in Figure 2, highlights that learned students consistently outperformed naive students across all categories, as expected. The AI-LLM set achieved the highest performance, with a mean score of 15.3 (SD 1.776). Conversely, the lowest mean was observed in the e-books set, in which learned students scored 10.3 (SD 3.1) and naive students scored 5.7 (SD 2.7).

Interestingly, the self-knowledge sets performed better than the e-books sets, with naive group students achieving a mean score of 6.0 (SD 1.9) and learned group students scoring a mean of 10.96 (SD 2.2). This may be due to the fact that students in the e-book sets spent sufficient time searching for relevant chapters instead of focusing on answering the questions. This led to some questions being left unanswered that likely contributed to their lower overall performance.

While ANCOVA analysis confirmed that the learned students consistently outperformed the naive students when comparing the same learning method (as shown in Table 1), an important cross-method comparison emerged: the naive students who used ChatGPT-4o achieved higher scores (mean 13.2, SD 3.3) than the learned students who used self-knowledge (mean 10.96, SD 2.5), e-books (mean 10.22, SD 3.12), or Google (mean 12.14, SD 2.01). This demonstrates that the advantage conferred by using ChatGPT-4o was sufficient to overcome the knowledge gap between the naive and learned students when the latter were restricted to conventional resources.

Figure 2. Box-violin plot showing the mean scores in the naive and learned groups.

Discussion

Principal Findings

AI is changing the face of medical education, as it suggests some new solutions to the old forms of teaching, increases the educational experience, and improves the results of students [14,15]. AI-based solutions, including LLMs, such as ChatGPT, virtual assistants, and AI-based simulation technologies, can change how medical students learn and apply their knowledge [16,17].

This study found that learned students consistently outperformed naive students across all learning methods. However, the naive students who used ChatGPT-4o achieved higher scores than the learned students who used conventional resources (ie, self-knowledge, e-books, or Google), suggesting that AI-LLM tools may help bridge knowledge gaps more effectively than traditional learning resources.

Our study demonstrates that AI-powered LLMs such as ChatGPT can notably improve students' performance on pharmacology MCQs, particularly for those without prior subject knowledge. Although our findings show that students achieved higher MCQ test scores when using ChatGPT-4o, we cannot conclude that this represents a genuine improvement in cognitive performance, critical thinking, or clinical decision-making skills. Higher scores may reflect the AI's ability to rapidly retrieve and present relevant information rather than deeper student understanding. The findings indicated that learned students have outperformed naive students in all 4 sets, and AI-LLM GPT recorded the best performance. The greatest effect size was

observed in the AI group (partial $\eta^2=0.328$), which indicates the tremendous influence of AI-LLM in ensuring the students improve their understanding and problem-solving skills. These results are consistent with those of previous research in that AI tools may enhance critical thinking and support learning under the condition that they are properly used by a student who already has a good knowledge grounding [18].

As opposed to other methods, the e-books set exhibited the lowest performance. This can be explained by the fact that it takes a lot of time to navigate and search relevant information in the e-books. This reveals the difficulties of self-learning with the help of conventional digital sources, especially in the case of naive students who are probably not able to manage their time efficiently. In contrast, the self-knowledge set outperformed the e-books set, which means that dependence on inherent knowledge, although restricted to a certain extent, can be more helpful than the effective utilization of external means.

Although our quantitative data demonstrate that ChatGPT-4o enhanced student performance, particularly for naive learners, the mechanisms underlying this effectiveness warrant deeper exploration. There may be several reasons for AI-LLM GPT (ChatGPT-4o) to outperform other methods of assistance.

First, ChatGPT-4o provides immediate, targeted responses without requiring students to navigate complex search results or textbook indices. Informal feedback from participants suggested that students who used Google often struggled to identify authoritative sources among numerous search results, whereas those who used e-books spent considerable time locating relevant chapters. In contrast, ChatGPT-4o delivered

direct answers to questions, reducing the cognitive load related to information seeking.

Second, ChatGPT-4o's conversational interface may facilitate iterative learning. Students could ask follow-up questions to clarify concepts, request examples, or seek a specific explanation for a specific MCQ—a dynamic interaction not possible with static e-books and more efficient than reformulating multiple Google searches.

Third, the structured, synthesized format of ChatGPT-4o responses may be particularly beneficial for naive learners who lack the prior knowledge framework to evaluate and integrate fragmented information from multiple sources. ChatGPT-4o essentially predigests information, whereas Google and e-books require students to perform this synthesis themselves—a task that may be especially challenging without foundational knowledge.

However, these remain as hypotheses. Future research should use mixed methods approaches, including think-aloud protocols during tool use, posttask interviews exploring student decision-making processes, and screen recordings analyzing search strategies and information evaluation patterns. Such qualitative data would provide richer insights into how students with different knowledge levels interact with various resources and why certain tools prove more effective. The study of recognizing the pattern of wrong responses in the AI-LLM GPTs output may also generate an “array of errors” for further training and improving LLMs.

These findings have substantial implications for both local and global medical education practices. In our institutional context, where students face high patient loads and limited access to senior clinicians for immediate consultation, AI-LLM GPTs could serve as readily available reference tools to support clinical decision-making during training. However, integration must be carefully structured to enhance rather than replace foundational learning.

Globally, these results are particularly relevant for resource-limited settings where access to comprehensive textbooks, updated references, and expert faculty may be constrained. AI-LLM GPTs could help democratize access to medical knowledge, potentially reducing disparities between well-resourced and under-resourced educational institutions. However, this assumes reliable internet connectivity and technological infrastructure, which remain barriers in many settings.

From a pedagogical perspective, our findings suggest a paradigm shift may be necessary in how we structure medical curricula. Rather than focusing exclusively on memorization of facts—information that AI can rapidly retrieve—educational programs should prioritize teaching students how to (1) formulate effective questions and search strategies, (2) critically evaluate AI-generated responses for accuracy and clinical appropriateness, (3) integrate AI-provided information with

clinical context and patient-specific factors, and (4) develop metacognitive skills to recognize the limitations of both their own knowledge and AI tools.

Importantly, although AI-LLMs demonstrated effectiveness for MCQ performance, medical practice requires competencies beyond factual knowledge, including physical examination skills, procedural expertise, empathetic patient communication, ethical reasoning under uncertainty, and team-based care coordination. Educational programs must ensure that AI integration enhances rather than diminishes these essential human dimensions of medical practice.

An important limitation is our focus on general-purpose ChatGPT-4o rather than specialized medical or pharmacology-specific AI agents. Domain-enriched LLMs that integrate pharmacology textbooks or medical databases may demonstrate different performance characteristics and potentially offer more accurate, context-specific responses. However, our choice reflects the current reality that most students have access to general LLMs such as ChatGPT-4o rather than specialized medical AI tools and makes the study more generalizable than using domain-enhanced LLMs. Second, the cross-sectional, single-day design prevents assessment of long-term learning outcomes, knowledge retention, or the development of clinical reasoning skills. Similarly, the sequential exposure to different resources (sets A through D) may have introduced order effects or learning transfer between sets. To minimize the learning transfer, the questions have been chosen from diverse topics and moderated to remove any repetition. However, the order effect cannot be removed completely in these types of cross-sectional-sequential studies.

AI will be able to deliver personalized learning and adjust to the needs of students. This individualization helps the learners make their progress through their priorities, repeating difficult concepts or working through material as they become familiar with each subject [19,20]. The AI-powered platforms can monitor student performance in real-time and can track the areas where the students might need more education or practice [21]. Such benefits help achieve the speed and the depth of knowledge acquisition with a higher level of efficiency and more targeted learning [22,23].

Conclusions

To conclude, this study shows that the learned group outperformed the naive group in all sets, including AI-LLM GPT, internet search engine, e-books, and self-knowledge in problem solving. Ironically, the AI-LLM GPT naive group outperformed the problem-solving skills of even the learned group augmented with an internet search engine, showcasing the disruptive potential of AI-LLM in medical education. Further longitudinal studies examining knowledge retention, clinical reasoning development, and the ability to solve novel problems without AI assistance are needed to determine whether AI tools genuinely enhance cognitive capabilities or primarily serve as effective reference tools.

Acknowledgments

The authors would like to express their gratitude to all the students who took part in this research. They would also like to thank Dr Binola for their constant motivation throughout the study. All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this paper.

Data Availability

The anonymized datasets generated and analyzed during this study are not publicly available due to institutional restrictions and the need to protect participant confidentiality; however, the data are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Post hoc results of the naive and learned groups.

[[DOCX File, 6001 KB - mededu_v12i1e81264_app1.docx](#)]

References

1. Arawi T, El Bachour J, El Khansa T. The fourth industrial revolution: its impact on artificial intelligence and medicine in developing countries. *Asian Bioeth Rev* 2024 May 25;16(3):513-526. [doi: [10.1007/s41649-024-00284-7](https://doi.org/10.1007/s41649-024-00284-7)] [Medline: [39022373](https://pubmed.ncbi.nlm.nih.gov/39022373/)]
2. Sharma S. Benefits or concerns of AI: a multistakeholder responsibility. *Futures* 2024 Mar;157:103328. [doi: [10.1016/j.futures.2024.103328](https://doi.org/10.1016/j.futures.2024.103328)]
3. Chaudhry MA, Kazim E. Artificial Intelligence in Education (AIEd): a high-level academic and industry note 2021. *AI Ethics* 2022;2(1):157-165. [doi: [10.1007/s43681-021-00074-z](https://doi.org/10.1007/s43681-021-00074-z)] [Medline: [34790953](https://pubmed.ncbi.nlm.nih.gov/34790953/)]
4. Xu LD, Lu Y, Li L. Embedding blockchain technology into IoT for security: a survey. *IEEE Internet Things J* 2021 Jul 1;8(13):10452-10473. [doi: [10.1109/JIOT.2021.3060508](https://doi.org/10.1109/JIOT.2021.3060508)]
5. Supriyono, Wibawa AP, Suyono, Kurniawan F. Advancements in natural language processing: implications, challenges, and future directions. *Telemat Inform Rep* 2024 Dec;16:100173. [doi: [10.1016/j.teler.2024.100173](https://doi.org/10.1016/j.teler.2024.100173)]
6. Popenici SA, Kerr S. Exploring the impact of artificial intelligence on teaching and learning in higher education. *Res Pract Technol Enhanc Learn* 2017;12(1):22. [doi: [10.1186/s41039-017-0062-8](https://doi.org/10.1186/s41039-017-0062-8)] [Medline: [30595727](https://pubmed.ncbi.nlm.nih.gov/30595727/)]
7. Ghory S, Ghafory H. The impact of modern technology in the teaching and learning process. *Int J Innov Res Sci Stud* 2021 Jun 18;4(3):219-227. [doi: [10.53894/ijirss.v4i3.73](https://doi.org/10.53894/ijirss.v4i3.73)]
8. Buja LM. Medical education today: all that glitters is not gold. *BMC Med Educ* 2019 Apr 16;19(1):110. [doi: [10.1186/s12909-019-1535-9](https://doi.org/10.1186/s12909-019-1535-9)] [Medline: [30991988](https://pubmed.ncbi.nlm.nih.gov/30991988/)]
9. Salih S, Husain O, Hamdan M, Abdelsalam S, Elshafie H, Motwakel A. Transforming education with AI: a systematic review of ChatGPT's role in learning, academic practices, and institutional adoption. *Res Eng* 2025 Mar;25:103837. [doi: [10.1016/j.rineng.2024.103837](https://doi.org/10.1016/j.rineng.2024.103837)]
10. Lawasi MC, Rohman VA, Shoreamanis M. The use of AI in improving student's critical thinking skills. *Proc Ser Soc Sci Humanit* 2024 Sep;18:366-370. [doi: [10.30595/pssh.v18i.1279](https://doi.org/10.30595/pssh.v18i.1279)]
11. Martindale J, Bizzaco M. GPT-4 vs GPT-35: how much difference is there? *Digital Trends*. 2023 Apr 4. URL: <https://www.digitaltrends.com/computing/gpt-4-vs-gpt-35> [accessed 2023-07-29]
12. Deutscher M. Microsoft researchers reveal neural network with 135B parameters. *SiliconANGLE*. 2021 Aug 4. URL: <https://siliconangle.com/2021/08/04/microsoft-researchers-reveal-neural-network-135b-parameters> [accessed 2023-07-29]
13. Almekinders S. Google Bard PaLM update: 540 billion parameters take fight to ChatGPT. *Techzine*. 2023 Apr 3. URL: <https://www.techzine.eu/news/applications/104043/google-bard-palm-update-540-billion-parameters-take-fight-to-chatgpt/> [accessed 2025-12-05]
14. Mir MM, Mir GM, Raina NT, et al. Application of artificial intelligence in medical education: current scenario and future perspectives. *J Adv Med Educ Prof* 2023 Jul;11(3):133-140. [doi: [10.30476/JAMP.2023.98655.1803](https://doi.org/10.30476/JAMP.2023.98655.1803)] [Medline: [37469385](https://pubmed.ncbi.nlm.nih.gov/37469385/)]
15. 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](https://doi.org/10.3233/SHTI230581)] [Medline: [37387115](https://pubmed.ncbi.nlm.nih.gov/37387115/)]
16. Mattas PS. ChatGPT: a study of AI language processing and its implications. *Int J Res Publ Rev* 2023;04(2):435-440. [doi: [10.55248/gengpi.2023.4218](https://doi.org/10.55248/gengpi.2023.4218)]
17. Alawida M, Mejri S, Mehmood A, Chikhaoui B, Isaac Abiodun O. A comprehensive study of ChatGPT: advancements, limitations, and ethical considerations in natural language processing and cybersecurity. *Information* 2023;14(8):462. [doi: [10.3390/info14080462](https://doi.org/10.3390/info14080462)]

18. Lijie H, Mat Yusoff S, Mohamad Marzaini AF. Influence of AI-driven educational tools on critical thinking dispositions among university students in Malaysia: a study of key factors and correlations. *Educ Inf Technol* 2025;30:8029-8053. [doi: [10.1007/s10639-024-13150-8](https://doi.org/10.1007/s10639-024-13150-8)]
19. Narayanan S, Ramakrishnan R, Durairaj E, Das A. Artificial intelligence revolutionizing the field of medical education. *Cureus* 2023 Nov 28;15(11):e49604. [doi: [10.7759/cureus.49604](https://doi.org/10.7759/cureus.49604)] [Medline: [38161821](https://pubmed.ncbi.nlm.nih.gov/38161821/)]
20. Jong M. Transforming medical education in the AI era: balancing technological expertise with humanistic care in tomorrow's doctors. *Ann Acad Med Singap* 2024 Dec 12;53(12):754-757. [doi: [10.47102/annals-acadmedsg.2024242](https://doi.org/10.47102/annals-acadmedsg.2024242)] [Medline: [39748174](https://pubmed.ncbi.nlm.nih.gov/39748174/)]
21. Gligoreia I, Cioca M, Oancea R, Gorski AT, Gorski H, Tudorache P. Adaptive learning using artificial intelligence in e-learning: a literature review. *Educ Sci* 2023;13(12):1216. [doi: [10.3390/educsci13121216](https://doi.org/10.3390/educsci13121216)]
22. Ahn HY. AI-powered e-learning for lifelong learners: impact on performance and knowledge application. *Sustainability* 2024;16(20):9066. [doi: [10.3390/su16209066](https://doi.org/10.3390/su16209066)]
23. Alqahtani T, Badreldin HA, Alrashed M, et al. The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Res Social Adm Pharm* 2023 Aug;19(8):1236-1242. [doi: [10.1016/j.sapharm.2023.05.016](https://doi.org/10.1016/j.sapharm.2023.05.016)] [Medline: [37321925](https://pubmed.ncbi.nlm.nih.gov/37321925/)]

Abbreviations

AI: artificial intelligence

ANCOVA: analysis of covariance

LLM: large language model

MCQ: multiple-choice question

Edited by J Gentges; submitted 25.Jul.2025; peer-reviewed by D Chartash, JM Santos-Gago, K Sakaguchi; revised version received 04.Nov.2025; accepted 06.Nov.2025; published 06.Jan.2026.

Please cite as:

Xavier A, Naeem SS, Rizwi W, Rabha H

Comparing AI-Assisted Problem-Solving Ability With Internet Search Engine and e-Books in Medical Students With Variable Prior Subject Knowledge: Cross-Sectional Study

JMIR Med Educ 2026;12:e81264

URL: <https://mededu.jmir.org/2026/1/e81264>

doi: [10.2196/81264](https://doi.org/10.2196/81264)

© Ajith Xavier, Syed Shariq Naeem, Waseem Rizwi, Hiramani Rabha. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 6.Jan.2026. 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.

AI Literacy Among Chinese Medical Students: Cross-Sectional Examination of Individual and Environmental Factors

Chunqing Li¹, PhD; Sian Hsiang-Te Tsuei², MD, CCFP, PhD; Hongbin Wu³, Prof Dr

¹Education Management Information Center, Ministry of Education of the People's Republic of China, Beijing, China

²Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

³Institute of Medical Education, Peking University, No 5 Yiheyuan Rd, Beijing, China

Corresponding Author:

Hongbin Wu, Prof Dr

Institute of Medical Education, Peking University, No 5 Yiheyuan Rd, Beijing, China

Abstract

Background: Artificial intelligence (AI) literacy is increasingly essential for medical students. However, without systematic characterization of the relevant components, designing targeted medical education interventions may be challenging.

Objective: This study aimed to systematically describe the levels of and factors associated with multidimensional AI literacy among Chinese medical students.

Methods: A cross-sectional, descriptive analysis was conducted using data from a nationwide survey of Chinese medical students (N=80,335) across 109 medical schools in 2024. AI literacy was assessed with a multidimensional instrument comprising three domains: knowledge, evaluating students' self-reported proficiency in core areas of medical AI applications; attitude, reflecting their self-perceived views on using AI for teaching and learning; and behavior, capturing the self-perceived usage frequency and application patterns. Multivariate linear regression was applied to examine the associations between individual factors (ie, demographic characteristics, family background, and enrollment motivation) and environmental factors (ie, educational phase, type of education program, and tier of education program) and AI literacy.

Results: Respondents showed moderate to high levels of AI knowledge (mean 76.0, SD 26.9), followed by moderate AI attitude scores (mean 71.6, SD, 24.4). In contrast, AI behavior scores were much lower (mean 32.5, SD, 28.5), indicating little usage of AI tools. Of the individual factors, male students reported higher levels of AI attitude and behavior; both intrinsic and extrinsic motivation were positively associated with all three dimensions; advantaged family background was positively related to AI attitude and behavior, but not knowledge. Among the environmental factors, attending the prestigious Double First-Class universities was positively associated with higher AI usage. Enrollment in long-track medical education programs was associated with higher AI attitude and behavior, while being in the clinical phase was negatively associated with both AI knowledge and behavior. Environmental factors moderated the associations between individual characteristics and AI literacy, potentially attenuating disparities.

Conclusions: Medical students reported moderate to high AI knowledge, moderate AI favorability, and low AI use. Individual characteristics and environmental factors were significantly associated with AI literacy, and environmental factors moderated the associations. The moderate AI literacy overall highlights the need for AI-related medical education, ideally with practical use and nuanced by socioeconomic factors.

(*JMIR Med Educ* 2026;12:e80604) doi:10.2196/80604

KEYWORDS

medical education; artificial intelligence; AI Literacy; medical student; multidimensional constructs; China

Introduction

Artificial intelligence's (AI's) rapid advancement in diagnostic support, treatment decision-making, personalized care, and health system management is quickly reshaping health care service delivery [1-4]. To prepare medical trainees in using these tools [5], medical training programs need to efficiently leverage their limited resources to set up the relevant training approaches and programs to meet the most important needs.

However, instead of identifying such gaps, research has mostly focused on applying AI in medical education across tasks, including, for example, admission, research, and evaluation [6-17]. The developments of AI-targeted curricula and competencies have so far advanced without high-quality evidence of trainees' needs [18-22], highlighting the questionable appropriateness of such proposals.

Among the various AI skills and competencies, clarifying trainees' AI literacy appears especially important. AI literacy

refers to a set of competencies that enable individuals to critically evaluate AI technologies, communicate and collaborate effectively with AI, and use AI as a tool across various contexts [23]. While various frameworks for AI literacy exist, there is growing consensus that it comprises cognitive knowledge, attitudes, and practical behaviors toward AI tools [5,24-26].

Building on this consensus, we defined AI literacy as a multidimensional construct encompassing knowledge, attitude, and behavior, integrating both classical and contemporary theoretical models. Specifically, this tripartite structure aligns with the ABC model of attitudes in social psychology (Affect–Behavior–Cognition), which conceptualizes human responses as comprising cognitive, affective, and behavioral components [27-29].

Current evidence on AI literacy in medical trainees has been hindered by theoretical and methodological deficiencies. First, the theoretical weakness stems from the lack of recognition that AI literacy has multiple dimensions. Most have focused on AI knowledge [30-33] or attitudes toward AI applications [34-38]. This misses the important dimension of AI behaviors [5,24,25,39,40], which account for the actual usage patterns of AI tools. Second, methodologically, the few studies that captured all three dimensions suffered from capturing only a small number of medical schools and sample size [41,42]. This limits their generalizability. Third, the studies often fail to examine the factors associated with subdimensions of AI literacy, which limits the ability to design targeted interventions within a limited budget. Existing studies have often focused mainly on individual-level variables, overlooking environmental factors associated with the digital divide [41,43-45].

We addressed these gaps by drawing on the comprehensive quantitative database of Chinese medical students, the 2024 China Medical Student Survey (CMSS). We report the distribution of AI literacy across cognitive, attitudinal, and behavioral dimensions among 80,355 Chinese medical students and their associated individual and environmental factors. This large sample size across all three dimensions of AI literacy provides a more meaningful representation of the potential educational needs. We further analyzed the potential factors associated with AI literacy, enabling medical educators to design targeted interventions.

Methods

Data Source and Study Sample

Since 2019, China's National Center for Health Professions Education Development has conducted an annual nationwide survey of medical students in China known as the CMSS. It aims to support medical education development by collecting comprehensive information across the entire educational journey, including students' demographic backgrounds, preadmission experiences, academic training, and postgraduation plans [46]. The survey takes place between May and July each year. Additional details on the context of medical education in China and the CMSS are provided in Items 1 and 2 in [Multimedia Appendix 1](#). All data analyses adhered to the

(STROBE) Strengthening the Reporting of Observational Studies in Epidemiology guidelines for cross-sectional studies.

This study used data from the 2024 CMSS, which included participants from 109 medical schools across 28 provinces in China. The survey achieved a 73.9% response rate (80,355/108,710 surveyed) and captured 109 of 202 medical schools (109/202, 53.9%). The geographical distribution of participating institutions roughly reflected the national pattern. Comparing the dataset and the national distribution, respectively, 49.5% (54/109) and 45.0% (91/202) of the schools were from the eastern region, 28.4% (31/109) and 30.7% (62/202) from the central region, and 22.0% (24/109) and 24.3% (49/202) from the western region. As for the tiers of institutions, the data included universities' designation as 'Double First-Class' (DFC) institutions. These institutions garner priority investment in China's bid to establish world-leading educational institutions [47] (see Item 3 in [Multimedia Appendix 1](#) for more details). Further, the proportion of DFC universities was similar in the dataset and nation was (25.7%, 28/109 vs 23.3%, 47/202), respectively.

Ethical Considerations

The project was approved by the Peking University Institutional Review Board (Beijing, China; approval Number: IRB00001052- 20069). Participation was voluntary, and completion and submission of the questionnaire were considered to constitute informed consent. To protect participant privacy and confidentiality, no personally identifiable information was collected, and all responses were analyzed in an anonymized form. The data were stored securely and were accessible only to the research team. Participants did not receive any financial or other compensation for their participation.

Measurement of AI Literacy

AI literacy is a multidimensional construct comprising cognitive, attitudinal, and behavioral dimensions [24,25,40]. In this study, we refer to the cognitive dimension as "AI knowledge" to improve clarity and align with common usage in educational research. AI knowledge was assessed through four items evaluating students' self-perceived proficiency in core areas of medical AI: basic concepts of AI, machine learning tools, multimodal medical data analysis, and ethics issues related to AI. Responses were rated on a three-point scale coded as 1 (low), 2 (medium), and 3 (high) proficiency. Students' attitude towards AI was measured using two items: students' self-perceived view of using AI tools in (1) teaching and (2) learning. Responses were rated on a three-point scale, coded as 1 (negative), 2 (neutral), and 3 (positive), respectively. AI behavior refers to actual use of AI tools rather than behavioral intention and was evaluated through self-perceived usage frequency and application patterns. Usage frequency was assessed on a six-point scale (ranging from "never used" to "multiple times daily," coded one to six, where higher scores indicated more frequent academic use). Usage patterns were measured with two items: extent of using AI for (1) professional knowledge learning and (2) writing tasks (options included "never used," "generating initial drafts," "writing specific sections," "editing and polishing," "outlining," and "brainstorming ideas"). The Cronbach α values for cognitive,

attitudinal, and behavioral dimension were 0.81, 0.86, and 0.79, respectively, indicating high internal consistency. The Kaiser-Meyer-Olkin values were 0.76, 0.70, and 0.75 for the three dimensions. Additionally, Bartlett's test of sphericity was significant ($P<.001$), supporting the construct validity of the measurement scales. Further confirmatory factor analysis showed that the measurement instrument loaded well onto the three identified dimensions of AI literacy. Relevant statistics include comparative fit index (CFI)=0.994, Tucker-Lewis index (TLI)=0.991, Root Mean Square Error of Approximation (RMSEA)=0.025, and statistically significant loading on all factors ($P<.001$).

Factors Associated With AI Literacy

Existing research suggests that both individual and environmental factors associate with access to, and use of digital technologies [45]. For individual factors, we examined demographic characteristics, family background, intellectual ability, and enrollment motivation [47-49]. Demographic characteristics included sex (female or male), ethnicity (Han Chinese or other), only child (yes or no). Family background included hometown (urban or rural), father's education, mother's education, having at least one parent as a physician (yes or no), and high-income families (total family income from the previous year>150,000 RMB, approximately US \$21,385, yes or no). We used the National College Entrance Examination scores to measure the respondents' intellectual ability. We assessed enrollment motivation influencing students' choice of clinical medicine following previous work [47]. Intrinsic motivation was measured by students' reported interest in medicine, confidence in achieving success in the field, and strong performance in relevant high school subjects. Extrinsic motivation was evaluated through responses about employment prospects, encouragement (or requirements) from significant others, and anticipated convenience in accessing health care resources for themselves and family members.

Regarding environmental factors, we considered students' affiliation with DFC universities, enrollment in long-track programs (ie, 5+3 or 8 y medical education program instead of 5 y medical education program; yes or no), and the educational phase (preclinical or clinical).

Statistical Analysis

We used the principal component analysis to aggregate the AI knowledge and AI attitude items, followed by Min-Max

normalization to standardize the scores. Based on expert consultation and author consensus, we aggregated the AI behavior index using a weighted arithmetic mean of three key indicators: (1) usage frequency (50% weight), (2) extent of AI integration in professional knowledge learning (30% weight), and (3) AI utilization in writing tasks (20% weight). For robustness check, we also applied equal weights across all three indicators. We then applied the Min-Max normalization method, scaling the data so that the minimum and maximum values correspond to 0 and 100, respectively. Given that students were clustered within schools, we first fitted an unconditional (null) multilevel model to estimate the intraclass correlation coefficient (ICC) and assess the proportion of variance in the standardized AI literacy scores attributable to between-school differences. The ICCs for the three dimensions of AI literacy—knowledge, attitude, and behavior—were 4.0%, 1.0%, and 4.1%, respectively. These results indicate that only a small proportion of the total variance was explained by school-level clustering, suggesting minimal between-school effects. Therefore, we proceeded with multivariate linear regression to examine the associations between individual factors (ie, demographic characteristics, family background, and enrollment motivation) and environmental factors (ie, educational phase, type of education program, and tier of education program) and AI literacy. We conducted subgroup analyses and examined formal interaction terms based on institutional tier. Both classical and school-clustered standard errors were used to calculate 95% CI, and statistical significance was set at a two-sided $P<0.05$. All analyses were performed using Stata (version 18.0; StataCorp LLC) between December 19, 2024, and March 18, 2025.

Results

Participant Characteristics

Of the 80,355 medical students included, most were Han (69,333 students, 86.3%) and female (41,227, 51.3%). A total of 28,885 students (35.9%) reported being the only child; 48,716 (60.6%) were from urban areas; 29,152 (36.3%) reported having at least one parent as a physician. In addition, 13,736 (17.1%) came from high-income families, 36,041 (44.9%) were in the clinical phase of training, and 5823 (7.2%) were enrolled in long-track medical education programs (Table 1).

Table . Characteristics of participants.

Variables	Participants, N=80,355			
	Non-DFC ^a (n=67,417)	DFC (n=12,938)	Total (N=80,355)	P value
Sex, n (%)				
Female	34563 (51.3)	6664 (51.5)	41227 (51.3)	.62
Male	32854 (48.7)	6274 (48.5)	39128 (48.7)	
Ethnicity, n (%)				
Others	9609 (14.3)	1413 (10.9)	11022 (13.7)	<.001
Han Chinese	57808 (85.7)	11525 (89.1)	69333 (86.3)	
Only child, n (%)				
No	44273 (65.7)	7197 (55.6)	51470 (64.1)	<.001
Yes	23144 (34.3)	5741 (44.4)	28885 (35.9)	
Hometown, n (%)				
Rural	27576 (40.9)	4063 (31.4)	31639 (39.4)	<.001
Urban	39841 (59.1)	8875 (68.6)	48716 (60.6)	
Father's education, mean (SD)	11.0 (3.7)	12.0 (4.0)	11.1 (3.8)	<.001
Mother's education, mean (SD)	10.0 (4.1)	11.1 (4.4)	10.2 (4.2)	<.001
Physician parent, n (%)				
No	43146 (64.0)	8057 (62.3)	51203 (63.7)	<.001
Yes	24271 (36.0)	4881 (37.7)	29152 (36.3)	
Family income, n (%)				
Middle and low	57022 (84.6)	9597 (74.2)	66619 (82.9)	<.001
High	10395 (15.4)	3341 (25.8)	13736 (17.1)	
NCEE ^b score, mean (SD)	546.9 (50.6)	593.7 (54.9)	554.5 (54.1)	<.001
Enrollment motivation, mean (SD)				
Intrinsic motivation	69.9 (18.4)	70.1 (18.6)	69.9 (18.4)	.15
Extrinsic motivation	68.4 (18.6)	67.7 (19.1)	68.3 (18.7)	<.001
Long-track programs ^c , n (%)				
No	64425 (95.6)	10107 (78.1)	74532 (92.8)	<.001
Yes	2992 (4.4)	2831 (21.9)	5823 (7.2)	
Educational phase, n (%)				
Preclinical	37439 (55.5)	6875 (53.1)	44314 (55.1)	<.001
Clinical	29978 (44.5)	6063 (46.9)	36041 (44.9)	

^aDFC: Double First-Class universities^bNCEE: National College Entrance Examination^cLong-track programs: the 5+3 medical education program and 8-year medical education program

When stratified by DFC status, students from DFCs came from relatively more advantaged families—with higher household incomes, more parental education, and greater likelihood of urban origin—and also had significantly higher college entrance examination scores (Table 1).

Distribution of the Three Dimensions of AI Literacy

In our analysis of the three dimensions of AI literacy among Chinese medical students (Table 2), the scores declined from knowledge (mean 76.0, SD 26.9) to attitude (mean 71.6, SD 24.4) and then further to behavior (mean 32.5, SD 28.5). When we stratified by institutional tier, students from DFC universities reported lower AI knowledge (mean 72.6, SD 28.6 vs mean

76.5, SD 26.7; $P<.001$) but higher attitude (mean 72.6, SD 24.6 vs mean 71.4, SD 24.4; $P<.001$), and behavior (mean 35.0, SD

Table 1. Summary results for three dimensions of AI literacy.

Variables	Non-DFC ^a (n=67,417)	DFC (n=12,938)	Total (N=80,355)	P value
AI knowledge, mean (SD)	76.5 (26.7)	72.6 (28.6)	76.0 (26.9)	<.001
AI attitude, mean (SD)	71.4 (24.4)	72.6 (24.6)	71.6 (24.4)	<.001
AI behavior, mean (SD) ^b	32.1 (28.5)	35.0 (28.4)	32.5 (28.5)	<.001
AI behavior, mean (SD) ^c	32.3 (29.5)	35.4 (29.5)	32.8 (29.5)	<.001

^aDFC: Double First-Class universities.

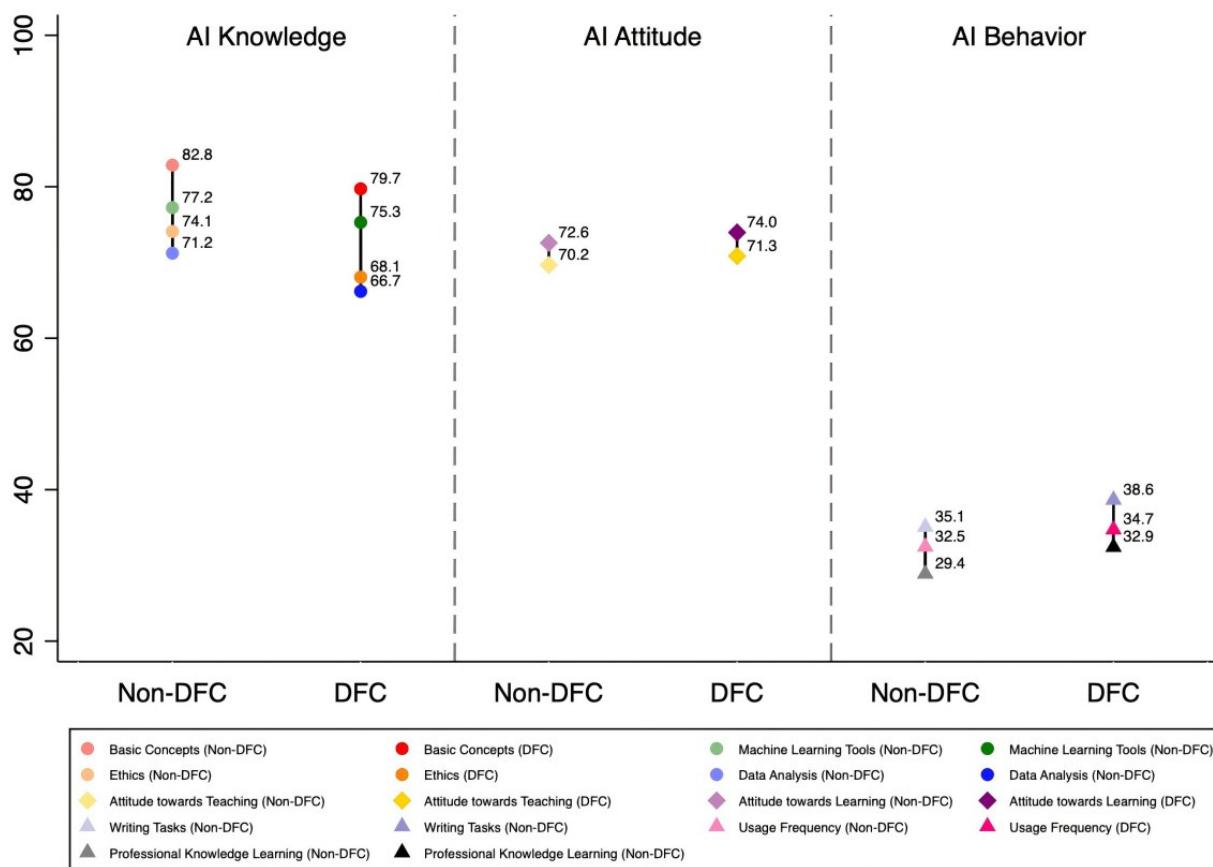
^b expert -assigned weights.

^cequal weights.

Figure 1 further illustrates the item-level patterns within each dimension of AI literacy. The descending trend—from higher scores in knowledge to lower scores in attitude and the lowest in behavior—was consistent across institutional tiers. Notably,

students from DFC universities scored higher on AI behavior subcomponents, especially in professional learning and writing tasks. More detailed results on each institution are provided in Figure S1 in [Multimedia Appendix 1](#).

Figure 1. Components of the three dimensions of AI literacy by institutional tier AI: artificial intelligence; DFC: Double First-Class universities. DFC represents advantaged universities, and non-DFC represents less advantaged universities.



The figure shows the distribution of the subcomponents of the three dimensions of AI literacy by institutional tier. The AI knowledge dimension comprised students' self-reported proficiency in core areas of medical AI, including (1) basic concepts of AI, (2) machine learning tools, (3) multimodal medical data analysis, and (4) ethics issues related to AI. The AI attitude dimension included students' view of using AI tools

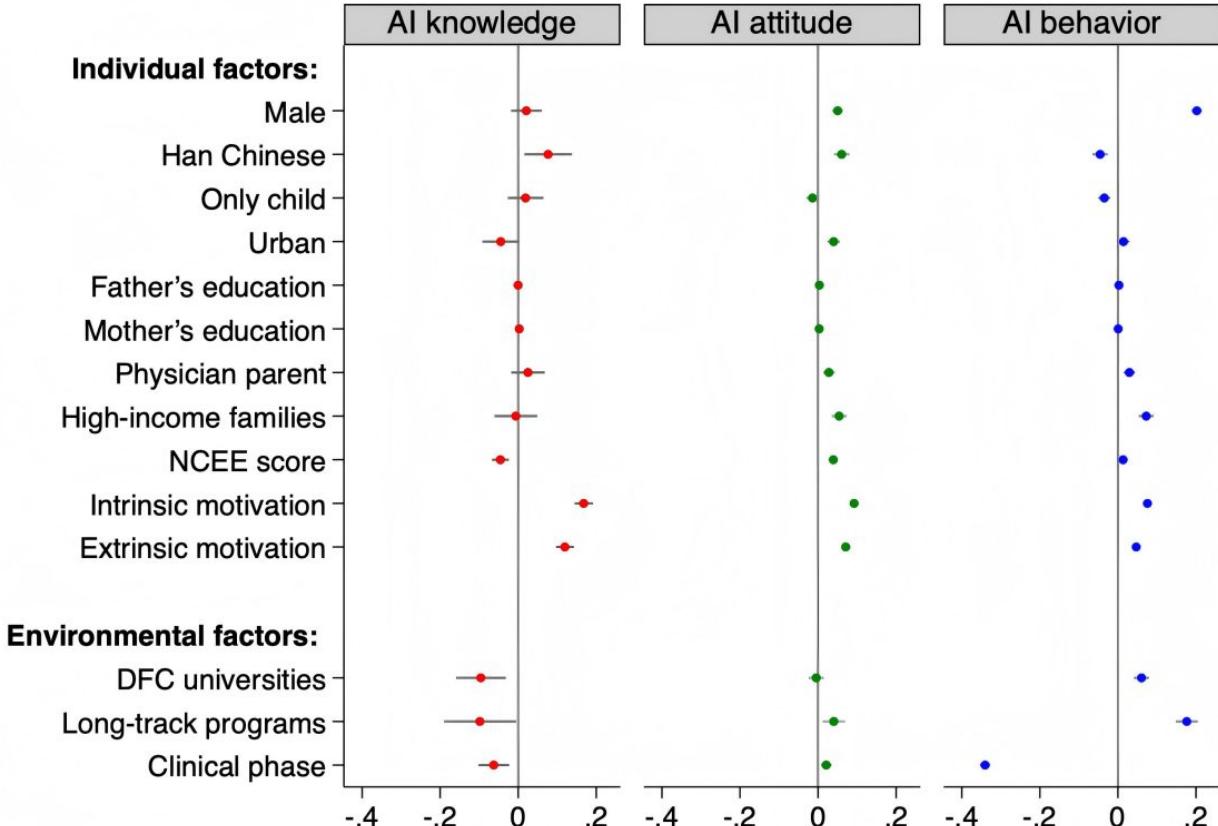
in (1) teaching and (2) learning. The AI behavior dimension was measured by usage frequency and application patterns, including AI use for professional knowledge learning and writing tasks.

Factors Associated With AI Knowledge

Figure 2 shows the results of regression analyzes of the individual and environmental factors associated with AI literacy (more details including robustness check results are available in Tables S1-S4 in [Multimedia Appendix 1](#)). Male students were slightly more likely to report higher AI knowledge ($\beta=0.02$, 95% CI -0.02 to 0.06 ; $P=.30$). Both intrinsic ($\beta=0.17$, 95% CI

0.14 to 0.19 ; $P<.001$) and extrinsic motivation ($\beta=0.12$, 95% CI, 0.10 to 0.14 ; $P<.001$) were positively associated with AI knowledge. Students enrolled in long-track programs reported lower AI knowledge than those in the five-year programs ($\beta=-0.10$, 95% CI -0.19 to -0.01 ; $P=.04$). When stratified by DFC status, these patterns remained largely consistent. However, students from DFC universities reported significantly less AI knowledge ($\beta=-0.10$, 95% CI -0.16 to -0.03 ; $P=.003$).

Figure 2. The individual and environmental factors associated with AI literacy in the full sample. AI: artificial intelligence; DFC: Double First-Class; NCEE: National College entrance examination. DFC represents advantaged universities, and non-DFC represents less advantaged universities. Long-track programs include the 5+3 medical education program and 8-year medical education program.



Factors Associated With AI Attitude

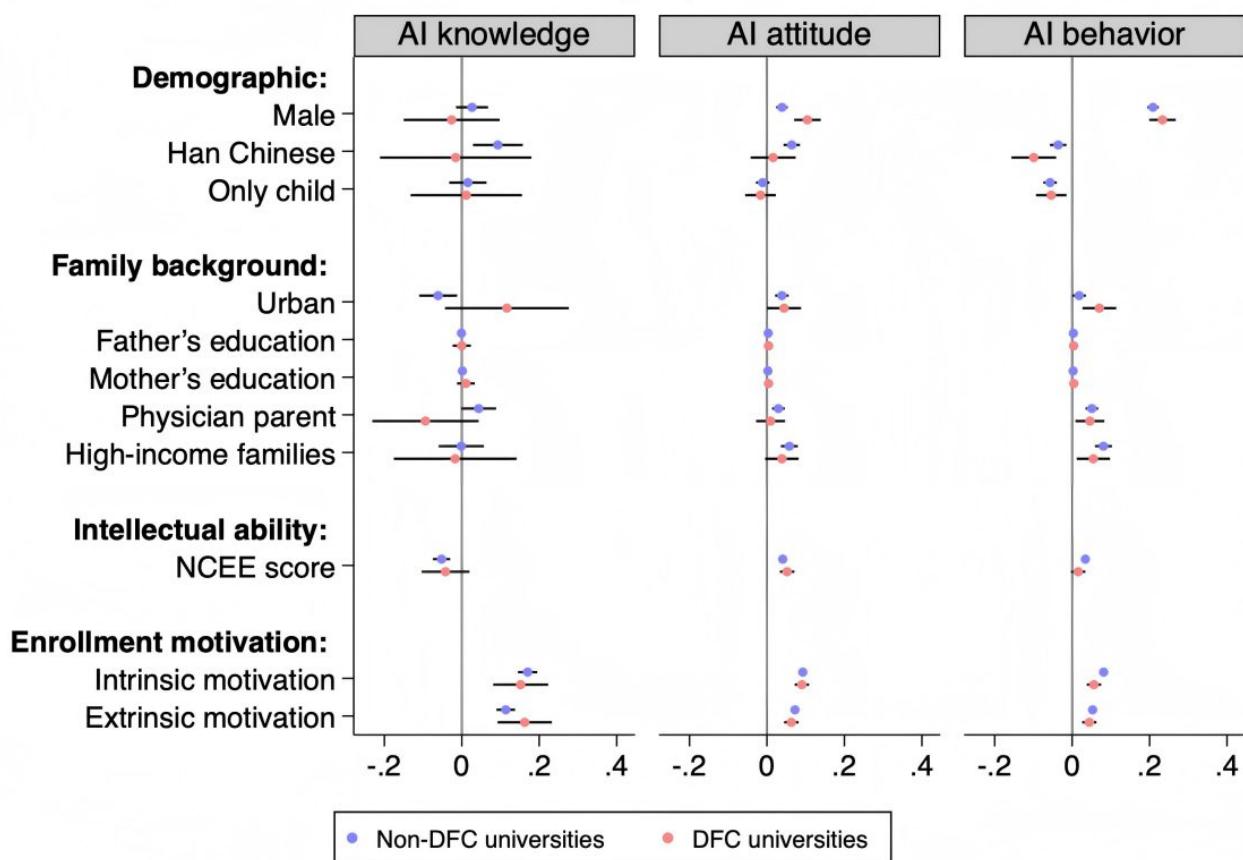
Male students ($\beta=0.05$, 95% CI 0.04 to 0.06 ; $P<.001$) demonstrated significantly more positive attitudes toward AI. Both intrinsic ($\beta=0.09$, 95% CI 0.08 to 0.10 ; $P<.001$) and extrinsic motivation ($\beta=0.07$, 95% CI 0.06 to 0.08 ; $P<.001$) were positively associated with AI attitude. Students from high-income families ($\beta=0.05$, 95% CI 0.03 to 0.07 ; $P<.001$), with at least one physician parent ($\beta=0.03$, 95% CI 0.01 to 0.04 ; $P<.001$), and urban backgrounds ($\beta=0.04$, 95% CI 0.02 to 0.06 ; $P<.001$) showed more positive AI attitude. As for environmental factors, students enrolled in long-track programs ($\beta=0.04$, 95% CI 0.01 to 0.07 ; $P<.006$) and in clinical phase of training ($\beta=0.02$, 95% CI 0.01 to 0.03 ; $P<.003$) showed significantly more positive attitudes toward AI. These patterns remained largely consistent in subgroup analyses by DFC status; however, among the non-DFC students, the associations of family income and parental medical background on AI attitude were significantly stronger.

Factors Associated With AI Behavior

Males ($\beta=0.20$, 95% CI 0.19 to 0.22 ; $P<.001$) demonstrated higher levels of AI behavior. Students from high-income families ($\beta=0.07$, 95% CI 0.05 to 0.09 ; $P<.001$) and those with at least one physician parent ($\beta=0.03$, 95% CI 0.01 to 0.04 ; $P<.001$) showed higher AI behavior scores. Both intrinsic ($\beta=0.08$, 95% CI 0.07 to 0.08 ; $P<.001$) and extrinsic motivation ($\beta=0.05$, 95% CI 0.04 to 0.05 ; $P<.001$) were positively associated with AI behavior. Students enrolled in long-track programs ($\beta=0.18$, 95% CI 0.15 to 0.20 ; $P<.001$) exhibited higher AI behavior. Again, these patterns were largely held in subgroup analysis by DFC status. In the non-DFC group, however, the effect of family income and parental medical background was again stronger.

Figure 3 illustrates the associations between individual factors and AI literacy across different institutional environments.

Figure 3. The individual and environmental factors associated with AI literacy. AI: artificial intelligence; DFC: Double First-Class; NCEE: National College entrance examination.



Discussion

Principal Findings

Despite the recognized importance of preparing medical students to use AI tools effectively, systematic evidence on the subcomponents of AI literacy and the associated factors is limited. To our knowledge, this is the first large-scale, multicenter, national-level study on the multidimensional constructs of AI literacy and their associated individual and environmental factors among medical students. Three key results emerged. First, medical students reported moderate to high AI knowledge, moderate AI attitude, but only low AI behavior scores. Second, both individual and environmental factors were associated with different dimensions of AI literacy. Third, compared to less favorable environmental conditions, the association between worse AI literacy and family income and parental medical background may be weaker.

This study provides several significant contributions. First, the use of a large, multicenter, and nationally representative sample—covering more than 50% (109/202) of the medical schools in China—expands beyond the small, single-center settings from previous studies. Second, by comprehensively examining the cognitive, attitudinal, and behavioral dimensions of AI literacy, this study provides a more holistic understanding of medical students' AI literacy profiles, identifying nuanced gaps and interventional needs within AI literacy, particularly around attitude and behaviors. Third, our comprehensive

examination of the individual differences and environmental factors associated with AI literacy provides an evidence-based foundation for developing tailored interventions. Understanding these associations enables medical educators to design targeted programs that meet the diverse needs of student populations, potentially reducing disparities in AI literacy and ensuring that all future physicians are adequately prepared to engage with an increasingly AI-integrated health care system.

Our findings consistently revealed that self-perceived AI knowledge scores surpassed attitude and behavior. This suggests that while students developed a foundational understanding of AI and generally perceived AI integration in medicine positively, they still lacked practical experience in its application. Although this defies common logic that deeper knowledge should portend more use [50], this trend aligns with previous international studies of medical staff and trainees who held generally positive attitudes toward AI despite low exposure in formal use [36,51].

The potential barriers underlying this knowledge-use gap are as follows. First, structural barriers—such as limited institutional infrastructure, insufficient curricular integration, and time constraints—may constrain students' ability to translate AI knowledge into practice [52]. Such challenges are particularly salient in the Chinese healthcare systems, where physician shortages drive each health care provider to carry significant patient load [53]. These contextual limitations can, in turn, dampen students' autonomous motivation to explore alternative tools [54]. Second, sociocultural norms may also constrain AI use [55]. A recent study showed that nearly half of Chinese

medical staff (1320/2705, 48.8%) expressed concerns about the ethical safety, accuracy, interpretability, and fairness of AI systems [56]. Such ethical reservations from staff providers to medical trainees, leading them to perceive AI use as potentially inappropriate, dishonest, unprofessional, or harmful. This may then limit the students' willingness to use the AI tools despite being knowledgeable of AI tools' capabilities, suggesting that moral and risk considerations may underlie the limited AI use observed among some medical students. Third, our survey was conducted between May and July 2024, when most mature and user-friendly AI tools were developed abroad. Chinese students may have faced barriers related to access [57] and cultural appropriateness [58].

At an individual level, students' demographics, family background, and motivation were all associated with AI literacy. The patterns generally aligned with the literature. Demographically, male medical students scored higher in AI attitude and behavior. This is consistent with previous findings [50,59] and the phenomenon of a "technological gender gap" [60]. Male students tended to perceive AI tools as more useful and easier to use [61-64]. Additionally, individuals from advantaged family background (ie, urban residents, those with at least one physician parent, and students from high-income families) perceived AI more favorably and used it more frequently. These advantaged groups hold greater material, cognitive, social, and cultural capital that can facilitate technological adoption [65]. Although technological innovations eventually reach broader populations, disparities in access and adoption persist during the diffusion process, with advantaged groups generally adopting innovations earlier and using them more effectively, thus creating a digital divide between socioeconomic groups [65,66].

Notably, both intrinsic and extrinsic motivation were positively associated with all dimensions of AI literacy, with intrinsic motivation showing particularly strong effects. According to self-determination theory, intrinsic motivation is associated with better performance and enhanced subjective well-being, while extrinsic motivation can also drive performance, even filtering into intrinsic motivation under certain conditions [67-69]. Intrinsic motivation serves as a natural source of AI learning and achievement, fostering greater autonomy, exploration, and reflection that promote deeper AI engagement. In contrast, extrinsically motivated students may preferentially recognize the instrumental value of AI but limit autonomous exploration of AI tools.

Our study highlights the significant role of the environment in shaping AI literacy. Specifically, DFC students demonstrated higher AI behavior scores. The DFC Project represents China's willingness to invest financial and human resources in these institutions to develop world-class universities and disciplines [47,70]. The stronger student-faculty and peer interactions and higher-quality student efforts may promote deeper engagement with AI [71]. The seemingly paradoxical finding that DFC students reported lower AI knowledge but higher AI behavior may stem from the Dunning-Kruger effect [72,73]. More competent individuals tend to recognize the complexity of a domain and their own knowledge limitations, leading to more cautious self-assessments, while those less competent may

overestimate their understanding. DFC students, exposed to higher levels of academic research, may be more aware of the challenges inherent in AI, whereas non-DFC students, with relatively limited exposure, may exhibit overconfidence in their basic knowledge. Similar cognitive bias patterns have been observed among health professions students; for example, low-performing medical students have been shown to substantially overestimate their academic performance relative to peers [74,75].

Importantly, in these DFC supportive environments, the students' family income and parental medical background showed less pronounced associations with AI attitude and behavior, suggesting that the supportive learning environment may have mitigated the detrimental relationship between students' background and AI literacy [48].

In addition, the students' program also appears to matter. Students in the long-track program demonstrated significantly higher AI attitude and behavior scores. They typically represent China's elite medical students, attending research-oriented universities that provide enhanced research training opportunities and funding [76], potentially including familiarization with AI technologies that led to more positive perceptions and use of AI tools. The students' phase of training also appeared to be associated with their AI literacy. Students undergoing clinical training reported higher attitude scores but lower knowledge and behavior scores. Perhaps the complexity of clinical practice and the pressure of professional development precipitated cognitive overload and a preferential reliance on traditional tools [77]. The technical limitations and evolving normative and legal thresholds around AI technologies may further exacerbate technology avoidance [78-82].

Tying together the findings, we provide five sets of recommendations to help medical schools maneuver through the current lack of structured AI education and training [83]. First, carefully integrating AI-related content into the curriculum to generate practical AI exposure may help bridge the knowledge-use gap. This should be mindful of the potentially detrimental association with excessive cognitive offloading [84] while ensuring a minimal knowledge and skills threshold [85]. As regulations and ethical frameworks surrounding AI use in health care continue to evolve to meet the technical, ethical, and legal challenges [83,86,87], developing interdisciplinary training on practical risks, benefits, and guidelines on AI usage may help establish norms around AI usage. Second, developing and implementing scalable AI education frameworks that enable targeted training based on learners' backgrounds can potentially reduce inequities in AI literacy driven by individual and environmental differences while maximizing educational impact.

The next three recommendations focus specifically on digital divide, and we draw on Han and Kumwenda's framework for mitigating digital divide. The framework considers first-order barriers relate to infrastructure and access; second-order barriers include motivation and engagement; and third-order barriers reflect skill and training disparities [88].

First, medical educators and policy makers should prioritize investments in digital infrastructure to reduce disparities in technological access across institutional and regional contexts

[89]. Targeted funding and digital resource allocation can ensure that students in under-resourced settings are not excluded from AI-driven learning environments.

Second, improving motivation and engagement may incorporate experiential learning opportunities [87,90,91]. Providing all students—particularly those from disadvantaged backgrounds—with the chance to work directly with AI tools and integrate them into their coursework can foster greater motivation, engagement, and confidence in using AI technologies.

Third, structured and scalable AI education frameworks that offer tiered and targeted training that adapts to learners' backgrounds. Such approaches would focus on ensuring that disadvantaged students can develop foundational AI literacy before advancing to higher-order skills. Moreover, the development of multilingual and culturally adaptive AI models [92] can further ensure inclusivity and contextual relevance in AI-enhanced medical education.

Limitations

First, the reliance on self-reported data potentially exposes the study to social desirability bias. It is necessary to develop objective instruments to measure medical students' AI literacy. Such a tool can help evaluate AI proficiency based on observable AI use. Performance- and scenario-based assessments might be potential avenues to operationalize such evaluation. Second, given the rapid development of AI technologies, the

AI literacy of medical students is likely to evolve rapidly. To capture these dynamic shifts of AI literacy, we are actively exploring another round of survey. Third, this study employed a cross-sectional design, which precludes causal inference. Although most individual and environmental factors were determined before the measurement of AI literacy, reverse causality cannot be fully excluded, and unmeasured confounders may have influenced both the predictors and outcomes, potentially biasing the observed associations. Fourth, even though the unique premedical training structure limits the generalizability of the findings, the methodological framework and multidimensional conceptualization of AI literacy (knowledge, attitude, and behavior) may still provide a useful reference for future comparative studies or adaptations in other educational systems.

Conclusions

This cross-sectional study revealed that medical students exhibited the highest performance in AI knowledge, followed by attitude, and then behavior. Both individual characteristics and environmental factors were significantly associated with AI literacy, and environmental factors moderated individual variations in AI literacy. Integrating practical AI-related training into medical curricula through interdisciplinary collaboration, coupled with targeted interventions for students according to their backgrounds, may help prepare future physicians to effectively engage with AI technologies in medical practice.

Acknowledgments

We are grateful to the National Center for Health Professions Education Development for their support of this study. We also extend our thanks to all the medical schools and medical students who participated in the CMSS survey. We also appreciate the insightful comments from Dr. Weimin Wang, Ms Ana Xie, and Dr. Chen Yu at the National Center for Health Professions Education Development, Peking University, as well as from Dr. Dragan Gasevic and Dr. Danijela Gasevic at Monash University.

CL was affiliated with the Institute of Medical Education at Peking University at the time of the study and is currently affiliated with the Education Management Information Center, Ministry of Education of the People's Republic of China.

Funding

This study was supported by the 2025 Undergraduate Teaching Reform Project of Peking University (JG2025170). This study was supported by the 2025 Undergraduate Teaching Reform Project of Peking University (JG2025170). The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclaimer

The views expressed herein are solely those of the authors and do not necessarily represent the views of their employers, funders, or affiliated institutions.

Data Availability

The parameters that we used in our model (text, tables, figures, models, and appendices) are available on reasonable request from the corresponding author under certain conditions (with the consent of all participating centers and with a signed data access agreement).

Authors' Contributions

HW had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: CL, HW.

Acquisition, analysis, or interpretation of data: CL, SHT, HW.

Drafting of the manuscript: CL, HW.

Critical review of the manuscript for important intellectual content: CL, SHT, HW.

Statistical analysis: CL.

Obtained funding: Wu.

Administrative, technical, or material support: HW.

Supervision: HW.

Conflicts of Interest

None declared.

Multimedia Appendix 1 [[DOC File, 1036 KB - mededu_v12i1e80604_app1.doc](#)]

References

1. Elemento O, Khozin S, Sternberg CN. The use of artificial intelligence for cancer therapeutic decision-making. *NEJM AI* 2025 May;2(5). [doi: [10.1056/aira2401164](https://doi.org/10.1056/aira2401164)] [Medline: [41112204](#)]
2. Sarkar U, Bates DW. Using artificial intelligence to improve primary care for patients and clinicians. *JAMA Intern Med* 2024 Apr 1;184(4):343-344. [doi: [10.1001/jamainternmed.2023.7965](https://doi.org/10.1001/jamainternmed.2023.7965)] [Medline: [38345801](#)]
3. 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](https://doi.org/10.1056/NEJMra2302038)] [Medline: [36988595](#)]
4. Unlu O, Varughese M, Shin J, et al. Manual vs AI-assisted prescreening for trial eligibility using large language models-a randomized clinical trial. *JAMA* 2025 Mar 25;333(12):1084-1087. [doi: [10.1001/jama.2024.28047](https://doi.org/10.1001/jama.2024.28047)] [Medline: [39960745](#)]
5. Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. *Acad Med* 2024;99(1):22-27. [doi: [10.1097/ACM.0000000000005439](https://doi.org/10.1097/ACM.0000000000005439)]
6. Preiksaitis C, Rose C. Opportunities, challenges, and future directions of generative artificial intelligence in medical education: scoping review. *JMIR Med Educ* 2023 Oct 20;9(1):e48785. [doi: [10.2196/48785](https://doi.org/10.2196/48785)] [Medline: [37862079](#)]
7. 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](https://doi.org/10.1080/0142159X.2024.2314198)] [Medline: [38423127](#)]
8. Burk-Rafel J, Reinstein I, Feng J, et al. Development and validation of a machine learning-based decision support tool for residency applicant screening and review. *Acad Med* 2021;96(11S):S54-S61. [doi: [10.1097/ACM.0000000000004317](https://doi.org/10.1097/ACM.0000000000004317)]
9. Baron T, Grossman RI, Abramson SB, et al. Signatures of medical student applicants and academic success. *PLoS One* 2020;15(1):e0227108. [doi: [10.1371/journal.pone.0227108](https://doi.org/10.1371/journal.pone.0227108)] [Medline: [31940377](#)]
10. Kind T, Olvet DM, Farina G, et al. Reading and study habits of medical students on clerkships and performance outcomes: a multi-institutional study. *Med Sci Educ* 2021 Dec;31(6):1957-1966. [doi: [10.1007/s40670-021-01409-5](https://doi.org/10.1007/s40670-021-01409-5)] [Medline: [34956707](#)]
11. Cutrer WB, Spickard WA 3rd, Triola MM, et al. Exploiting the power of information in medical education. *Med Teach* 2021 Jul;43(sup2):S17-S24. [doi: [10.1080/0142159X.2021.1925234](https://doi.org/10.1080/0142159X.2021.1925234)] [Medline: [34291714](#)]
12. Roberts LW. Addressing the novel implications of generative ai for academic publishing, education, and research. *Acad Med* 2024;99(5):471-473. [doi: [10.1097/ACM.0000000000005667](https://doi.org/10.1097/ACM.0000000000005667)]
13. Stokel-Walker C. ChatGPT listed as author on research papers: many scientists disapprove. *Nature New Biol* 2023 Jan;613(7945):620-621. [doi: [10.1038/d41586-023-00107-z](https://doi.org/10.1038/d41586-023-00107-z)] [Medline: [36653617](#)]
14. 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. [doi: [10.1001/jama.2023.1344](https://doi.org/10.1001/jama.2023.1344)]
15. Tolsgaard MG, Boscardin CK, Park YS, Cuddy MM, Sebok-Syer SS. The role of data science and machine learning in Health Professions Education: practical applications, theoretical contributions, and epistemic beliefs. *Adv Health Sci Educ Theory Pract* 2020 Dec;25(5):1057-1086. [doi: [10.1007/s10459-020-10009-8](https://doi.org/10.1007/s10459-020-10009-8)] [Medline: [33141345](#)]
16. Gin BC, Ten Cate O, O’Sullivan PS, Hauer KE, Boscardin C. Exploring how feedback reflects entrustment decisions using artificial intelligence. *Med Educ* 2022 Mar;56(3):303-311. [doi: [10.1111/medu.14696](https://doi.org/10.1111/medu.14696)] [Medline: [34773415](#)]
17. Schaye V, Guzman B, Burk-Rafel J, et al. Development and validation of a machine learning model for automated assessment of resident clinical reasoning documentation. *J Gen Intern Med* 2022 Jul;37(9):2230-2238. [doi: [10.1007/s11606-022-07526-0](https://doi.org/10.1007/s11606-022-07526-0)] [Medline: [35710676](#)]
18. Triola MM, Rodman A. Integrating generative artificial intelligence into medical education: curriculum, policy, and governance strategies. *Acad Med* 2025;100(4):413-418. [doi: [10.1097/ACM.0000000000005963](https://doi.org/10.1097/ACM.0000000000005963)]
19. 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;98(3):348-356. [doi: [10.1097/ACM.0000000000004963](https://doi.org/10.1097/ACM.0000000000004963)]
20. Ong QC, Ang CS, Lai NM, Neves AL, Car J. Dearth of digital health education: the need for an accelerated medical curriculum reform in Malaysia. *Lancet Reg Health West Pac* 2025 Feb;55:101476. [doi: [10.1016/j.lanwpc.2025.101476](https://doi.org/10.1016/j.lanwpc.2025.101476)] [Medline: [39902151](#)]

21. Succi MD, Chang BS, Rao AS. Building the AI-enabled medical school of the future. *JAMA* 2025 May 20;333(19):1665-1666. [doi: [10.1001/jama.2025.2789](https://doi.org/10.1001/jama.2025.2789)] [Medline: [40163081](https://pubmed.ncbi.nlm.nih.gov/40163081/)]
22. Lee J, Wu AS, Li D, Kulasegaram K. Artificial intelligence in undergraduate medical education: a scoping review. *Acad Med* 2021;96(11S):S62-S70. [doi: [10.1097/ACM.00000000000004291](https://doi.org/10.1097/ACM.00000000000004291)]
23. Long D, Magerko B. What is AI literacy? Competencies and design considerations. *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* 2020 Apr 21:1-16. [doi: [10.1145/3313831.3376727](https://doi.org/10.1145/3313831.3376727)]
24. Ng DTK, Wu W, Leung JKL, Chiu TKF, Chu SKW. Design and validation of the ai literacy questionnaire: the affective, behavioural, cognitive and ethical approach. *Brit J Educational Tech* 2024 May;55(3):1082-1104. [doi: [10.1111/bjet.13411](https://doi.org/10.1111/bjet.13411)]
25. Ng DTK, Leung JKL, Chu SKW, Qiao MS. Conceptualizing AI literacy: an exploratory review. *Computers and Education: Artificial Intelligence* 2021;2:100041. [doi: [10.1016/j.caai.2021.100041](https://doi.org/10.1016/j.caai.2021.100041)]
26. Aktoprak A, Hursen C. A bibliometric and content analysis of critical thinking in primary education. *Think Skills Creat* 2022 Jun;44:101029. [doi: [10.1016/j.tsc.2022.101029](https://doi.org/10.1016/j.tsc.2022.101029)]
27. Gruber T, Bazhydai M, Sievers C, Clément F, Dukes D. The ABC of social learning: affect, behavior, and cognition. *Psychol Rev* 2022 Nov;129(6):1296-1318. [doi: [10.1037/rev0000311](https://doi.org/10.1037/rev0000311)] [Medline: [34292022](https://pubmed.ncbi.nlm.nih.gov/34292022/)]
28. Boyle MJ, Williams B, Brown T, et al. Attitudes of undergraduate health science students towards patients with intellectual disability, substance abuse, and acute mental illness: a cross-sectional study. *BMC Med Educ* 2010 Oct 21;10(1):71. [doi: [10.1186/1472-6920-10-71](https://doi.org/10.1186/1472-6920-10-71)] [Medline: [20964840](https://pubmed.ncbi.nlm.nih.gov/20964840/)]
29. Van Harreveld F. The ABC of ambivalence: affective, behavioral, and cognitive consequences of attitudinal conflict. In: *Adv Exp Soc Psychol* 2015, Vol. 52:285-324. [doi: [10.1016/bs.aesp.2015.01.002](https://doi.org/10.1016/bs.aesp.2015.01.002)]
30. Patino GA, Amiel JM, Brown M, Lypson ML, Chan TM. The promise and perils of artificial intelligence in health professions education practice and scholarship. *Acad Med* 2024;99(5):477-481. [doi: [10.1097/ACM.00000000000005636](https://doi.org/10.1097/ACM.00000000000005636)]
31. Barreiro-Ares A, Morales-Santiago A, Sendra-Portero F, Souto-Barayri M. Impact of the rise of artificial intelligence in radiology: What do students think? *Int J Environ Res Public Health* 2023 Jan 16;20(2):1589. [doi: [10.3390/ijerph20021589](https://doi.org/10.3390/ijerph20021589)] [Medline: [36674348](https://pubmed.ncbi.nlm.nih.gov/36674348/)]
32. Stewart J, Lu J, Gahungu N, et al. Western Australian medical students' attitudes towards artificial intelligence in healthcare. *PLoS One* 2023;18(8):e0290642. [doi: [10.1371/journal.pone.0290642](https://doi.org/10.1371/journal.pone.0290642)] [Medline: [37651380](https://pubmed.ncbi.nlm.nih.gov/37651380/)]
33. Wood EA, Ange BL, Miller DD. Are we ready to integrate artificial intelligence literacy into medical school curriculum: students and faculty survey. *J Med Educ Curric Dev* 2021;8:23821205211024078. [doi: [10.1177/23821205211024078](https://doi.org/10.1177/23821205211024078)] [Medline: [34250242](https://pubmed.ncbi.nlm.nih.gov/34250242/)]
34. Ejaz H, McGrath H, Wong BL, Guise A, Vercauteren T, Shapey J. Artificial intelligence and medical education: a global mixed-methods study of medical students' perspectives. *Digit Health* 2022;8:20552076221089099. [doi: [10.1177/20552076221089099](https://doi.org/10.1177/20552076221089099)] [Medline: [35521511](https://pubmed.ncbi.nlm.nih.gov/35521511/)]
35. 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](https://doi.org/10.3389/frai.2022.1015418)] [Medline: [36406470](https://pubmed.ncbi.nlm.nih.gov/36406470/)]
36. 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](https://doi.org/10.1186/s13244-019-0830-7)] [Medline: [32025951](https://pubmed.ncbi.nlm.nih.gov/32025951/)]
37. Mehta N, Harish V, Bilimoria K, et al. Knowledge of and attitudes on artificial intelligence in healthcare: a provincial survey study of medical students. *Medical Education*. Preprint posted online on 2021. [doi: [10.1101/2021.01.14.21249830](https://doi.org/10.1101/2021.01.14.21249830)]
38. 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](https://doi.org/10.1080/10872981.2023.2182659)] [Medline: [36855245](https://pubmed.ncbi.nlm.nih.gov/36855245/)]
39. 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](https://doi.org/10.2196/33390)] [Medline: [35099397](https://pubmed.ncbi.nlm.nih.gov/35099397/)]
40. Wang B, Rau PLP, Yuan T. Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology* 2023 Jul 4;42(9):1324-1337. [doi: [10.1080/0144929X.2022.2072768](https://doi.org/10.1080/0144929X.2022.2072768)]
41. 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](https://doi.org/10.1186/s12909-024-05400-7)] [Medline: [38600457](https://pubmed.ncbi.nlm.nih.gov/38600457/)]
42. Swed S, Alibrahim H, Elkalagi NKH, et al. Knowledge, attitude, and practice of artificial intelligence among doctors and medical students in Syria: A cross-sectional online survey. *Front Artif Intell* 2022;5:1011524. [doi: [10.3389/frai.2022.1011524](https://doi.org/10.3389/frai.2022.1011524)] [Medline: [36248622](https://pubmed.ncbi.nlm.nih.gov/36248622/)]
43. Holstein K, Doroudi S. Equity and artificial intelligence in education: will "AIED" amplify or alleviate inequities in education? *arXiv*. Preprint posted online on Apr 27, 2021. [doi: [10.48550/arXiv.2104.12920](https://doi.org/10.48550/arXiv.2104.12920)]
44. van Deursen AJ, van Dijk JA. The digital divide shifts to differences in usage. *New Media & Society* 2014 May;16(3):507-526. [doi: [10.1177/1461444813487959](https://doi.org/10.1177/1461444813487959)]

45. van Dijk J. Digital divide: impact of access. In: Rössler P, Hoffner CA, Zoonen L, editors. The International Encyclopedia of Media Effects, 1st edition 2017:1-11. [doi: [10.1002/9781118783764.wbime0043](https://doi.org/10.1002/9781118783764.wbime0043)]

46. Ma X, Shen Z, Xiao R, Wu H. Perceived mistreatment and professional identity of medical students in China. *JAMA Netw Open* 2024 Nov 4;7(11):e2444245. [doi: [10.1001/jamanetworkopen.2024.44245](https://doi.org/10.1001/jamanetworkopen.2024.44245)]

47. 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. [doi: [10.1080/10872981.2020.1742964](https://doi.org/10.1080/10872981.2020.1742964)] [Medline: [32180537](https://pubmed.ncbi.nlm.nih.gov/32180537/)]

48. Loh RSM, Kraaykamp G, van Hek M. Plugging in at school: do schools nurture digital skills and narrow digital skills inequality? *Computers & Education* 2025 Mar;226:105195. [doi: [10.1016/j.compedu.2024.105195](https://doi.org/10.1016/j.compedu.2024.105195)]

49. Hargittai E. Digital na(t)ives? Variation in internet skills and uses among members of the "Net Generation". *Sociol Inq* 2010 Feb;80(1):92-113. [doi: [10.1111/j.1475-682X.2009.00317.x](https://doi.org/10.1111/j.1475-682X.2009.00317.x)]

50. 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](https://doi.org/10.1007/s00330-018-5601-1)]

51. Sarwar S, Dent A, Faust K, et al. Physician perspectives on integration of artificial intelligence into diagnostic pathology. *NPJ Digit Med* 2019;2:28. [doi: [10.1038/s41746-019-0106-0](https://doi.org/10.1038/s41746-019-0106-0)] [Medline: [31304375](https://pubmed.ncbi.nlm.nih.gov/31304375/)]

52. Kennedy T, Regehr G, Rosenfield J, Roberts SW, Lingard L. Exploring the gap between knowledge and behavior: a qualitative study of clinician action following an educational intervention. *Acad Med* 2004 May;79(5):386-393. [doi: [10.1097/00001888-200405000-00006](https://doi.org/10.1097/00001888-200405000-00006)] [Medline: [15107277](https://pubmed.ncbi.nlm.nih.gov/15107277/)]

53. Wang X, Sanders HM, Liu Y, et al. ChatGPT: promise and challenges for deployment in low- and middle-income countries. *The Lancet Regional Health - Western Pacific* 2023 Dec;41:100905. [doi: [10.1016/j.lanwpc.2023.100905](https://doi.org/10.1016/j.lanwpc.2023.100905)]

54. Ryan RM, Deci EL. Self-determination theory. In: Encyclopedia of Quality of Life and Well-Being Research: Springer, Vol. 2024:6229-6235. [doi: [10.1007/978-3-031-17299-1_2630](https://doi.org/10.1007/978-3-031-17299-1_2630)]

55. Kahlke RM, McConnell MM, Wisener KM, Eva KW. The disconnect between knowing and doing in health professions education and practice. *Adv in Health Sci Educ* 2020 Mar;25(1):227-240. [doi: [10.1007/s10459-019-09886-5](https://doi.org/10.1007/s10459-019-09886-5)]

56. Dai Q, Li M, Yang M, et al. Attitudes, perceptions, and factors influencing the adoption of AI in health care among medical staff: nationwide cross-sectional survey study. *J Med Internet Res* 2025 Aug 8;27:e75343. [doi: [10.2196/75343](https://doi.org/10.2196/75343)] [Medline: [40779308](https://pubmed.ncbi.nlm.nih.gov/40779308/)]

57. Meaningful connectivity: a new target to raise the bar for internet access. Alliance for Affordable Internet. URL: https://a4ai.org/wp-content/uploads/2021/02/meaningful-connectivity_Public-.pdf [accessed 2025-12-23]

58. Cao Y, Zhou L, Lee S, Cabello L, Chen M, Hershcovich D. Assessing cross-cultural alignment between chatgpt and human societies: an empirical study. . 2023. [doi: [10.18653/v1/2023.c3nlp-1.7](https://doi.org/10.18653/v1/2023.c3nlp-1.7)]

59. Akca Sumengen A, Ozcevik Subasi D, Cakir GN. Nursing students' attitudes and literacy toward artificial intelligence: a cross-sectional study. *Teaching and Learning in Nursing* 2025 Jan;20(1):e250-e257. [doi: [10.1016/j.teln.2024.10.022](https://doi.org/10.1016/j.teln.2024.10.022)]

60. Canada K, Brusca F. The technological gender gap: evidence and recommendations for educators and computer-based instruction designers. *ETR&D* 1991 Jun;39(2):43-51. [doi: [10.1007/BF02298153](https://doi.org/10.1007/BF02298153)]

61. Cai Z, Fan X, Du J. Gender and attitudes toward technology use: a meta-analysis. *Comput Educ* 2017 Feb;105:1-13. [doi: [10.1016/j.compedu.2016.11.003](https://doi.org/10.1016/j.compedu.2016.11.003)]

62. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q* 1989 Sep 1;13(3):319-340. [doi: [10.2307/249008](https://doi.org/10.2307/249008)]

63. Ong CS, Lai JY. Gender differences in perceptions and relationships among dominants of e-learning acceptance. *Comput Human Behav* 2006 Sep;22(5):816-829. [doi: [10.1016/j.chb.2004.03.006](https://doi.org/10.1016/j.chb.2004.03.006)]

64. Ackerman PL, Wolman SD. Determinants and validity of self-estimates of abilities and self-concept measures. *J Exp Psychol Appl* 2007 Jun;13(2):57-78. [doi: [10.1037/1076-898X.13.2.57](https://doi.org/10.1037/1076-898X.13.2.57)] [Medline: [17535132](https://pubmed.ncbi.nlm.nih.gov/17535132/)]

65. van Dijk J. : Polity Press; 2020.

66. Norris P. Digital divide: civic engagement, information poverty, and the internet worldwide. *Canadian Journal of Communication* 2003 Jan 1;28(1). [doi: [10.22230/cjc.2003v28n1a1352](https://doi.org/10.22230/cjc.2003v28n1a1352)]

67. 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](https://doi.org/10.1006/ceps.1999.1020)] [Medline: [10620381](https://pubmed.ncbi.nlm.nih.gov/10620381/)]

68. Deci EL, Ryan RM. Intrinsic Motivation and Self-Determination in Human Behavior: Springer US; 1985. [doi: [10.1007/978-1-4899-2271-7](https://doi.org/10.1007/978-1-4899-2271-7)]

69. Ryan RM, Stiller J. The social contexts of internalization: parent and teacher influences on autonomy, motivation and learning. In: *Advances in Motivation and Achievement* 1991, Vol. 7:115-149.

70. Qi W. Comment: Programmed to fulfill global ambitions. *Nature New Biol* 2017 May 25;545(7655):S53-S53. [doi: [10.1038/545S53a](https://doi.org/10.1038/545S53a)]

71. Pascarella ET, Terenzini PT. How College Affects Students: A Third Decade of Research: ERIC; 2005, Vol. 2. URL: <https://eric.ed.gov/?id=ED498537> [accessed 2025-06-26]

72. 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](https://doi.org/10.1037//0022-3514.77.6.1121)] [Medline: [10626367](https://pubmed.ncbi.nlm.nih.gov/10626367/)]

73. Dunning D. The Dunning–Kruger effect: on being ignorant of one’s own ignorance. In: *Adv Exp Soc Psychol*: Elsevier; 2011, Vol. 44:247-296. [doi: [10.1016/B978-0-12-385522-0.00005-6](https://doi.org/10.1016/B978-0-12-385522-0.00005-6)]

74. Edwards RK, Kellner KR, Sistrom CL, Magyari EJ. Medical student self-assessment of performance on an obstetrics and gynecology clerkship. *Am J Obstet Gynecol* 2003 Apr;188(4):1078-1082. [doi: [10.1067/mob.2003.249](https://doi.org/10.1067/mob.2003.249)] [Medline: [12712114](https://pubmed.ncbi.nlm.nih.gov/12712114/)]

75. Austin Z, Gregory PAM, Galli M. “I just don’t know what I’m supposed to know”: Evaluating self-assessment skills of international pharmacy graduates in Canada. *Research in Social and Administrative Pharmacy* 2008 Jun;4(2):115-124. [doi: [10.1016/j.sapharm.2007.03.002](https://doi.org/10.1016/j.sapharm.2007.03.002)]

76. Zhang G, Wu H, Xie A, Cheng H. The association between medical student research engagement with learning outcomes. *Med Educ Online* 2022 Dec;27(1):2100039. [doi: [10.1080/10872981.2022.2100039](https://doi.org/10.1080/10872981.2022.2100039)] [Medline: [35818325](https://pubmed.ncbi.nlm.nih.gov/35818325/)]

77. Sweller J. Cognitive load during problem solving: effects on learning. *Cogn Sci* 1988 Jun;12(2):257-285. [doi: [10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7)]

78. Bernstein MH, Sheppard B, Bruno MA, Lay PS, Baird GL. Randomized study of the impact of AI on perceived legal liability for radiologists. *NEJM AI* 2025 May 22;2(6):A1oa2400785. [doi: [10.1056/A1oa2400785](https://doi.org/10.1056/A1oa2400785)]

79. Tsuei SHT. How are Canadians regulating artificial intelligence for healthcare? A brief analysis of the current legal directions, challenges and deficiencies. *Healthc Pap* 2025 Apr;22(4):44-51. [doi: [10.12927/hcpap.2025.27571](https://doi.org/10.12927/hcpap.2025.27571)] [Medline: [40391478](https://pubmed.ncbi.nlm.nih.gov/40391478/)]

80. Blumenthal D, Patel B. The regulation of clinical artificial intelligence. *NEJM AI* 2024 Jul 25;1(8):A1pc2400545. [doi: [10.1056/A1pc2400545](https://doi.org/10.1056/A1pc2400545)]

81. 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](https://doi.org/10.1038/s41591-018-0300-7)] [Medline: [30617339](https://pubmed.ncbi.nlm.nih.gov/30617339/)]

82. Reddy S. Explainability and artificial intelligence in medicine. *Lancet Digit Health* 2022 Apr;4(4):e214-e215. [doi: [10.1016/S2589-7500\(22\)00029-2](https://doi.org/10.1016/S2589-7500(22)00029-2)] [Medline: [35337639](https://pubmed.ncbi.nlm.nih.gov/35337639/)]

83. Car J, Ong QC, Erlikh Fox T, et al. The digital health competencies in medical education framework: an international consensus statement based on a Delphi study. *JAMA Netw Open* 2025 Jan 2;8(1):e2453131. [doi: [10.1001/jamanetworkopen.2024.53131](https://doi.org/10.1001/jamanetworkopen.2024.53131)] [Medline: [39888625](https://pubmed.ncbi.nlm.nih.gov/39888625/)]

84. Kosmyna N, Hauptmann E, Yuan YT, et al. Your brain on ChatGPT: accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv*. Preprint posted online on Jun 10, 2025. [doi: [10.48550/arXiv.2506.08872](https://doi.org/10.48550/arXiv.2506.08872)]

85. Gin BC, LaForge K, Burk-Rafel J, Boscardin CK. Macy Foundation Innovation Report Part II: from hype to reality: innovators’ visions for navigating AI integration challenges in medical education. *Acad Med* 2025;100(9S):S22-S29. [doi: [10.1097/ACM.0000000000006117](https://doi.org/10.1097/ACM.0000000000006117)]

86. Zhai X, Chu X, Chai CS, et al. A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity* 2021 Jan;2021(1):8812542. [doi: [10.1155/2021/8812542](https://doi.org/10.1155/2021/8812542)]

87. 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](https://doi.org/10.1038/s41746-020-0294-7)] [Medline: [32577533](https://pubmed.ncbi.nlm.nih.gov/32577533/)]

88. Han SP, Kumwenda B. Bridging the digital divide: promoting equal access to online learning for health professions in an unequal world. *Med Educ* 2025 Jan;59(1):56-64. [doi: [10.1111/medu.15455](https://doi.org/10.1111/medu.15455)] [Medline: [38860836](https://pubmed.ncbi.nlm.nih.gov/38860836/)]

89. 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 Mar 7;11:e64768. [doi: [10.2196/64768](https://doi.org/10.2196/64768)] [Medline: [40053774](https://pubmed.ncbi.nlm.nih.gov/40053774/)]

90. 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](https://doi.org/10.2196/16048)] [Medline: [31793895](https://pubmed.ncbi.nlm.nih.gov/31793895/)]

91. Kolachalama VB, Garg PS. Machine learning and medical education. *NPJ Digit Med* 2018;1(1):54. [doi: [10.1038/s41746-018-0061-1](https://doi.org/10.1038/s41746-018-0061-1)] [Medline: [31304333](https://pubmed.ncbi.nlm.nih.gov/31304333/)]

92. Fitas R. Inclusive education with AI: supporting special needs and tackling language barriers. *SSRN*. Preprint posted online on Apr 19, 2025. [doi: [10.2139/ssrn.5366398](https://doi.org/10.2139/ssrn.5366398)]

Abbreviations

AI: artificial intelligence
CMSS: China Medical Student Survey
DFC: Double First-Class
ICC: intraclass correlation coefficient

Edited by J Moen; submitted 14.Jul.2025; peer-reviewed by MA Amin, X Liang; revised version received 10.Nov.2025; accepted 05.Dec.2025; published 06.Jan.2026.

Please cite as:

Li C, Tsuei SHT, Wu H

AI Literacy Among Chinese Medical Students: Cross-Sectional Examination of Individual and Environmental Factors

JMIR Med Educ 2026;12:e80604

URL: <https://mededu.jmir.org/2026/1/e80604>

doi:10.2196/80604

© Chunqing Li, Sian Hsiang-Te Tsuei, Hongbin Wu. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 6.Jan.2026. 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.

GPT-4o and OpenAI o1 Performance on the 2024 Spanish Competitive Medical Specialty Access Examination: Cross-Sectional Quantitative Evaluation Study

Pau Benito¹, MD; Mikel Isla-Jover², MD; Pablo González-Castro³, MD; Pedro José Fernández Esparcia⁴, MD; Manuel Carpio⁵, MD; Iván Blay-Simón⁶, MD; Pablo Gutiérrez-Bedia⁷, MD; María J Lapastora⁸, MD; Beatriz Carratalá⁹, LLB, MEd; Carlos Carazo-Casas¹⁰, MD

¹Department of Preventive Medicine and Epidemiology, Clinical Institute of Medicine and Dermatology (ICMiD), Hospital Clínic de Barcelona, Rosselló, 138, ground floor, Barcelona, Spain

¹⁰Department of Otolaryngology, Hospital Universitario Ramón y Cajal, Madrid, Spain

²Department of Radiology, Hospital de Cruces, Barakaldo, Spain

³Department of Plastic and Reconstructive Surgery, Hospital Universitario Virgen del Rocío, Sevilla, Spain

⁴Department of Dermatology, Hospital Universitario Ramón y Cajal, Madrid, Spain

⁵Department of Endocrinology and Nutrition, Santa Lucía University General Hospital, Cartagena, Spain

⁶Department of Dermatology, Hospital Universitario Doctor Peset, Valencia, Spain

⁷Department of Neurology, Hospital Clínico San Carlos, Madrid, Spain

⁸Department of Intensive Care Medicine, Hospital Universitario 12 De Octubre, Madrid, Spain

⁹Innovation and Digital Projects Academic Department, Healthcademia, Madrid, Spain

Corresponding Author:

Pau Benito, MD

Department of Preventive Medicine and Epidemiology, Clinical Institute of Medicine and Dermatology (ICMiD), Hospital Clínic de Barcelona, Rosselló, 138, ground floor, Barcelona, Spain

Abstract

Background: In recent years, generative artificial intelligence and large language models (LLMs) have rapidly advanced, offering significant potential to transform medical education. Several studies have evaluated the performance of chatbots on multiple-choice medical examinations.

Objective: The study aims to assess the performance of two LLMs—GPT-4o and OpenAI o1—on the *Médico Interno Residente* (MIR) 2024 examination, the Spanish national medical test that determines eligibility for competitive medical specialist training positions.

Methods: A total of 176 questions from the MIR 2024 examination were analyzed. Each question was presented individually to the chatbots to ensure independence and prevent memory retention bias. No additional prompts were introduced to minimize potential bias. For each LLM, response consistency under verification prompting was assessed by systematically asking, “Are you sure?” after each response. Accuracy was defined as the percentage of correct responses compared to the official answers provided by the Spanish Ministry of Health. It was assessed for GPT-4o, OpenAI o1, and, as a benchmark, for a consensus of medical specialists and for the average MIR candidate. Subanalyses included performance across different medical subjects, question difficulty (quintiles based on the percentage of examinees correctly answering each question), and question types (clinical cases vs theoretical questions; positive vs negative questions).

Results: Overall accuracy was 89.8% (158/176) for GPT-4o and 90% (160/176) after verification prompting, 92.6% (163/176) for OpenAI o1 and 93.2% (164/176) after verification prompting, 94.3% (166/176) for the consensus of medical specialists, and 56.6% (100/176) for the average MIR candidate. Both LLMs and the consensus of medical specialists outperformed the average MIR candidate across all 20 medical subjects analyzed, with $\geq 80\%$ LLMs’ accuracy in most domains. A performance gradient was observed: LLMs’ accuracy gradually declined as question difficulty increased. Slightly higher accuracy was observed for clinical cases compared to theoretical questions, as well as for positive questions compared to negative ones. Both models demonstrated high response consistency, with near-perfect agreement between initial responses and those after the verification prompting.

Conclusions: These findings highlight the excellent performance of GPT-4o and OpenAI o1 on the MIR 2024 examination, demonstrating consistent accuracy across medical subjects and question types. The integration of LLMs into medical education presents promising opportunities and is likely to reshape how students prepare for licensing examinations and change our

understanding of medical education. Further research should explore how the wording, language, prompting techniques, and image-based questions can influence LLMs' accuracy, as well as evaluate the performance of emerging artificial intelligence models in similar assessments.

(*JMIR Med Educ* 2026;12:e75452) doi:10.2196/75452

KEYWORDS

accuracy; artificial intelligence; GPT-4o; large language models; medical education; medical examination; Médico Interno Residente; MIR 2024 examination; OpenAI o1

Introduction

The continuous developments in recent years have positioned generative artificial intelligence (AI) as a topic of paramount public and scientific interest. These developments have resulted in the creation of gradually more sophisticated and efficient large language models (LLMs) [1].

Some of the most prominent examples are the increasingly advanced models derived from the GPT family, developed by OpenAI, which rely on deep neural networks [2]. In 2024, OpenAI released 2 highly promising models. Out of which one was GPT-4o, launched in May 2024, a multimodal model capable of processing text and image inputs and generating text outputs in real time. GPT-4o stands out in terms of rapid response times and efficiency [3]. The other one was OpenAI o1, launched in September 2024, a model only capable of processing and generating text, but trained with large-scale reinforcement learning (RL) to reason using chain of thought (CoT) and so possessing advanced reasoning capabilities, surpassing GPT-4o in competitive programming, mathematics, and scientific reasoning [4]. Despite the absence of comprehensive benchmark sets providing consistent evidence, it is reasonable to expect that the presence of sophisticated built-in reasoning-optimized mechanisms in LLMs such as OpenAI o1—trained with RL and CoT—diminishes the relative impact of complex prompting strategies. In such cases, simple zero-shot prompting may prove more effective, or at least equally effective, compared to few-shot and chain-of-thought prompting [4,5].

Chatbots for daily use have emerged to provide virtual assistance, personalized solutions, and task automation in a wide range of fields, including medical education, which has also embraced this trend [6]. Chatbots can be used as a learning aid to improve clinical skills at the undergraduate, residency training, and postgraduate levels of continuous medical education [7].

There are several previous experiences evaluating the performance of chatbots answering multiple-choice questions [7], including medical board examinations like the United States Medical Licensing Examination (USMLE) [8-10]. These assessments are helpful to understand the state of the art regarding LLMs' performance in medical examinations. Furthermore, they pose questions about and provide insightful information to shape the content and characteristics of medical education and examinations.

In Spain, similarly to the USMLE, doctors are examined prior to the beginning of their specialized training. The test is called

the *Médico Interno Residente* (MIR) examination and consists of a 4.5-hour-long examination that includes 210 multiple-choice questions with 4 options and only 1 correct answer. It is held on a yearly basis. The examination serves a double purpose. On the one hand, it is used to rank physicians to assess their eligibility for competitive medical specialist training positions. On the other hand, it ensures minimum requirements are met among candidates.

Evidence suggests a strong correlation between LLM performance across different input languages and the representativeness of each language in the pre-training corpus, a relationship that extends to retrieval-augmented generation LLMs [11,12]. To our knowledge, no study has evaluated GPT-4o's performance on the Spanish MIR examination to date and, more importantly, no study has compared it to OpenAI o1, which owns enhanced reasoning capabilities that could potentially be an advantage when taking the MIR examination [4]. In a broader sense, there are few published studies that have evaluated the performance of LLMs when responding to medical questions in the Spanish language. Among them, the study by Guillen-Grima et al [13] reported a remarkable accuracy rate of 87% for GPT-4 on the 2022 MIR examination, while the study by Flores-Cohaila et al [14] showed an accuracy rate of 86% for GPT-4 on the 2022 Peruvian National Licensing Medical Examination. Other studies posing questions in Spanish from specific medical subjects have shown similar results, with performance rates of 83.7% for GPT-4o in anesthesiology [15] and 93.7% for GPT-4 in rheumatology [16].

The primary aim of this study is to assess the performance of GPT-4o and OpenAI o1 LLMs in passing the MIR examination and to compare them with the expert consensus from instructors of one of the largest MIR preparation academies (Academia AMIR) and the students' mean results. The secondary aim of this study is to compare the performance of GPT-4o, OpenAI o1, expert consensus from AMIR instructors and students by medical subjects, question difficulty, and type of question (clinical case vs theoretical question and positive vs negative question) to better characterize AI chatbots' capabilities, limitations, strengths, and weaknesses.

Methods

Study Design

This is a cross-sectional study assessing the performance of 2 LLMs (GPT-4o and OpenAI o1) in answering the MIR 2024 examination questions. The study compares the models' performance against each other and against specifically trained

humans (expert consensus from AMIR instructors and the mean results from MIR 2024 examination candidates).

MIR 2024 Examination

In Spain, there are 46 medical specialties, each requiring a specific training period of 4 to 5 years as a resident physician (MIR) in an accredited health care institution. Access to each specialty training spot depends on the national ranking of candidates. The ranking is based on a final grade which comes from the MIR examination score (90%) and the candidate's academic record (10%) [17,18].

A total of 15,114 candidates were admitted to the MIR 2024 examination, of whom 13,711 sat for the test, competing for 9007 specialty positions available in accredited healthcare institutions across Spain [18].

The examination, held on January 25, 2025, consisted of 200 multiple-choice questions, each with 4 answer choices, with only 1 correct option. The first 25 questions included linked images that were part of the questions' content and could help or be necessary to answer them. Additionally, 10 reserve questions were included to replace any disputed questions due to typographical errors, ambiguous wording, or issues with multiple or missing correct answers. Participants were given 4 hours and 30 minutes to complete the examination [17,18].

As a safeguard against academic misconduct (ie, cheating), the MIR examination is administered in several different versions each year. Each version comprises an identical question set with a varied sequence. Version 0 is established as the canonical version for scoring and for the publication of the official answer key.

Version 0 of the MIR 2024 examination was obtained from the Spanish Ministry of Health website [19]. In the final analysis, the 25 questions requiring image interpretation were excluded since one of the LLMs evaluated (OpenAI o1) does not accept image inputs. This decision aimed to ensure fair comparability across all study arms, as providing images only to human participants and the other LLM (GPT-4o) would have introduced a systematic advantage for them. The 5 questions whose objections were accepted by the Spanish Ministry of Health were also excluded from the final analysis in order to approximate the real examination as closely as possible (the Spanish Ministry of Health accepted the objection for 6 questions but one of them was linked to an image, so it was already eliminated from the analysis). Nonetheless, the performance of the 4 study arms on these questions is also reported and discussed later in the study. Of the reserve questions, the 4 questions that did not replace any challenged question were also discarded from the analysis. Final analysis included 176 questions.

Study Arms

This study compares the 4 distinct arms as follows: GPT-4o, OpenAI o1, expert consensus from AMIR instructors (henceforth "AMIR consensus"), and mean results of the MIR 2024 examination candidates (henceforth "students").

GPT-4o

This model, OpenAI's flagship in 2024, is characterized by rapid response times and efficiency. Although it is a multimodal model, in this study, only its text processing and generation capabilities were used. Image-linked questions were excluded to ensure comparability with OpenAI o1 model.

OpenAI o1

Designed to use large-scale RL and CoT reasoning, this model exhibits advanced reasoning capabilities, which could be particularly useful for answering questions in the MIR examination.

AMIR Consensus

Academia AMIR is a private, for-profit educational company operating in Spain, Portugal, and several Latin American countries, providing postgraduate health sciences training. Its core activities include preparing candidates for official examinations such as the MIR examination. The company employs faculty members who deliver these courses but are independent from the MIR examination process. They neither contribute to the examination's development nor belong to any public organization involved in its preparation. A panel of these faculty—at least 2 per medical specialty—collaboratively answered the entire MIR 2024 examination after its administration and official content release by the Spanish Ministry of Health. Meeting in a hybrid format (combining in-person and remote participation), they established a consensus through discussion answering within 4.5 hours, mirroring the time allotted to candidates. The goal was to give candidates prompt performance feedback before the official answer key was published. Therefore, the faculty had unrestricted access to textbooks, scientific literature, and LLMs, though reported use was minimal and reserved for clarifying ambiguous questions prior to group consensus. This process produced an expert consensus answer list for the MIR 2024 examination.

Students

Candidates who took the examination were encouraged to submit their answer templates to EstimAMIR, an online platform developed by Academia AMIR. This platform provides students with a preliminary assessment of their results, initially using "AMIR consensus" answers and later incorporating the provisional and definitive correct answers published by the Spanish Ministry of Health. The platform also estimated each student's ranking position based on the sample of candidates available. The mean results of the "students", as well as the percentage of correct answers for each question, were obtained from this platform (based on 5066 answer templates submitted). All data were appropriately anonymized and aggregated.

Data Collection

The 185 text-based questions from the MIR 2024 examination were collected and transcribed verbatim in Spanish into the dialogue interface of both GPT-4o and OpenAI o1. A ChatGPT Plus license was used to access the GPT-4o and OpenAI o1. The models were used with their default settings, with no modifications to parameters such as temperature or output variation. Each multiple-choice question was followed by the

4 possible answer choices (1, 2, 3, and 4), which were manually entered and separated by single spaces. No pretraining or standardized instructions were provided, adhering strictly to a zero-shot prompting approach to minimize potential bias. Henceforth, the results generated using this prompt will be designated as the first iteration results.

Questions were presented to the chatbots individually, with a new dialogue initiated for each question to ensure independence and prevent memory retention bias. To assess response consistency, chatbots were systematically challenged with the verification prompt “Are you sure?” after each answer, which served as a single CoT prompt. Hereafter, we referred to the results obtained with this prompt as the second iteration results. For GPT-4o, internet access was disabled during testing.

All responses were recorded in a spreadsheet. Once the definitive official answers were published by the Spanish Ministry of Health, any challenged or unused reserve questions were excluded from the final analysis.

Main Endpoint and Additional Analysis

Overview

The primary endpoint was the percentage of correct answers per study arm. The definitive official answers published by the Spanish Ministry of Health served as the gold standard for determining accuracy within each study arm. The secondary endpoints included comparisons of study arm performance based on medical subjects, question difficulty, and question type (clinical case vs theoretical question and positive vs negative question). Categorization was conducted as described below.

Medical Subject

Questions were classified into the following categories such as gastroenterology and general surgery, endocrinology, infectious diseases and microbiology, miscellaneous and basic sciences, neurology and neurosurgery, cardiology and cardiovascular surgery, gynecology and obstetrics, orthopedic surgery, pediatrics, nephrology, respiratory medicine and thoracic surgery, rheumatology, hematology, psychiatry, immunology, urology, dermatology, ophthalmology, otorhinolaryngology, and statistics and epidemiology.

Question Difficulty

Difficulty was categorized based on the percentage of examinees who correctly answered each question, using data from EstimAMIR (very difficult: 0% - 20% correct responses; difficult: 21% - 40% correct responses; intermediate: 41% - 60% correct responses; easy: 61% - 80% correct responses; very easy: 81% - 100% correct responses).

Theoretical Question Versus Clinical Case

Questions were classified as theoretical (requiring a direct answer based exclusively on theoretical knowledge) and clinical case (presenting a clinical scenario from which the possible answers emerged).

Positive Versus Negative Questions

Questions were classified based on whether they asked for the correct answer (or the next appropriate step) or the incorrect answer (or the step that should not be taken).

Statistical Analysis

Comparisons between study arms were performed using chi-squared tests or the Fisher exact test, where applicable. The Benjamini-Hochberg method was applied to statistically adjust for multiple comparisons. Differences between groups were considered statistically significant if $P < .05$. We assessed the consistency of responses from GPT-4o and OpenAI o1 to a verification prompt (“Are you sure?”). Consistency was measured using both the simple agreement percentage and the Cohen κ coefficient between the answers provided before and after the prompt. All statistical analyses were performed using R version 4.4.1 (R Foundation for Statistical Computing).

Ethical Considerations

In the absence of a formal ethics committee at Academia AMIR, an ad hoc data ethical oversight panel approved a comprehensive data use protocol (internal reference: AMIR-ETH-2025-11-05-v1.0). At enrollment, all students provided data-use consent, acknowledging that their irreversibly anonymized and aggregated data could be used by Academia AMIR for statistical, commercial, educational, research, and product improvement purposes. The panel determined that the study qualified for exemption from institutional review board approval, as it involved secondary data that were aggregated, processed without human intervention, and contained no identifiable information, in accordance with principles of risk proportionality and data minimization. Participant privacy and confidentiality were safeguarded through irreversible anonymization and aggregation prior to investigator access, data minimization, role-based access controls, and encryption of data both in transit and at rest within corporate repositories. No participants received compensation for their participation in the study.

Results

Global Performance

A flowchart of the exclusion criteria for the question selection process is shown in [Figure 1](#).

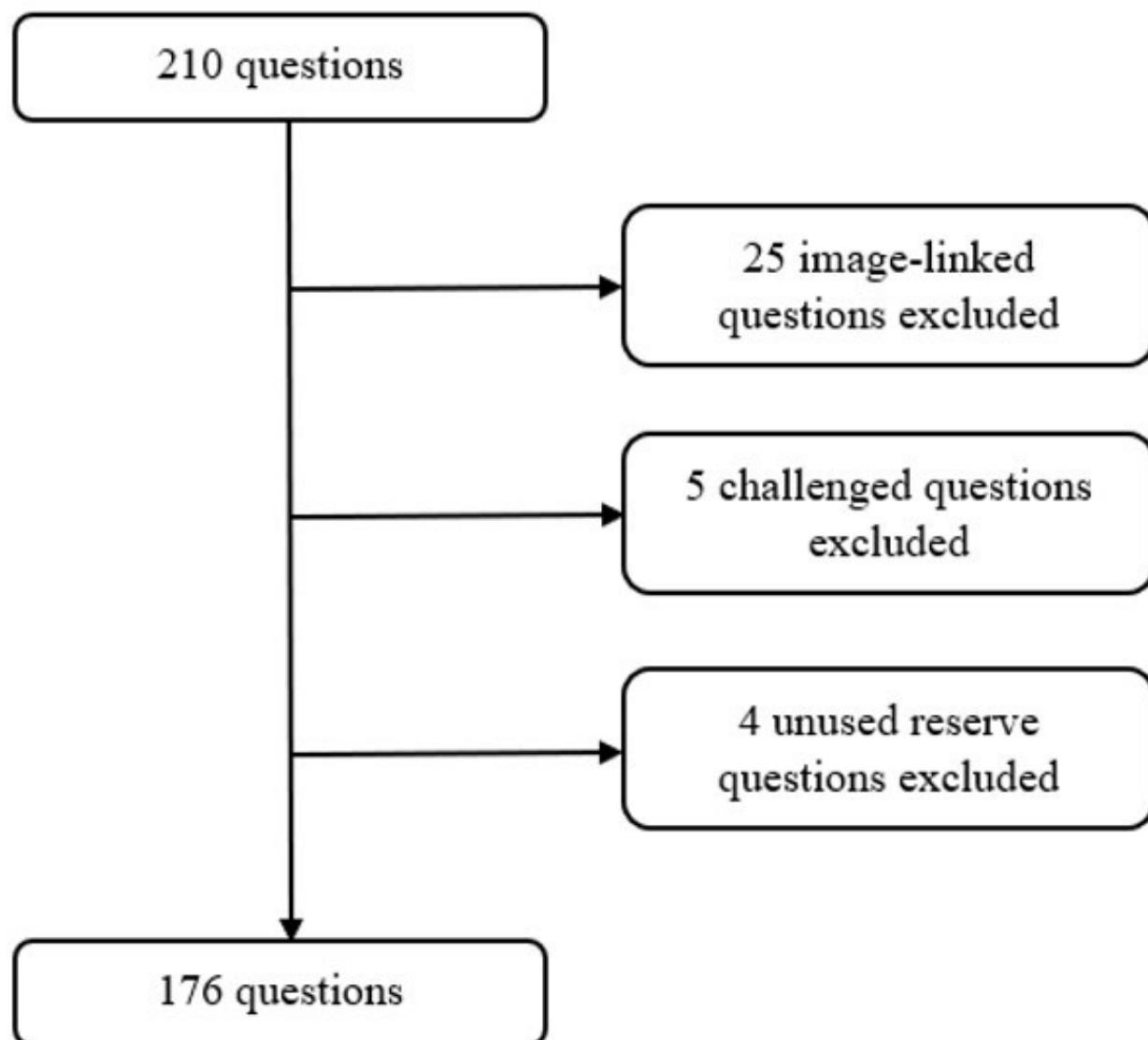
Figure 1. Flowchart of the exclusion criteria for the question selection process.

Table 1 presents the accuracy of GPT-4o, OpenAI o1, the AMIR consensus, and students when taking the entire examination, excluding the discarded questions. GPT-4o achieved an accuracy of 89.8% (158/176) in the first iteration, which slightly increased to 90.9% (160/176) in the second iteration. Similarly, OpenAI

o1's accuracy was 92.6% (163/176) in the first iteration and improved to 93.2% (164/176) in the second iteration. The AMIR instructors' consensus obtained the highest score among the study arms, with 94.3% (166/176). The mean score of the EstimAMIR-submitted templates was 56.6% (100/176).

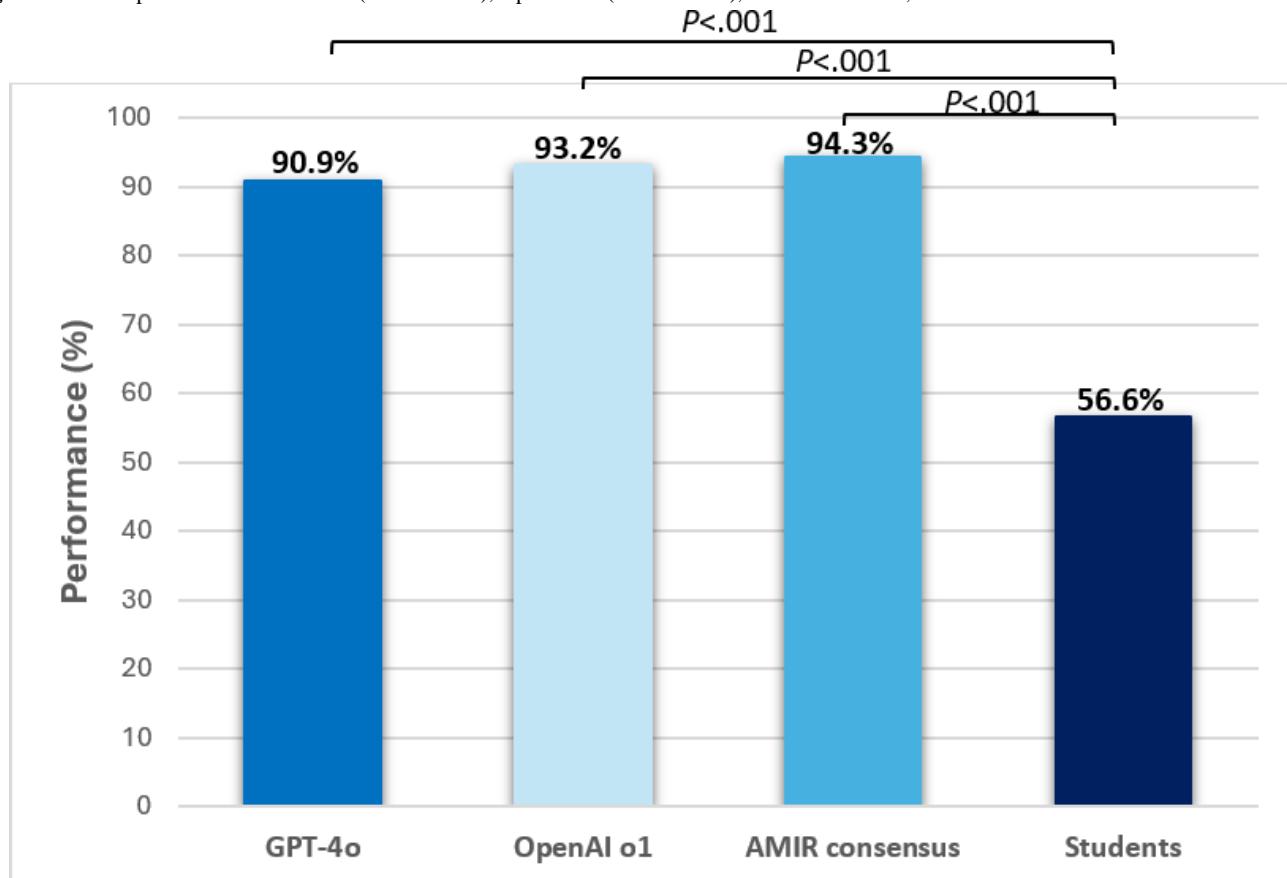
Table 1. Global performance of GPT-4o (first and second iteration), OpenAI o1 (first and second iteration), AMIR consensus, and students.

	Absolute and relative performance	
	Correct answers, n (%)	Incorrect answers, n (%)
GPT-4o		
First iteration	158 (89.8)	18 (10.2)
Second iteration	160 (90.9)	16 (9.1)
OpenAI o1		
First iteration	163 (92.6)	13 (7.4)
Second iteration	164 (93.2)	12 (6.8)
AMIR consensus	166 (94.3)	10 (5.7)
Students	100 (56.6)	76 (43.4)

Figure 2 compares the accuracy of GPT-4o and OpenAI o1 in their second iterations, along with the AMIR consensus and students. GPT-4o, OpenAI o1, and the AMIR consensus achieved significantly higher accuracy scores than the average

student (in all cases $P<.001$); however, differences between these 3 arms did not reach statistical significance ($P=.22$ for GPT-4o vs OpenAI o1; $P=.07$ for GPT-4o vs AMIR consensus; $P=.75$ for OpenAI o1 vs AMIR consensus).

Figure 2. Global performance of GPT-4o (2nd iteration), OpenAI o1 (2nd iteration), AMIR consensus, and students.

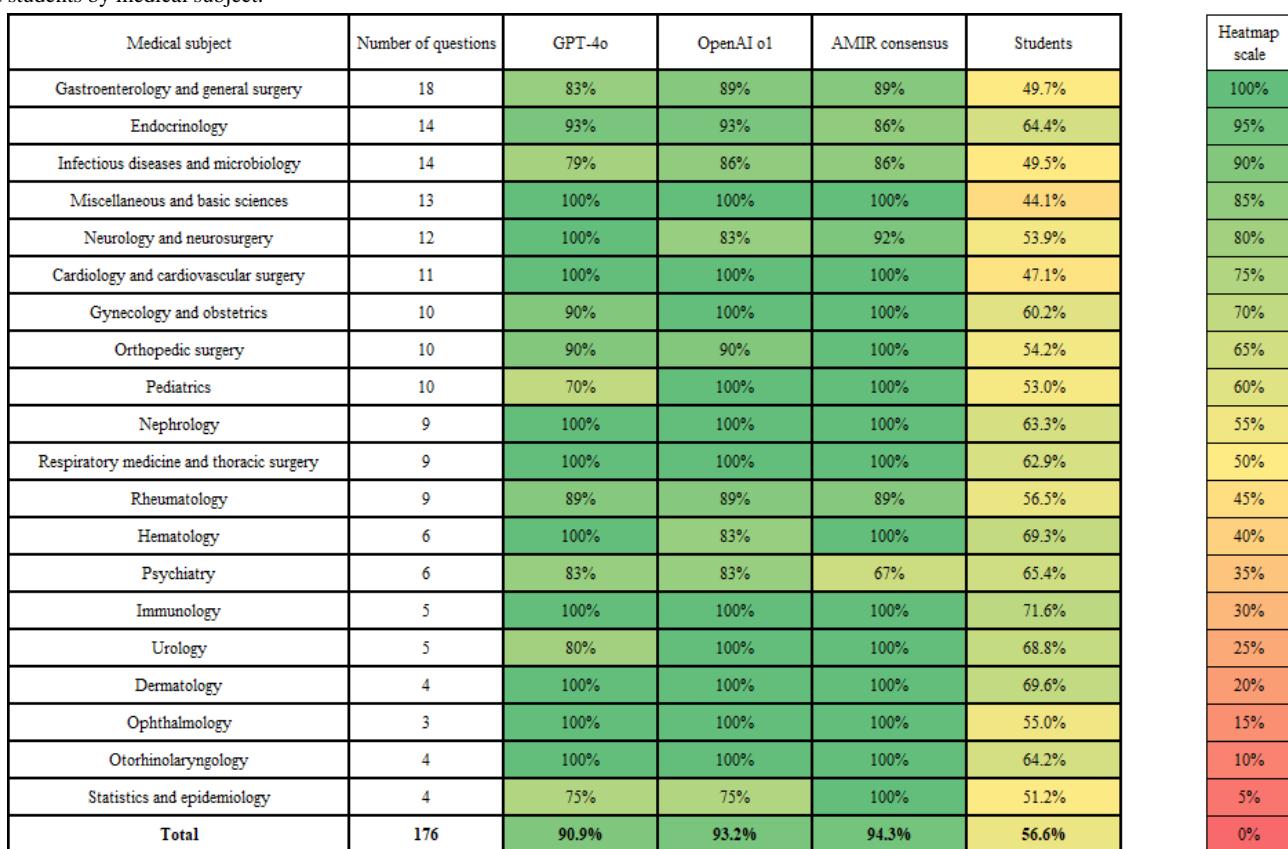


From the final analysis, 5 challenged questions were excluded. On these items, GPT-4o achieved an accuracy of 100% (5/5) across both iterations. OpenAI o1 scored 80% (4/5) in both iterations, without modifying any of its responses. The AMIR consensus achieved an accuracy of 40% (2/5). When restricted to these 5 questions, the students' mean score was 31.3% (SD 16.7%).

Medical Subjects

The heatmap in **Figure 3** presents a comparative analysis of study arm performance across different medical subjects. Both LLMs (GPT-4o and OpenAI o1, in their second iterations) and the AMIR consensus outperformed the average student in all 20 medical subjects analyzed.

Figure 3. Heatmap comparing the performance of generative pre-trained transformer 4o (2nd iteration), OpenAI o1 (2nd iteration), and AMIR consensus and students by medical subject.



GPT-4o achieved an accuracy below 80% in only 3 subjects: infectious diseases and microbiology (11/14, 79%), pediatrics (7/10, 70%), and statistics and epidemiology (3/4, 75%). OpenAI o1 fell below 80% accuracy only in statistics and epidemiology, with a single error (3/4, 75%). The AMIR consensus exhibited an accuracy lower than 80% in psychiatry (4/6, 67%). Average student performance, based on 5066 EstimAMIR-submitted templates, ranged from 44.1% in miscellaneous and basic sciences to 71.6% in immunology.

Question Difficulty

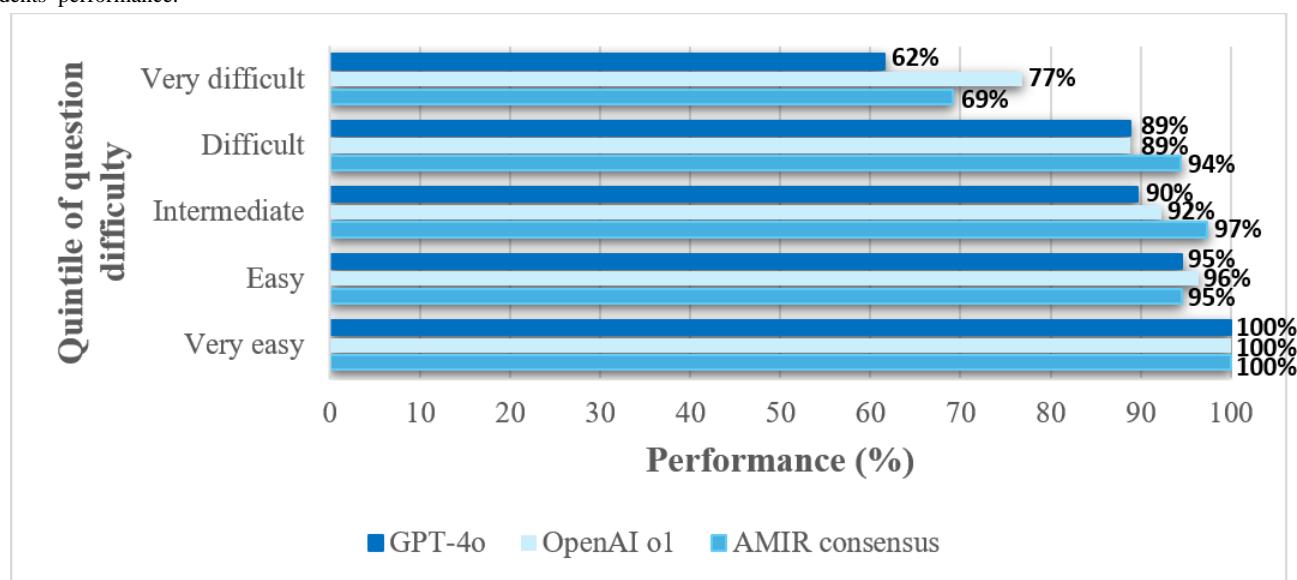
As shown in **Table 2** and **Figure 4** the performance of GPT-4o, OpenAI o1, and the AMIR consensus across quintiles of

question difficulty, defined based on student performance. A statistically significant gradient is observed among these 3 study arms, with accuracy decreasing as question difficulty increases (crude *P* values: GPT-4o, *P*=.003; OpenAI o1, *P*=.04; AMIR consensus, *P*=.008; Benjamini-Hochberg adjusted *P* values: GPT-4o, *P*=.03; OpenAI o1, *P*=.10—the only case not reaching statistical significance; AMIR consensus, *P*=.03). The decline in performance is particularly pronounced in the highest difficulty quintile. Differences between study arms within each difficulty quintile do not reach statistical significance.

Table 4. Absolute and relative performance of GPT-4o (second iteration), OpenAI o1 (second iteration), and AMIR consensus by quintiles of question difficulty defined by students' performance.

Question difficulty	Number of questions	GPT-4o, n (%)	OpenAI o1, n (%)	AMIR consensus, n (%)
Very easy	32	32 (100)	32 (100)	32 (100)
Easy	56	53 (95)	54 (96)	53 (95)
Intermediate	39	35 (90)	36 (92)	38 (97)
Difficult	36	32 (89)	32 (89)	34 (94)
Very difficult	13	8 (62)	10 (77)	9 (69)
Total	176	160 (90.9)	164 (93.2)	166 (94.3)

Figure 4. Performance of GPT-4o (2nd iteration), OpenAI o1 (2nd iteration), and AMIR consensus by quintiles of question difficulty defined by students' performance.



Clinical Cases versus Theoretical Questions

As shown in [Table 3](#) the performance of GPT-4o, OpenAI o1, the AMIR consensus, and students when answering clinical cases versus theoretical questions. Overall, a slightly higher accuracy was observed for clinical cases compared to theoretical

Table 3. Performance of GPT-4o (second iteration), OpenAI o1 (second iteration), AMIR consensus, and students by questions being clinical cases or theoretical questions.

	Clinical cases (n=105), n (%)	Theoretical questions (n=71), n (%)	P value
GPT-4o	98 (93.3)	62 (87)	.55
OpenAI o1	99 (94.2)	65 (92)	.68
AMIR consensus	99 (94.2)	67 (94)	.99
Students	62 (59)	38 (54)	.68

Positive Versus Negative Questions

[Table 4](#) shows the performance of GPT-4o, OpenAI o1, the AMIR consensus, and students when answering positive versus negative questions. Overall, accuracy was higher for positive

questions, although these differences do not reach statistical significance in any study arm. Statistically significant differences between study arms both for clinical cases and theoretical questions were observed only when the students' arm was included in the analysis.

Table 4. Performance of GPT-4o (second iteration), OpenAI o1 (second iteration), AMIR consensus, and students by questions being positive or negative.

	Positive questions (n=140), n (%)	Negative questions (n=36), n (%)	P value
GPT-4o	132 (94.2)	28 (78)	.01
OpenAI o1	132 (94.2)	32 (89)	.40
AMIR consensus	132 (94.2)	34 (94)	1
Students	82 (58.5)	18 (49)	.55

Response Consistency

Response consistency was assessed using the simple agreement percentage between the first and second iterations of GPT-4o (172/176, 97.7%) and OpenAI o1 (170/176, 96.6%). The Cohen κ coefficient was 0.97 for GPT-4o and 0.95 for OpenAI o1, indicating almost perfect agreement ($P<.001$ in both cases).

questions than for negative ones, with the difference reaching statistical significance only for GPT-4o ($P=.01$). Statistically significant differences between study arms for both positive and negative questions are observed only when the students' arm is included in the analysis.

When analyzing individually the questions in which there was no concordance between the first and second iterations, it was observed that, for GPT-4o, 4 initially incorrect responses were modified: in 2 cases, the second response was also incorrect, while in the other 2 cases, the second response became correct. For OpenAI o1, 5 initially incorrect responses were modified: in 3 cases, the second response was again incorrect, and in 2 cases, the second response became correct. In addition, 1

initially correct response was modified, with the second response becoming incorrect.

Discussion

Principal Results

This study highlights the exceptional performance of both LLMs analyzed—GPT-4o and OpenAI o1—on the MIR 2024 examination. Both models achieved or exceeded a 90% accuracy rate, significantly outperforming the average human candidate as well as the top 10% of examinees [18]. The expert consensus from AMIR instructors yielded even higher accuracy. Although this result should be interpreted in the context of unrestricted access to textbooks, scientific literature, and AI tools such as GPT, the reported use of these resources was minimal and reserved for clarifying ambiguous questions, never substituting for the group discussion and consensus process for each item. Results from the expert consensus suggest the added value of human expertise when synergistically combined with AI capabilities.

The challenged questions were excluded from the final analysis to remain faithful to the actual examination. Upon examination, these proved to be difficult items (on average, candidates answered them correctly in 31.3% of cases, compared with 56.6% for the other questions), which the LLMs managed more accurately than the human experts (5/5 for GPT-4o, 4/5 for OpenAI o1, and 2/5 for AMIR instructors).

These results were consistent across the different medical subjects analyzed. Interestingly, when question difficulty was assessed based on human performance, a similar trend was observed in the LLMs, with accuracy decreasing as question difficulty increased. Additionally, a slightly higher accuracy was observed for clinical cases compared to theoretical questions, as well as for positive questions compared to negative ones. This resemblance to human reasoning and performance could be rooted in the input used to train LLMs.

Both GPT-4o and OpenAI o1 demonstrated great consistency in their answers, with statistically significant near-perfect agreement between the first and second iterations. Furthermore, it is particularly interesting to note that, of the 10 responses that were altered between the first and second iterations (4 for GPT-4o and 6 for OpenAI o1), 9 were initially incorrect (with 4 of these changing to a correct response), and only 1 initially correct response was changed to an incorrect one. It is remarkable in favor of these LLMs that, considering their exceptional accuracy in the first iteration, the few changes occurring in the second iteration almost exclusively involved some of the few initially incorrect responses.

Comparison With Prior Work

Several previous studies have evaluated the performance of GPT-3.5 and GPT-4 across different medical disciplines, as well as on national medical board examinations [7]. For instance, Gilson et al assessed ChatGPT's performance on various sets of USMLE step 1 and step 2 questions, reporting an accuracy range of 42% to 64% [8], which aligns with another study that found a 56% accuracy rate on a set of USMLE step 1-style questions [9]. Knoedler et al examined ChatGPT-3.5 and

ChatGPT-4 on USMLE step 3 questions, reporting 57% accuracy for GPT-3.5 and 85% for GPT-4 [10]. Takagi et al evaluated these models on the Japanese Medical Licensing Examination, finding 51% accuracy for GPT-3.5 and 80% for GPT-4 [20]. Meyer et al conducted a similar study on the written German medical licensing examination, with accuracy rates of 58% for GPT-3.5 and 85% for GPT-4 [21]. A study by Prazeres [22] on the Portuguese national examination for access to specialized training reported 54% accuracy for GPT-3.5 turbo and 65% for GPT-4o mini. Guillen-Grima et al [13] published the perhaps most comparable study, as they compared GPT-3.5 and GPT-4 on the MIR 2022 examination. Accuracy rates were 63% for GPT-3.5 and 87% for GPT-4. Interestingly, our results show the highest accuracy for LLMs among all the aforementioned studies, almost matching the consensus from expert human instructors. This may be a result of the gradual development of LLMs with time. It poses relevant questions regarding how medical education and examinations should be shaped in the future, both in terms of content and the skills that are underscored. As stated in a previous editorial, these results make it important to consider the necessity of more emphasis on soft skills and critical thinking rather than plain memorization [23].

Strengths

This study offers new insights into the accuracy of GPT-4o and OpenAI o1 in a national medical specialty access examination. To date, the only comparable research published in an indexed journal that we have identified is a recent study by Liu et al [24], which evaluated GPT-4o's performance on the Japanese national medical examination, reporting an accuracy of 89%.

Moreover, this study reinforces the trend that each newly developed LLM exhibits improved accuracy compared to its predecessors. Additionally, our secondary analysis proved that the accuracy of both GPT-4o and OpenAI o1 aligned with difficulty levels of questions based on human candidates' performance.

Limitations

This study has several limitations. First, increasing the sample size—that is, including a larger number of questions in the analysis—would have provided a more robust insight into the different subanalyses performed. For instance, it would have allowed us to investigate whether performance differences of LLMs compared to human experts—such as those suggested in unique medical subjects like psychiatry—are truly meaningful or just the results of random variation. Second, additional secondary analyses could have been of interest, such as examining the relationship between the number of words or characters in each question and the performance of LLMs; the influence of specific expressions or wording styles on model accuracy; the impact of different languages on performance; and the effect of alternative prompting formulas on accuracy. Third, although image-based questions are part of the MIR examination, they were not included in this study because OpenAI o1 does not support image inputs, and fair comparability between LLMs was prioritized. This decision reduces methodological bias—LLMs are not artificially penalized for lacking multimodal capabilities—and increases the internal

validity of between-arms comparisons. However, it may reduce the representativeness of the performance evaluation and limit the generalizability of our findings to the actual test setting, where visual interpretation is an integral component of clinical reasoning. Previous studies suggest LLMs may exhibit reasonable performance on image-based questions even without access to the image itself [25]. Fourth, the AMIR consensus may not represent a pure benchmark of human expert knowledge, as experts had access to textbooks and generative AI. However, the faculty use of these resources was minimal and strictly advisory, with all final answers determined by expert discussion and consensus, indicating that the potential for significant bias was low. Fifth, the field of LLMs is continuously evolving, and several new models have been released in recent months that were not analyzed in this study, including GPT-4.5 by OpenAI [26], DeepSeek-R1 by DeepSeek-AI [27], Qwen 2.5 by Alibaba [28], LLaMa 3.2 by Meta AI [29], and Claude 3.7 Sonnet by Anthropic [30], among others. Sixth, student results were self-reported, which could be a source of bias. Finally, caution should be exercised when generalizing LLM

accuracy on the MIR examination to other national medical licensing examinations or to different fields and tasks within medical education.

Conclusions

This study highlights the excellent performance of the two analyzed LLMs—GPT-4o and OpenAI o1—on the MIR 2024 examination, demonstrating strong consistency across different medical subjects and types of questions, as well as between first and second iterations.

The integration of LLMs into medical education is promising and likely to revolutionize the field and change our understanding of medical education. Further research is needed to explore how wording, language, prompting techniques, and image-based questions influence LLM accuracy in national medical licensing examinations, as well as to assess the performance of other emerging models. More research is also needed to better understand the potential usefulness of these tools as learning assistants in broader educational contexts.

Acknowledgments

The authors thank the students who entered their MIR examination templates into the EstimAMIR application, which provided a benchmark for assessing the performance of the LLMs evaluated in this study. The authors also appreciate the interest of the AMIR Academy faculty in this study, especially the instructors, whose responses to the MIR 2024 examination questions allowed us to design the comparison group called AMIR consensus.

Funding

The expenses associated with the preparation and publication of this study were funded by Healthcademia.

Data Availability

The data from this study are available upon reasonable request.

Authors' Contributions

Conceptualization: CCC (lead); PB and MIJ (equal)

Data curation: PB and CCC (lead); MIJ, PGC, PJFE, MCP, IBS, PGB, MJL, and BCO (equal)

Formal analysis: PB (lead), MIJ (supporting)

Project administration: CCC (lead), PB (supporting)

Writing – original draft: PB

Writing – review & editing: PB, CCC, MIJ, PGC, PJFE, MCP, IBS, PGB, MJL, BCO (equal)

Conflicts of Interest

None declared.

References

1. Zhao WX, Zhou K, Li J, Tang T, Wang X, Hou Y, et al. A survey of large language models. arXiv. Preprint posted online on Mar 31, 2023. [doi: [10.48550/arXiv.2303.18223](https://doi.org/10.48550/arXiv.2303.18223)]
2. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. : OpenAI; 2018 URL: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf [accessed 2025-12-10]
3. OpenAI, Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, et al. GPT-4 technical report. arXiv. Preprint posted online on Mar 15, 2023. [doi: [10.48550/arXiv.2303.08774](https://doi.org/10.48550/arXiv.2303.08774)]
4. OpenAI, Jaech A, Kalai A, Lerer A, Richardson A, El-Kishky A, et al. OpenAI o1 system card. arXiv. Preprint posted online on Dec 21, 2024. [doi: [10.48550/arXiv.2412.16720](https://doi.org/10.48550/arXiv.2412.16720)]
5. Wang G, Sun Z, Ye S, et al. Do advanced language models eliminate the need for prompt engineering in software engineering? ACM Trans Softw Eng Methodol 2025. [doi: [10.1145/3771933](https://doi.org/10.1145/3771933)]

6. Calderini G, Jaf S, McGarry K. A literature survey of recent advances in chatbots. *Information* 2022;13(1):41. [doi: [10.3390/info13010041](https://doi.org/10.3390/info13010041)]
7. Hallquist E, Gupta I, Montalbano M, Loukas M. Applications of artificial intelligence in medical education: a systematic review. *Cureus* 2025 Mar;17(3):e79878. [doi: [10.7759/cureus.79878](https://doi.org/10.7759/cureus.79878)] [Medline: [40034416](https://pubmed.ncbi.nlm.nih.gov/40034416/)]
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](https://doi.org/10.2196/45312)] [Medline: [36753318](https://pubmed.ncbi.nlm.nih.gov/36753318/)]
9. Knoedler L, Knoedler S, Hoch CC, et al. In-depth analysis of ChatGPT's performance based on specific signaling words and phrases in the question stem of 2377 USMLE step 1 style questions. *Sci Rep* 2024 Jun 12;14(1):13553. [doi: [10.1038/s41598-024-63997-7](https://doi.org/10.1038/s41598-024-63997-7)] [Medline: [38866891](https://pubmed.ncbi.nlm.nih.gov/38866891/)]
10. Knoedler L, Alfertshofer M, Knoedler S, et al. Pure wisdom or Potemkin villages? A comparison of ChatGPT 3.5 and ChatGPT 4 on USMLE step 3 style questions: quantitative analysis. *JMIR Med Educ* 2024 Jan 5;10:e51148. [doi: [10.2196/51148](https://doi.org/10.2196/51148)] [Medline: [38180782](https://pubmed.ncbi.nlm.nih.gov/38180782/)]
11. Li Z, Shi Y, Liu Z, Yang F, Liu N, Du M. Quantifying multilingual performance of large language models across languages. *arXiv*. Preprint posted online on Apr 17, 2024. [doi: [10.48550/arXiv.2404.11553](https://doi.org/10.48550/arXiv.2404.11553)]
12. Gao Y, Xiong Y, Gao X, Jia K, Pan J, Bi Y, et al. Retrieval-augmented generation for large language models: a survey. *arXiv*. Preprint posted online on Dec 18, 2023. [doi: [10.48550/arXiv.2312.10997](https://doi.org/10.48550/arXiv.2312.10997)]
13. 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](https://doi.org/10.3390/clinpract13060130)] [Medline: [37987431](https://pubmed.ncbi.nlm.nih.gov/37987431/)]
14. 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](https://doi.org/10.2196/48039)] [Medline: [37768724](https://pubmed.ncbi.nlm.nih.gov/37768724/)]
15. Altermatt FR, Neyem A, Sumonte NI, et al. Evaluating GPT-4o in high-stakes medical assessments: performance and error analysis on a Chilean anesthesiology exam. *BMC Med Educ* 2025 Oct 27;25(1):1499. [doi: [10.1186/s12909-025-08084-9](https://doi.org/10.1186/s12909-025-08084-9)] [Medline: [41146119](https://pubmed.ncbi.nlm.nih.gov/41146119/)]
16. Madrid-García A, Rosales-Rosado Z, Freites-Nuñez D, et al. Harnessing ChatGPT and GPT-4 for evaluating the rheumatology questions of the Spanish access exam to specialized medical training. *Sci Rep* 2023 Dec 13;13(1):22129. [doi: [10.1038/s41598-023-49483-6](https://doi.org/10.1038/s41598-023-49483-6)] [Medline: [38092821](https://pubmed.ncbi.nlm.nih.gov/38092821/)]
17. Order SND/888/2024 of 14 august, approving the number of available positions and announcing the 2024 competitive selection examinations for access in 2025 to specialized health training positions for university degree holders in medicine, pharmacy, nursing, and in the fields of psychology, chemistry, biology, and physics [Article in Spanish]. : Boletín Oficial del Estado; 2024 URL: <https://www.boe.es/boe/dias/2024/08/23/pdfs/BOE-A-2024-17246.pdf> [accessed 2025-01-07]
18. The Ministry of Health publishes the final list of results of the specialized health training examinations [Website in Spanish]. Ministry of Health, Spain. URL: <https://www.sanidad.gob.es/gabinete/notasPrensa.do?id=6632> [accessed 2025-03-12]
19. Consulting previous examination booklets—search by examination session [Website in Spanish]. Ministry of Health, Spain. URL: <https://fse.sanidad.gob.es/fseweb/#/principal/datosAnteriores/cuadernosExamen> [accessed 2025-01-26]
20. 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](https://doi.org/10.2196/48002)] [Medline: [37384388](https://pubmed.ncbi.nlm.nih.gov/37384388/)]
21. Meyer A, Riese J, Streichert T. Comparison of the performance of GPT-3.5 and GPT-4 with that of medical students on the written German medical licensing examination: observational study. *JMIR Med Educ* 2024 Feb 8;10:e50965. [doi: [10.2196/50965](https://doi.org/10.2196/50965)] [Medline: [38329802](https://pubmed.ncbi.nlm.nih.gov/38329802/)]
22. 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 Mar 5;11:e65108. [doi: [10.2196/65108](https://doi.org/10.2196/65108)] [Medline: [40043219](https://pubmed.ncbi.nlm.nih.gov/40043219/)]
23. Mbakwe AB, Lourentzou I, Celi LA, Mechanic OJ, Dagan A. ChatGPT passing USMLE shines a spotlight on the flaws of medical education. *PLOS Digit Health* 2023 Feb;2(2):e0000205. [doi: [10.1371/journal.pdig.0000205](https://doi.org/10.1371/journal.pdig.0000205)] [Medline: [36812618](https://pubmed.ncbi.nlm.nih.gov/36812618/)]
24. Liu M, Okuhara T, Dai Z, et al. Evaluating the effectiveness of advanced large language models in medical knowledge: a comparative study using the Japanese national medical examination. *Int J Med Inform* 2025 Jan;193:105673. [doi: [10.1016/j.ijmedinf.2024.105673](https://doi.org/10.1016/j.ijmedinf.2024.105673)] [Medline: [39471700](https://pubmed.ncbi.nlm.nih.gov/39471700/)]
25. Gravina AG, Pellegrino R, Palladino G, Imperio G, Ventura A, Federico A. Charting new AI education in gastroenterology: cross-sectional evaluation of ChatGPT and perplexity AI in medical residency exam. *Dig Liver Dis* 2024 Aug;56(8):1304-1311. [doi: [10.1016/j.dld.2024.02.019](https://doi.org/10.1016/j.dld.2024.02.019)] [Medline: [38503659](https://pubmed.ncbi.nlm.nih.gov/38503659/)]
26. OpenAI GPT-4.5 system card. : OpenAI; 2025 URL: <https://cdn.openai.com/gpt-4-5-system-card-2272025.pdf> [accessed 2026-01-06]
27. DeepSeek-AI, Guo D, Yang D, Zhang H, Song J, Zhang R, et al. DeepSeek-R1: incentivizing reasoning capability in llms via reinforcement learning. *arXiv*. Preprint posted online on Jan 22, 2025. [doi: [10.48550/arXiv.2501.12948](https://doi.org/10.48550/arXiv.2501.12948)]
28. Qwen, Yang A, Yang B, Zhang B, et al. Qwen2.5 technical report. *arXiv*. Preprint posted online on Dec 19, 2024. [doi: [10.48550/arXiv.2412.15115](https://doi.org/10.48550/arXiv.2412.15115)]

29. Grattafiori A, Dubey A, et al. The llama 3 herd of models. arXiv. Preprint posted online on Jul 31, 2024. [doi: [10.48550/arXiv.2407.21783](https://doi.org/10.48550/arXiv.2407.21783)]
30. Claude 3.7 sonnet system card. : Anthropic URL: <https://assets.anthropic.com/m/785e231869ea8b3b/original/clause-3-7-sonnet-system-card.pdf> [accessed 2026-01-07]

ABBREVIATIONS

AI: artificial intelligence

CoT: chain of thought

LLM: large language model

MIR: Médico Interno Residente

RL: reinforcement learning

USMLE: United States medical licensing examination

Edited by B Lesselroth; submitted 03.Apr.2025; peer-reviewed by R Pellegrino, T Watari; revised version received 17.Nov.2025; accepted 25.Nov.2025; published 12.Jan.2026.

Please cite as:

Benito P, Isla-Jover M, González-Castro P, Fernández Esparcia PJ, Carpio M, Blay-Simón I, Gutiérrez-Bedia P, Lapastora MJ, Carratalá B, Carazo-Casas C

GPT-4o and OpenAI o1 Performance on the 2024 Spanish Competitive Medical Specialty Access Examination: Cross-Sectional Quantitative Evaluation Study

JMIR Med Educ 2026;12:e75452

URL: <https://mededu.jmir.org/2026/1/e75452>

doi:10.2196/75452

© Pau Benito, Mikel Isla-Jover, Pablo González-Castro, Pedro José Fernández Esparcia, Manuel Carpio, Iván Blay-Simón, Pablo Gutiérrez-Bedia, María J Lapastora, Beatriz Carratalá, Carlos Carazo-Casas. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 12.Jan.2026. 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.

Adaptation of the Japanese Version of the 12-Item Attitudes Towards Artificial Intelligence Scale for Medical Trainees: Multicenter Development and Validation Study

Hirohisa Fujikawa^{1,2,3}, MD, PhD; Hirotake Mori¹, MD, PhD; Kayo Kondo⁴, MA, PhD; Yuji Nishizaki^{1,5}, MD, MPH, PhD; Yuichiro Yano¹, MD, PhD; Toshio Naito¹, MD, PhD

¹Department of General Medicine, Juntendo University Faculty of Medicine, 2-1-1 Hongo, Bunkyo-ku, Tokyo, Japan

²Department of Medical Education Studies, International Research Center for Medical Education, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

³Center for General Medicine Education, School of Medicine, Keio University, Shinjuku-ku, Tokyo, Japan

⁴School of Modern Languages and Cultures, Durham University, Durham, United Kingdom

⁵Division of Medical Education, Juntendo University School of Medicine, Bunkyo-ku, Tokyo, Japan

Corresponding Author:

Hirohisa Fujikawa, MD, PhD

Department of General Medicine, Juntendo University Faculty of Medicine, 2-1-1 Hongo, Bunkyo-ku, Tokyo, Japan

Abstract

Background: In the current era of artificial intelligence (AI), use of AI has increased in both clinical practice and medical education. Nevertheless, it is probable that perspectives on the prospects and risks of AI vary among individuals. Given the potential for attitudes toward AI to significantly influence its integration into medical practice and educational initiatives, it is essential to assess these attitudes using a validated tool. The recently developed 12-item Attitudes Towards Artificial Intelligence scale has demonstrated good validity and reliability for the general population, suggesting its potential for extensive use in future studies. However, to our knowledge, there is currently no validated Japanese version of the scale. The lack of a Japanese version hinders research and educational efforts aimed at understanding and improving AI integration into the Japanese health care and medical education system.

Objective: We aimed to develop the Japanese version of the 12-item Attitudes Towards Artificial Intelligence scale (J-ATTARI-12) and investigate whether it is applicable to medical trainees.

Methods: We first translated the original English-language scale into Japanese. To examine its psychometric properties, we then conducted a validation survey by distributing the translated version as an online questionnaire to medical students and residents across Japan from June 2025 to July 2025. We assessed structural validity through factor analysis and convergent validity by computing the Pearson correlation coefficient between the J-ATTARI-12 scores and scores on attitude toward robots. Internal consistency reliability was assessed using Cronbach α values.

Results: We included 326 participants in our analysis. We used a split-half validation approach, with exploratory factor analysis (EFA) on the first half and confirmatory factor analysis on the second half. EFA suggested a 2-factor solution (factor 1: AI anxiety and aversion; factor 2: AI optimism and acceptance). Confirmatory factor analysis revealed that the model fitness indexes of the 2-factor structure suggested by the EFA were good (comparative fit index=0.914 [>0.900]; root mean square error of approximation=0.075 [<0.080]; standardized root mean square residual=0.056 [<0.080]) and superior to those of the 1-factor structure. The value of the Pearson correlation coefficient between the J-ATTARI-12 scores and the attitude toward robots scores was 0.52, which indicated good convergent validity. The Cronbach α for all 12 items was 0.84, which indicated a high level of internal consistency reliability.

Conclusions: We developed and validated the J-ATTARI-12. The developed instrument had good structural validity, convergent validity, and internal consistency reliability for medical trainees. The J-ATTARI-12 is expected to stimulate future studies and educational initiatives that can effectively assess and enhance the integration of AI into clinical practice and medical education systems.

(*JMIR Med Educ* 2026;12:e81986) doi:10.2196/81986

KEYWORDS

artificial intelligence; AI; attitudes toward artificial intelligence; attitudes toward AI; medical education; medical resident; medical student

Introduction

Artificial intelligence (AI) is undergoing rapid development and integration into health care [1,2]. Although the use of AI has raised a number of potential concerns, including a lack of control over the rapidly progressing technology [3], issues of privacy and data protection [4], ethical problems [5], and the potential loss of human jobs [6], the benefits of AI are expected to far outweigh these concerns. AI contributes to improving patient care through enhancing diagnostic accuracy and providing more personalized therapeutic plans [7]. Moreover, AI has the potential to enhance the training and education of medical professionals by offering benefits including virtual simulation and training, remote education, and the recording of teaching videos [8,9]. Consequently, the use of AI has increased in both clinical practice and medical education.

Individuals may hold divergent views on the prospects and risks of AI and adopt varied attitudes toward it [10]. The attitudes of physicians and medical trainees toward AI will have a major impact on its integration into medical practice and educational activities [11]. Negative attitudes toward AI can lead to skepticism and concerns, thereby impeding its adoption [12]. Conversely, positive attitudes toward AI likely foster trust [13], leading individuals to embrace AI systems and amplify their benefits and possibly resulting in their integration into health care [12].

Thus, assessing physicians' and medical trainees' attitudes toward AI is critical to identifying potential barriers and fostering acceptance. Accordingly, instruments with robust psychometric properties are required. Although several measures for assessing attitudes toward AI have been developed, only a few have been comprehensively evaluated for reliability and validity. In 2024, Stein et al [10] developed the 12-item Attitudes Towards Artificial Intelligence scale (ATTARI-12). This is a unidimensional scale that integrates affective, behavioral, and cognitive facets into a single measure. The authors subsequently confirmed its psychometric properties for the general population [10], and its wide use in future studies is anticipated.

Despite the potential global applicability of the ATTARI-12, there is currently no validated Japanese version. Japan has a distinct cultural context, and attitudes toward AI are likely influenced by cultural and societal norms. For instance, Japanese culture's high uncertainty avoidance tends to result in only a gradual adoption of new technologies [14]. In the absence of a culturally adapted measure, assessments may not accurately reflect the true sentiments of Japanese health care workers and trainees. This gap hinders research and educational efforts aimed at understanding and improving AI integration into the Japanese health care system.

In this study, we developed a Japanese version of the ATTARI-12, which was originally created for the general population, and examined whether it is applicable to medical students and resident physicians. We anticipated that the development of this scale in Japan would promote future research and educational courses that effectively assess and

improve the integration of AI into clinical practice and the medical education system.

Methods

This study formed part of a larger research project regarding AI education for medical trainees. We conducted this study in the following 2 steps.

Step 1: Translation and Cross-Cultural Adaptation

In accordance with an international guideline [15], we translated the original ATTARI-12 into Japanese. First, the first author (HF) asked the creator of the original ATTARI-12 and corresponding author of the article that reported it to allow us to develop the Japanese version. The creator willingly provided permission for our translation of the scale. Second, 2 translators (HF and KK) conducted forward translations independently. Both translators were familiar with Western and Japanese cultures and had rich experience in developing translated versions of scales in the field of health profession education. In particular, KK is a fluent speaker of Japanese and English. Third, the translators performed a synthesis of their translations. Discrepancies were resolved through discussion (version 1). Fourth, HF asked professional bilingual translators who were not involved in our study to translate the Japanese text back into English. HF and KK then compared the back-translated and original English versions item by item and then revised the Japanese version (version 2). Fifth, an expert review of version 2 was conducted by an AI expert (YY) and a health profession education expert (YN) at HF's request. These experts concluded that no amendment was required. Sixth, HF contacted the creator of the original scale again and asked him to check version 2. The creator concluded that no revision was necessary. Seventh, pilot testing was performed with 2 medical trainees, which indicated no problematic items. Finally, version 2 was confirmed as the finalized Japanese version of the ATTARI-12 (J-ATTARI-12).

Step 2: Investigation of Psychometric Properties

Participants

Between June 2025 and July 2025, we recruited study participants from 5 universities and 9 hospitals across Japan, each of which varied in type and location (Multimedia Appendix 1). We asked the medical education directors of the universities and residency training directors of the hospitals to distribute our anonymous online self-administered questionnaire. An email with a link to the online questionnaire was sent to medical students at the universities and resident physicians at the hospitals via the respective directors. The participants were provided with a brief explanation of the study and indicated their consent to take part by checking the consent box. They were then able to access and complete the questionnaire. The survey duration was 1 month. To maximize the response rate, we sent reminders several times during the survey period.

Measures

The original version of the ATTARI-12 has 12 items [10], with responses on a 5-point Likert scale (1="strongly disagree"; 5="strongly agree") [10]. Items 2, 4, 7, 8, 10, and 12 are reverse

items and, therefore, are reverse scored [10]. The sum of all item scores is averaged to create a total score ranging from 1 to 5, with higher scores indicating more positive attitudes toward AI.

Statistical Analysis

To investigate the structural validity of the instrument, we performed exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Because conducting these 2 types of factor analysis on the same dataset would potentially raise concerns [16], we adopted the split-half validation approach, in which the participants were randomly divided into 2 groups, half (group A) for EFA and the other half (group B) for CFA. As mentioned above, we aimed to develop the scale in a manner that was culturally adapted to the Japanese medical education context. Consequently, EFA was conducted first, followed by CFA.

To determine the appropriateness of EFA, we checked the Kaiser-Meyer-Olkin (KMO) coefficient and Bartlett sphericity test. Running EFA requires a KMO value over 0.80 and a significant result in the Bartlett test [17,18]. We applied EFA to the responses of group A to explore the factor structure of the J-ATTARI-12 using maximum likelihood estimation and Promax rotation. We determined the final factor solution using the results of parallel analysis and the factor loading values (cutoff value=0.30).

We subsequently conducted CFA on group B to confirm the model obtained in EFA. CFA was performed using the maximum likelihood estimator method. Model fitness of CFA is commonly conducted using indexes, including the comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR), with higher CFI (>0.900), lower RMSEA (<0.080), and lower SRMR (<0.080) values indicating a good fit [19-21]. In this study, we compared these indexes between the 2 models, namely, the model suggested by the EFA (2-factor model; described in detail below) and the 1-factor model.

Convergent validity was evaluated through hypothesis testing. On the basis of a previous finding that ATTARI-12 scores were positively linked with specific attitudes toward robots [10], we investigated this relationship using Pearson correlation coefficients. With reference to previous studies [10,22], attitudes toward robots were assessed using 3 items (eg, “Robots are necessary as they can do jobs that are too hard or too dangerous for people”), each of which was answered on a 4-point Likert scale (Multimedia Appendix 2). The scores were computed by averaging the responses to the 3 items (Cronbach α =0.78; mean

3.25, SD 0.55 in our dataset), with higher scores indicating a more positive attitude toward robots. The Pearson correlation coefficients are deemed meaningful if they are greater than 0.30 [23].

We used Cronbach α values to evaluate the internal consistency reliability of the scale. In this study, we computed the values using the responses of the entire sample (N=326; described in detail below). Cronbach α values above 0.70 are acceptable [24].

The possible influence of participant gender and year group on ATTARI-12 scores was explored using descriptive statistics and comparison analyses (independent 2-tailed *t* test or 1-way ANOVA). With regard to gender, as outlined later, the “Others” category comprised only 4 participants, rendering it inadequate for inclusion in a 1-way ANOVA. Given the high likelihood of unstable estimates and heterogeneity of variance associated with such a small group, we compared scores between only the 2 larger groups (man and woman) via independent *t* test. The undergraduate medical curriculum in Japan has traditionally consisted of 2 phases: the first 4 years of preclinical education and the subsequent 2 years of clinical education (ie, clinical clerkship) [25]. Accordingly, medical trainees were divided into preclinical medical students (ie, first- to fourth-year students), clinical medical students (ie, fifth- and sixth-year students), and medical residents. All statistical analyses were performed using R (version 4.5.1; R Foundation for Statistical Computing) and SPSS (version 30.0; IBM Corp), with 2-sided *P* values of *<.05* considered statistically significant.

Ethical Considerations

This study was conducted according to the ethical standards and principles of the Declaration of Helsinki and was approved by the ethics committee of Juntendo University Faculty of Medicine (E25-0028). All participants checked the consent box at the beginning of the questionnaire to indicate their informed consent to take part in the study. To ensure confidentiality, all participant data were anonymized before analysis. Participants were entered into a draw for 1 of 10 ¥5000 (approximately US \$30) gift cards.

Results

Overview

In total, 9.2% (326/3551) of the eligible participants responded to the survey. There were no missing data. Table 1 shows the participants’ characteristics, and Table 2 shows the responses to each item.

Table . Characteristics of the participants (N=326).

Characteristic	Participants, n (%)
Gender	
Woman	142 (43.6)
Man	180 (55.2)
Nonbinary	4 (1.2)
Year group	
Medical students	
First	103 (31.6)
Second	19 (5.8)
Third	51 (15.6)
Fourth	60 (18.4)
Fifth	13 (4.0)
Sixth	20 (6.1)
Medical residents	
First	32 (9.8)
Second	28 (8.6)

Table . Responses to the 12-item Attitudes Towards Artificial Intelligence scale (N=326).

Item (as in the original English-language version)	Responses, n (%) ^a	1	2	3	4	5
Item 1: "AI will make this world a better place."	1 (0.3)	11 (3.4)	72 (22.1)	201 (61.7)	41 (12.6)	
Item 2: "I have strong negative emotions about AI." ^b	49 (15.0)	163 (50.0)	71 (21.8)	37 (11.3)	6 (1.8)	
Item 3: "I want to use technologies that rely on AI."	0 (0.0)	3 (0.9)	36 (11.0)	200 (61.3)	87 (26.7)	
Item 4: "AI has more disadvantages than advantages." ^b	28 (8.6)	171 (52.5)	101 (31.0)	22 (6.7)	4 (1.2)	
Item 5: "I look forward to future AI developments."	2 (0.6)	4 (1.2)	28 (8.6)	157 (48.2)	135 (41.4)	
Item 6: "AI offers solutions to many world problems."	2 (0.6)	24 (7.4)	73 (22.4)	175 (53.7)	52 (16.0)	
Item 7: "I prefer technologies that do not feature AI." ^b	19 (5.8)	124 (38.0)	120 (36.8)	51 (15.6)	12 (3.7)	
Item 8: "I am afraid of AI." ^b	17 (5.2)	62 (19.0)	77 (23.6)	147 (45.1)	23 (7.1)	
Item 9: "I would rather choose a technology with AI than one without it."	5 (1.5)	16 (4.9)	111 (34.0)	154 (47.2)	40 (12.3)	
Item 10: "AI creates problems rather than solving them." ^b	19 (5.8)	115 (35.3)	137 (42.0)	47 (14.4)	8 (2.5)	
Item 11: "When I think about AI, I have mostly positive feelings."	2 (0.6)	45 (13.8)	89 (27.3)	157 (48.2)	22 (6.7)	
Item 12: "I would rather avoid technologies that are based on AI." ^b	38 (11.7)	177 (54.3)	84 (25.8)	23 (7.1)	4 (1.2)	

^a5-point Likert scale (1="strongly disagree" to 5="strongly agree")^bThese are reverse-scored items.

Structural Validity Analysis

We performed EFA on group A using maximum likelihood estimation with Promax rotation (154/326, 47.2% of the participants). The KMO value was 0.83, and the Bartlett test

was significant ($P<.001$). **Table 3** shows the results of the EFA. EFA suggested a 2-factor solution. After discussion among the team members, we named the factors as follows: factor 1 was AI anxiety and aversion, and factor 2 was AI optimism and acceptance.

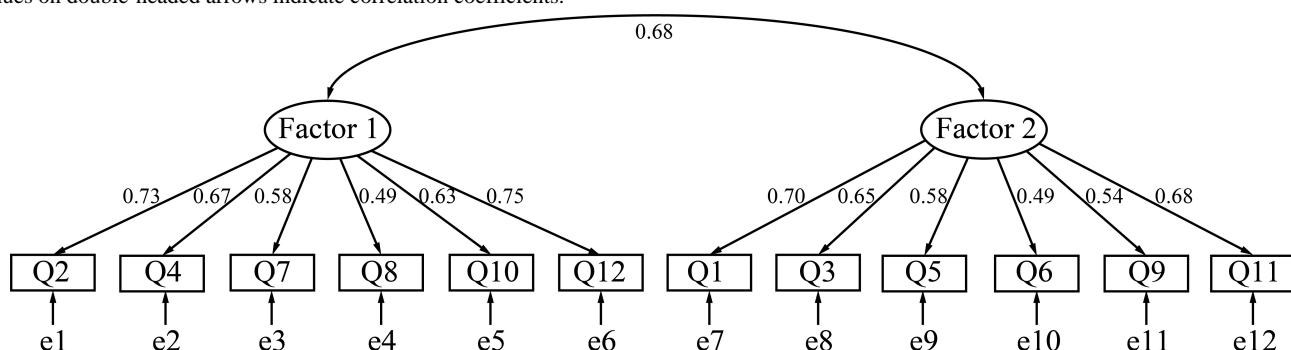
Table . Results of the exploratory factor analysis.

	Factor 1	Factor 2
Item 1 factor loading	-0.16	0.73
Item 2 factor loading	0.59	0.18
Item 3 factor loading	0.21	0.52
Item 4 factor loading	0.73	-0.07
Item 5 factor loading	-0.05	0.85
Item 6 factor loading	0.05	0.35
Item 7 factor loading	0.73	-0.06
Item 8 factor loading	0.30	0.12
Item 9 factor loading	-0.05	0.47
Item 10 factor loading	0.78	-0.15
Item 11 factor loading	0.12	0.44
Item 12 factor loading	0.64	0.07
Eigenvalue	3.79	0.76
Percentage of the variance explained	22	18

We then performed CFA on group B (the remaining 172/326, 52.8% of the participants) using the maximum likelihood estimator method. The 2-factor model suggested by the EFA yielded better goodness-of-fit results (CFI=0.914;

RMSEA=0.075; SRMR=0.056) than the 1-factor model (CFI=0.804; RMSEA=0.113; SRMR=0.078). [Multimedia Appendix 3](#) shows the CIs for the RMSEA. [Figure 1](#) shows the path diagram. Therefore, we adopted the 2-factor model.

Figure 1. Factor structure of the Japanese version of the 12-item Attitudes Towards Artificial Intelligence scale (confirmatory factor analysis). Ellipses indicate latent variables (factors). Rectangles indicate observed variables (items). Values on single-headed arrows indicate standardized factor loadings. Values on double-headed arrows indicate correlation coefficients.



Convergent Validity Analysis

We computed the Pearson correlation coefficient between the ATTARI-12 scores and the attitude toward robots scores. The coefficient value was 0.52 ($P<.001$), which indicated a positive correlation between these 2 scores.

Internal Consistency Reliability Analysis and Descriptive Statistics

[Table 4](#) shows the descriptive statistics with the Cronbach α values. The Cronbach α for all 12 items was 0.84. The Cronbach α values for factors 1 and 2 were 0.80 and 0.76, respectively. These values were above the cutoff (0.70).

Table . Internal consistency reliability analysis and descriptive statistics.

	Number of items	Mean (SD)	Observed range	Cronbach α
Total	12	3.61 (0.50)	2.08 - 5.00	0.84
Factor 1	6	3.36 (0.63)	1.00 - 5.00	0.80
Factor 2	6	3.87 (0.52)	2.17 - 5.00	0.76

[Table 5](#) shows descriptive data and comparison by gender and year group. The comparison analyses did not show any significant differences in ATTARI-12 scores.

Table . Descriptive data and comparison by gender and year group.

	Score on ATTARI-12 ^a , mean (SD)	P value
Gender ^b		.71 ^c
Woman	3.60 (0.47)	
Man	3.62 (0.53)	
Year group		.21 ^d
First- to fourth-year medical students	3.58 (0.51)	
Fifth- and sixth-year medical students	3.67 (0.55)	
Medical residents	3.70 (0.42)	

^aATTARI-12: 12-item Attitudes Towards Artificial Intelligence scale.

^bThe “Nonbinary” group included only 4 participants, which is too small to be appropriately included in a 1-way ANOVA. Given the high likelihood of unstable estimates and heterogeneity of variance associated with such a small group, we compared scores between only the 2 larger groups (man and woman).

^cP value from independent-sample *t* test.

^dP value from 1-way ANOVA.

Discussion

Principal Findings

In this study, we translated the ATTARI-12, originally developed for the general population, into Japanese in accordance with an international guideline [15] and then validated its structural and convergent validity and internal consistency reliability for medical trainees. Applying this scale in the context of medical trainees in Japan has the potential to stimulate future research and educational interventions. Such initiatives would serve to effectively assess and enhance the integration of AI in clinical practice and the medical education system.

Our study found that the internal consistency reliability of the J-ATTARI-12 was good. The findings were consistent with those of the original ATTARI-12 developmental study [10]. The original study reported Cronbach α values of 0.93 for a US sample and 0.90 for a German sample [10], which suggested that the ATTARI-12 is likely a helpful measure with good internal consistency reliability across countries.

Factor analysis indicated that the J-ATTARI-12 had a 2-factor structure, in contrast to the original English-language version’s unidimensional structure. The original ATTARI-12 was conceptualized as a unidimensional scale. The original article by Stein et al [10] suggested that the developers intentionally balanced positively and negatively worded items. This design makes it unsurprising that the Japanese version yielded 2 factors corresponding to negative (factor 1: AI anxiety and aversion) and positive (factor 2: AI optimism and acceptance) item valence. Therefore, we should acknowledge that, although 2 factors emerged, they simply reflect positive versus negative wording and that, for the purpose of international comparison, using a total score based on a 1-factor assumption may remain preferable. At the same time, the content of the 2 factors (ie, AI anxiety and aversion vs AI optimism and acceptance) may reflect attitudinal domains that are influenced by broader cultural characteristics, such as uncertainty avoidance, technophilia, or collectivism. To determine whether a 2-factor structure offers

conceptual or psychometric advantages beyond a unidimensional model and whether this pattern is observed across different cultural contexts, future cross-cultural validation studies with larger and more diverse samples are required.

We note a couple of potential limitations of our study. First, the response rate was relatively low. In addition, the survey was conducted using convenience sampling and included only 14 institutions, and it is likely that participating institutions or individuals had a greater interest in AI. It is increasingly challenging to obtain high response rates to online surveys, and rates frequently drop to 10% [26,27]. Nevertheless, the literature suggests that the response rate to our survey may have been sufficient to provide reliable data [28,29]. Additionally, it should be noted that, despite the absence of accurate statistics on the demographic variables of medical trainees across Japan, recent reports have indicated that the proportion of female physicians aged ≤ 29 years stands at approximately 40%, with a steady upward trend observed in recent years [30]. Thus, the finding that 43.6% (142/326) of the respondents were women does not generate concerns regarding the representativeness of the sample. Second, in this study, although we evaluated the structural validity, convergent validity, and internal consistency reliability of the scale, other determinants of validity (eg, discriminant validity and predictive validity) and reliability (eg, test-retest reliability) have yet to be examined. Future studies should examine these psychometric properties. Third, the use of a gift card lottery can raise concerns about bias. Nevertheless, this method is commonly used and appears to be acceptable [31-34]. Fourth, our results may indicate a ceiling effect. For example, 41.4% (135/326) of the respondents selected the highest score (5) for item 5. However, according to the original ATTARI-12 paper, responses tend to be skewed toward a score of 4, suggesting that the scale inherently elicits generally positive attitudes toward AI. Therefore, these findings do not represent an unusual deviation but are consistent with the response patterns reported in the original study.

Despite these limitations, this study produced the first Japanese version of the ATTARI-12, which is likely to be used as a novel

measure for assessing AI attitudes among medical students and residents. Two outcomes are expected. First, by incorporating the J-ATTARI-12 into medical curricula, medical educators will be able to develop a more comprehensive understanding of the trainees' readiness for AI adoption. This will likely be advantageous in the design of customized educational interventions using AI based on the level of each trainee's attitude toward AI [35] and will support curriculum development by enabling medical educators to identify learners who may need more foundational AI exposure or targeted support. Second, the J-ATTARI-12 will facilitate medical education research in Japan. The scale allows researchers to examine the impact of specific educational interventions on attitude change concerning AI, and it is suitable for repeated administration in longitudinal studies to track how attitudes evolve over the course of training. In addition, the availability of the J-ATTARI-12 will promote cross-cultural comparison of attitudes toward AI

and its possible outcomes internationally. For these reasons, we expect that the J-ATTARI-12 will play a pivotal role in facilitating data-informed curriculum development and contribute to the expanding body of medical education research in this era of AI.

Conclusions

In this study, we developed the J-ATTARI-12, originally developed for the general population, in accordance with an international guideline. A validation survey revealed that the structural and convergent validity, as well as the internal consistency reliability, were good for medical trainees in Japan. The developed measure can be used for customized educational initiatives using AI based on the level of each trainee's attitude toward AI. It will also provide helpful information to medical education researchers in Japan in this era of AI.

Acknowledgments

The authors would like to thank all the participants. They would also like to thank ChatGPT-5.1 by OpenAI for its valuable assistance in refining their academic writing.

Funding

This work was funded by the "Creating training hubs for advanced medical personnel (supporting the fostering of doctors with advanced clinical and research capabilities)" project and was supported by the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

HF conceived the study with input from HM, KK, YN, YY, and TN. HF conducted the data analysis. HF wrote the first draft of the manuscript. All authors reviewed, discussed, proofread, and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Characteristics of the participating institutions.

[[DOCX File, 18 KB - mededu_v12i1e81986_app1.docx](#)]

Multimedia Appendix 2

Measure for attitudes toward robots.

[[DOCX File, 18 KB - mededu_v12i1e81986_app2.docx](#)]

Multimedia Appendix 3

CIs for the root mean square error of approximation.

[[DOCX File, 18 KB - mededu_v12i1e81986_app3.docx](#)]

References

1. Buch VH, Ahmed I, Maruthappu M. Artificial intelligence in medicine: current trends and future possibilities. *Br J Gen Pract* 2018 Mar;68(668):143-144. [doi: [10.3399/bjgp18X695213](https://doi.org/10.3399/bjgp18X695213)] [Medline: [29472224](https://pubmed.ncbi.nlm.nih.gov/29472224/)]
2. Wu K, Wu E, Theodorou B, et al. Characterizing the clinical adoption of medical AI devices through U.S. insurance claims. *NEJM AI* 2024 Nov 9;1(1). [doi: [10.1056/AIoa2300030](https://doi.org/10.1056/AIoa2300030)]

3. Habli I, Lawton T, Porter Z. Artificial intelligence in health care: accountability and safety. *Bull World Health Organ* 2020 Feb 25;98(4):251-256. [doi: [10.2471/BLT.19.237487](https://doi.org/10.2471/BLT.19.237487)] [Medline: [32284648](https://pubmed.ncbi.nlm.nih.gov/32284648/)]
4. Bavli I, Ho A, Mahal R, McKeown MJ. Ethical concerns around privacy and data security in AI health monitoring for Parkinson's disease: insights from patients, family members, and healthcare professionals. *AI Soc* 2024 Jan 23;40:155-165. [doi: [10.1007/s00146-023-01843-6](https://doi.org/10.1007/s00146-023-01843-6)]
5. Katirai A. The ethics of advancing artificial intelligence in healthcare: analyzing ethical considerations for Japan's innovative AI hospital system. *Front Public Health* 2023 Jul 17;11:1142062. [doi: [10.3389/fpubh.2023.1142062](https://doi.org/10.3389/fpubh.2023.1142062)] [Medline: [37529426](https://pubmed.ncbi.nlm.nih.gov/37529426/)]
6. Khan B, Fatima H, Qureshi A, et al. Drawbacks of artificial intelligence and their potential solutions in the healthcare sector. *Biomed Mater Devices* 2023 Feb 8:1-8. [doi: [10.1007/s44174-023-00063-2](https://doi.org/10.1007/s44174-023-00063-2)] [Medline: [36785697](https://pubmed.ncbi.nlm.nih.gov/36785697/)]
7. Andrade-Arenas L, Yactayo-Arias C. A bibliometric analysis of the advance of artificial intelligence in medicine. *Int J Electr Comput Eng* 2024;14(3):3350. [doi: [10.11591/ijece.v14i3.pp3350-3361](https://doi.org/10.11591/ijece.v14i3.pp3350-3361)]
8. Dave M, Patel N. Artificial intelligence in healthcare and education. *Br Dent J* 2023 May 26;234:761-764. [doi: [10.1038/s41415-023-5845-2](https://doi.org/10.1038/s41415-023-5845-2)]
9. Mir MM, Mir GM, Raina NT, et al. Application of artificial intelligence in medical education: current scenario and future perspectives. *J Adv Med Educ Prof* 2023 Jul;11(3):133-140. [doi: [10.30476/JAMP.2023.98655.1803](https://doi.org/10.30476/JAMP.2023.98655.1803)] [Medline: [37469385](https://pubmed.ncbi.nlm.nih.gov/37469385/)]
10. Stein JP, Messingschläger T, Gnambs T, Hutmacher F, Appel M. Attitudes towards AI: measurement and associations with personality. *Sci Rep* 2024 Feb 5;14(1):2909. [doi: [10.1038/s41598-024-53335-2](https://doi.org/10.1038/s41598-024-53335-2)] [Medline: [38316898](https://pubmed.ncbi.nlm.nih.gov/38316898/)]
11. Horowitz MC, Kahn L. Leading in artificial intelligence through confidence building measures. *Wash Q* 2021 Oct 2;44(4):91-106. [doi: [10.1080/0163660X.2021.2018794](https://doi.org/10.1080/0163660X.2021.2018794)]
12. Naiseh M, Babiker A, Al-Shakhs S, et al. Attitudes towards AI: the interplay of self-efficacy, well-being, and competency. *J Technol Behav Sci* 2025. [doi: [10.1007/s41347-025-00486-2](https://doi.org/10.1007/s41347-025-00486-2)]
13. Sindermann C, Sha P, Zhou M, et al. Assessing the attitude towards artificial intelligence: introduction of a short measure in German, Chinese, and English language. *Künstl Intell* 2020;35:109-118. [doi: [10.1007/s13218-020-00689-0](https://doi.org/10.1007/s13218-020-00689-0)]
14. Hofstede GJ. Adoption of communication technologies and national culture. *Syst Inf Manag* 2001;6(3):55-69 [[FREE Full text](#)]
15. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)* 2000 Dec 15;25(24):3186-3191. [doi: [10.1097/00007632-200012150-00014](https://doi.org/10.1097/00007632-200012150-00014)] [Medline: [11124735](https://pubmed.ncbi.nlm.nih.gov/11124735/)]
16. Lorenzo-Seva U. SOLOMON: a method for splitting a sample into equivalent subsamples in factor analysis. *Behav Res Methods* 2022 Dec;54(6):2665-2677. [doi: [10.3758/s13428-021-01750-y](https://doi.org/10.3758/s13428-021-01750-y)] [Medline: [34918226](https://pubmed.ncbi.nlm.nih.gov/34918226/)]
17. Kaiser HF. An index of factorial simplicity. *Psychometrika* 1974;39(1):31-36. [doi: [10.1007/BF02291575](https://doi.org/10.1007/BF02291575)]
18. Karimian Z, Chahartangi F. Development and validation of a questionnaire to measure educational agility: a psychometric assessment using exploratory factor analysis. *BMC Med Educ* 2024 Nov 9;24:1284. [doi: [10.1186/s12909-024-06307-z](https://doi.org/10.1186/s12909-024-06307-z)]
19. Brown TA. *Confirmatory Factor Analysis for Applied Research*: The Guilford Press; 2015.
20. Tabachnick BG, Fidell LS. *Using Multivariate Statistics*: Pearson Education; 2013.
21. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organ Res Methods* 2000;3(1):4-70. [doi: [10.1177/109442810031002](https://doi.org/10.1177/109442810031002)]
22. Gnambs T, Appel M. Are robots becoming unpopular? Changes in attitudes towards autonomous robotic systems in Europe. *Comput Human Behav* 2019 Apr;93:53-61. [doi: [10.1016/j.chb.2018.11.045](https://doi.org/10.1016/j.chb.2018.11.045)]
23. Cohen J. Statistical power analysis. *Curr Dir Psychol Sci* 1992;1(3):98-101. [doi: [10.1111/1467-8721.ep10768783](https://doi.org/10.1111/1467-8721.ep10768783)]
24. DeVellis RF. *Scale Development: Theory and Applications*: Sage Publications; 2003.
25. Nishigori H. Medical education in Japan. *Med Teach* 2024 Sep;46(sup1):S4-S10. [doi: [10.1080/0142159X.2024.2372108](https://doi.org/10.1080/0142159X.2024.2372108)] [Medline: [39545499](https://pubmed.ncbi.nlm.nih.gov/39545499/)]
26. 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](https://doi.org/10.1016/j.chb.2009.10.015)]
27. Van Mol C. Improving web survey efficiency: the impact of an extra reminder and reminder content on web survey response. *Int J Soc Res Methodol* 2017;20(4):317-327. [doi: [10.1080/13645579.2016.1185255](https://doi.org/10.1080/13645579.2016.1185255)]
28. Fosnacht K, Sarraf SA, Howe E, Peck LK. How important are high response rates for college surveys? *Rev High Educ* 2017;40(2):245-265. [doi: [10.1353/rhe.2017.0003](https://doi.org/10.1353/rhe.2017.0003)]
29. Wu MJ, Zhao K, Fils-Aime F. Response rates of online surveys in published research: a meta-analysis. *Comput Hum Behav Rep* 2022 Aug;7:100206. [doi: [10.1016/j.chbr.2022.100206](https://doi.org/10.1016/j.chbr.2022.100206)]
30. Overview of statistics on doctors, dentists, and pharmacists in 2022 [Web page in Japanese]. Ministry of Health, Labor and Welfare, Japan. 2022. URL: <https://www.mhlw.go.jp/toukei/saikin/hw/ishi/22/index.html> [accessed 2025-12-10]
31. Fujikawa H, Ando T, Kondo K, Hayashi M. Development and psychometric properties of the Japanese version of the capacity for wonder scale. *Int J Med Educ* 2025 May 29;16:100-106. [doi: [10.5116/ijme.6819.fcae](https://doi.org/10.5116/ijme.6819.fcae)] [Medline: [40449508](https://pubmed.ncbi.nlm.nih.gov/40449508/)]
32. Fujikawa H, Aoki T, Eto M. Associations between workplace social capital, well-being, and work engagement in medical residents: a multicenter cross-sectional study. *BMC Med Educ* 2024 Sep 28;24(1):1063. [doi: [10.1186/s12909-024-06055-0](https://doi.org/10.1186/s12909-024-06055-0)] [Medline: [39342158](https://pubmed.ncbi.nlm.nih.gov/39342158/)]

33. Fujikawa H, Hayashi M, Son D, Kondo K, Eto M. Translating, adapting, and validating the medical student version of the patient care ownership scale for use in Japan. *BMC Med Educ* 2024 Jun 28;24(1):706. [doi: [10.1186/s12909-024-05704-8](https://doi.org/10.1186/s12909-024-05704-8)] [Medline: [38943114](#)]
34. Shinohara K, Aoki T, So R, et al. Influence of overstated abstract conclusions on clinicians: a web-based randomised controlled trial. *BMJ Open* 2017 Dec 14;7(12):e018355. [doi: [10.1136/bmjopen-2017-018355](https://doi.org/10.1136/bmjopen-2017-018355)] [Medline: [29247102](#)]
35. Sami A, Tanveer F, Sajwani K, et al. Medical students' attitudes toward AI in education: perception, effectiveness, and its credibility. *BMC Med Educ* 2025 Jan 17;25(1):82. [doi: [10.1186/s12909-025-06704-y](https://doi.org/10.1186/s12909-025-06704-y)] [Medline: [39833834](#)]

Abbreviations

AI: artificial intelligence

ATTARI-12: 12-item Attitudes Towards Artificial Intelligence scale

CFA: confirmatory factor analysis

CFI: comparative fit index

EFA: exploratory factor analysis

J-ATTARI-12: Japanese version of the 12-item Attitudes Towards Artificial Intelligence scale

KMO: Kaiser-Meyer-Olkin

RMSEA: root mean square error of approximation

SRMR: standardized root mean square residual

Edited by Z Khlaif; submitted 07.Aug.2025; peer-reviewed by T Iwayama, T Yusuff; revised version received 11.Dec.2025; accepted 14.Dec.2025; published 14.Jan.2026.

Please cite as:

Fujikawa H, Mori H, Kondo K, Nishizaki Y, Yano Y, Naito T

Adaptation of the Japanese Version of the 12-Item Attitudes Towards Artificial Intelligence Scale for Medical Trainees: Multicenter Development and Validation Study

JMIR Med Educ 2026;12:e81986

URL: <https://mededu.jmir.org/2026/1/e81986>

doi: [10.2196/81986](https://doi.org/10.2196/81986)

© Hirohisa Fujikawa, Hirotake Mori, Kayo Kondo, Yuji Nishizaki, Yuichiro Yano, Toshio Naito. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 14.Jan.2026. 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.

Original Paper

Evaluation of a Problem-Based Learning Program's Effect on Artificial Intelligence Ethics Among Japanese Medical Students: Mixed Methods Study

Yuma Ota¹, MSN, RN; Yoshikazu Asada², PhD; Saori Kubo³, PhD; Takeshi Kanno^{4,5}, MD, PhD; Machiko Saeki Yagi², RN, CHSE, PhD; Yasushi Matsuyama⁶, MD, MHPE, PhD

¹Medical Education Center, Jichi Medical University Graduate School, Shimotsuke, Japan

²Medical Education Center, Jichi Medical University, Shimotsuke, Japan

³Graduate School of Education, Tohoku University, Sendai, Japan

⁴R & D Division of Career Education for Medical Professionals, Medical Education Center, Jichi Medical University, Shimotsuke, Japan

⁵Division of Gastroenterology, Graduate School of Medicine, Tohoku University, Sendai, Japan

⁶Medical Education Center, Jichi Medical University, Tochigi, Japan

Corresponding Author:

Yasushi Matsuyama, MD, MHPE, PhD

Medical Education Center

Jichi Medical University

3311-1 Yakushiji, Shimotsuke

Tochigi, 329-0498

Japan

Phone: 81 285 58 7067

Email: yasushim@jichi.ac.jp

Abstract

Background: The rapid advancement of artificial intelligence (AI) has had a substantial impact on medicine, necessitating the integration of AI education into medical school curricula. However, such integration remains limited. A key challenge is the discrepancy between medical students' positive perceptions of AI and their actual competencies, with research in Japan identifying specific gaps in the students' competencies in understanding regulations and discussing ethical issues.

Objective: This study evaluates the effectiveness of an educational program designed to improve medical students' competencies in understanding legal and ethical AI-related issues. It addresses the following research questions: (1) Does this educational program improve students' knowledge of AI and its legal and ethical issues, and what is each program element's contribution to this knowledge? (2) How does this educational program qualitatively change medical students' thoughts on these issues from an abstract understanding to a concrete and structured thought process?

Methods: This mixed methods study used a single-group pretest and posttest framework involving 118 fourth-year medical students. The 1-day intervention comprised a lecture and problem-based learning (PBL) session centered on a clinical case. A 24-item multiple-choice questionnaire (MCQ) was administered at 3 time points (pretest, midtest, and posttest), and descriptive essays were collected before and after the intervention. Data were analyzed using linear mixed-effects models, the Wilcoxon signed-rank test, and text mining, including comparative frequency analysis and cooccurrence network analysis with Jaccard coefficients. An optional survey on student perceptions based on the attention, relevance, confidence, and satisfaction model was conducted (n=76, 64.4%).

Results: Objective knowledge scores increased significantly from the pretest (median 17, IQR 15-18) to posttest (median 19, IQR 17-21; $\beta=1.42$; $P<.001$). No significant difference was observed between score gains during the lecture and PBL phases ($P=.54$). Qualitative text analysis revealed the significant transformation of cooccurrence network structures (Jaccard coefficients 0.116 and 0.121) from fragmented clusters to integrated networks. Students also used professional and ethical terminology more frequently. For instance, use of the term "bias" in patient explanations increased from 10 (8.5%) at pretest to 25 (21.2%) at posttest, while references to "personal information" in physician precautions increased from 36 (30.5%) to 50 (42.4%). The optional survey indicated that students' confidence (mean 3.78, SD 0.87) was significantly lower than their perception of the program's relevance (mean 4.20, SD 0.71; $P<.001$).

Conclusions: This PBL-based program was associated with the improvements in knowledge and, more importantly, a structural transformation in students' thinking about AI ethics from an abstract level to a concrete, clinically grounded reasoning. The discrepancy between quantitative and qualitative findings suggests limitations of MCQs in assessing higher-order skills fostered by PBL. Overall, this study indicates the potential of PBL as an effective pedagogical method for AI ethics education.

(*JMIR Med Educ* 2026;12:e84535) doi:10.2196/84535

KEYWORDS

artificial intelligence; undergraduate medical education; medical ethics; problem-based learning; mixed methods research; text mining

Introduction

The rapid advancement of artificial intelligence (AI) has significantly affected the medical domain, which uses applications ranging from diagnostic support to treatment selection [1-3]. This necessitates the integration of AI education into medical school curricula to ensure that future physicians have the necessary competencies to use these technologies effectively and ethically [4-6]. Both educators and students recognize this necessity, and many studies support the implementation of systematic AI education [7,8]. However, despite the establishment of international recommendations, the integration of AI education into medical school curricula remains limited [3,6]. One of the key challenges in AI education identified by earlier research is the discrepancy between medical students' perceptions of and competencies in AI [6,7,9]. Recent systematic reviews consistently indicate that despite generally maintaining a positive attitude toward AI's health care potential, students have a moderate to low level of objective knowledge and practical skills [7]. This gap between the positive perceptions and insufficient competency acquisition of students indicates the need to integrate systematic and effective educational programs into the medical curriculum [9-11].

Furthermore, AI education poses not only technical challenges but also legal and ethical ones [1,2,12]. The use of AI can give rise to new legal and ethical issues, such as the risk of algorithmic bias resulting in patient discrimination, challenges related to patient privacy, and problems associated with accountability for AI-driven decisions [2,3,13]. Further, expert consensus studies indicate that nontechnical competencies such as ethics, law, communication, and collaboration are considered more important than advanced technical skills such as programming for physicians [8,11]. However, practical, case-based ethics education and standardized frameworks to teach these important topics are currently lacking [3,12,13].

This global context is highly relevant in Japan. In response to these trends, Japan's Ministry of Education, Culture, Sports, Science, and Technology revised the national Model Core Curriculum for medical education in 2022, establishing "Competency in Utilizing Information Science and Technology" as a core competency for all future physicians [14,15]. This necessitated an investigation of the specific educational needs of Japanese medical students. A recent national cross-sectional study of sixth-year medical students in Japan clarified the acquisition status of specific competencies [16]. The study revealed particularly low self-assessment in three key learning objectives: (1) understanding the regulations, laws, and

guidelines for information science and technology (IST) adoption in medicine; (2) discussing ethical issues, such as social disparities; and (3) understanding and discussing the potential applications of specific technologies, including AI, in medical care [16]. These results highlight the necessity of implementing focused educational interventions to address these gaps among Japanese medical students. To address these identified gaps, we chose a problem-based learning (PBL) framework for our educational intervention. This decision was directly informed by the findings of the aforementioned national study [16], which not only highlighted these specific areas of low confidence but also revealed a significant positive association between prior PBL experience and students' self-assessed competence in these very domains. This suggests that the PBL methodology—which promotes active, case-based, and collaborative problem-solving—is a particularly effective approach for navigating the complex, nontechnical challenges of AI ethics and law.

Accordingly, this study aims to evaluate the effectiveness of an educational program designed to improve medical students' competencies in addressing the legal and ethical issues associated with AI use. To this end, we developed and implemented a PBL program centered on a clinical case requiring fourth-year medical students to navigate AI-related ethical dilemmas. The use of the term "artificial intelligence" by Japan's Model Core Curriculum highlights AI's development period [14,15]. However, the educational program developed in this study considers recent technological advancements and the societal landscape to primarily address the ethical challenges posed by generative AI. Therefore, although this study uses the term AI in a broader sense in alignment with earlier research and the core curriculum, its content and examples mainly focus on generative AI. Finally, we address the following research questions (RQs):

RQ1: Does this educational program improve students' knowledge of AI and its legal and ethical issues? Furthermore, which program elements improve knowledge, and how?

RQ2: How does the educational program qualitatively change medical students' thoughts on the legal and ethical issues of AI? In particular, does their thinking shift from an abstract understanding to a more concrete and structured thought process that is based on the clinical context?

Methods

Study Design and Participants

This study adopted a mixed methods design using a single-group pretest and posttest framework to evaluate an educational intervention. Specifically, we used a convergent approach integrating quantitative assessment (multiple-choice questionnaire [MCQ] scores) and quantitative text analysis of qualitative data (student essays). The participants were fourth-year medical students (n=124) at Jichi Medical University who undertook the intervention program as a mandatory part of their curriculum during the 2025 academic year. The program was conducted immediately before the students commenced their full-time clinical clerkship; this time was selected because they were expected to soon encounter ethical and legal issues related to AI's adoption in clinical practice.

Educational Intervention

The intervention in this study was a 1-day (approximately 7 hours) program based on the PBL model [17]. The program's learning objectives were selected such that they address the 3 key competency gaps (understanding regulations, discussing ethical issues, and understanding AI applications) that had low self-assessment scores among Japanese medical students in an earlier national cross-sectional study. The program required students to achieve the following three competencies based on a rubric proposed by earlier research [18]:

1. For standard clinical settings and cases, be able to take action based on regulations (laws, guidelines, etc) associated with IST's use in medicine at a certain rate.
2. Be able to exemplify and discuss the ethical issues, such as the social disparities caused by the digital divide, potentially arising from the use of digital information and scientific technology in medicine and welfare.
3. Be able to understand the IST (AI) related to medicine and discuss it with the support of practitioners.

Table 1. Overview of the program components and schedule.

Content	Time	Activity format
Introduction: presentation of learning objectives and case	15 minutes	Individual
Assessment 1 (pretest): MCQ ^a and descriptive questions	35 minutes	Individual
Lecture 1: foundational knowledge of AI ^b	35 minutes	Individual
Assessment 2 (midtest): MCQ	15 minutes	Individual
PBL ^c session 1: confirmation of schedule and learning process	5 minutes	Group
PBL session 2: problem clarification	20 minutes	Group
PBL session 3: self-learning (including lunch break)	120 minutes	Individual
PBL session: problem-solving group work	70 minutes	Group
Lecture 2: case review and debriefing	40 minutes	Individual
Assessment 3 (posttest): MCQ and descriptive questions	35 minutes	Individual

^aMCQ: multiple-choice questionnaire.

^bAI: artificial intelligence.

^cPBL: problem-based learning.

The PBL process proceeded in the following stages: (1) PBL learning process confirmation, (2) problem clarification, (3) self-directed learning (including a lunch break), and (4) a 70-minute problem-solving group work session. PBL was conducted in 16 small groups of 7-8 students each. The session comprised problem clarification, self-directed learning, and problem-solving phases and included a final 70-minute group work session where students used shared Google slides to collaboratively devise an explanation for the patient and his son. Five tutors facilitated the process using a standardized tutor guide and, to enhance discussion quality, key domestic AI guidelines [19,20] were distributed as group work materials.

At the end of the program, in lecture 2, an instructor reviewed the case to consolidate important concepts regarding AI's legal and ethical issues. Immediately thereafter, the final assessment (posttest: an MCQ and descriptive questions) was administered. A detailed description of the PBL program's development and theoretical underpinnings has been published [21].

Data Collection

To evaluate the program's effectiveness, we collected both quantitative and qualitative data using the university's learning management system.

Using the MCQ

A 24-item MCQ, with its content aligned with lecture 1 materials, was used to assess students' knowledge. Each item scored 1 point for a correct answer, the total possible score being 24 points. The questionnaire comprised 8 subscales, each with 3 items, covering the mechanisms of AI, the strengths of AI, machine learning, deep learning, overfitting, prompts, personal information protection, and challenges associated with AI use. To assess knowledge acquisition over time, the same set of questions was administered at 3 time points: before the intervention (pretest), after the lecture (midtest), and after the PBL program's completion (posttest). To minimize recall bias, the order of the questions and their options was randomized at each administration.

Descriptive Essay Questions

Two descriptive essay questions were administered before the intervention (pretest) and after the program (posttest): "A patient literally believes the results diagnosed by a personal AI consultation tool. As a medical student, how will you explain the dangers of believing the results diagnosed by AI?" (question 1) and "When using AI as a physician, what points do you think require caution?" (question 2). These questions were designed to evaluate changes in students' patient communication strategies and their understanding of the professional and ethical responsibilities associated with AI use.

Program Perception Survey

After the program, an optional survey was administered to assess students' perceptions of the program based on the attention, relevance, confidence, and satisfaction (ARCS) model [22]. Students rated their agreement with statements for each component on a 5-point Likert scale (ranging from 1=strongly disagree to 5=strongly agree).

Data Analysis

All statistical analyses were performed using R software (version 4.4.3; R Foundation for Statistical Computing) [23].

MCQ Data

The normality of data distribution for score gains was assessed using the Shapiro-Wilk test to determine the appropriate statistical test to compare educational phases. To assess the overall trajectory of the increase in score across the time points pretest, midtest, and posttest, a linear mixed-effects model was used. To compare the score gains between the lecture phase (midtest-pretest) and the PBL phase (posttest-midtest), a paired-samples 2 tailed *t* test was performed. Any violation of the assumption of normality was prespecified to result in the use of the nonparametric Wilcoxon signed-rank test as an alternative.

Descriptive Essay Data

To quantitatively and objectively analyze the content of free-text data, a quantitative text analysis (text mining) approach using several R packages was adopted. The analysis targeted the free-text responses to 2 questions at 2 time points: question 1 (patient explanation) at pretest and posttest and question 2 (physician precautions) at pretest and posttest. Further, data handling, aggregation, and visualization were performed using the *dplyr*, *tidyverse*, *tidytext*, *widyr*, *igraph*, and *ggraph* R packages [24].

Morphological Analysis for Word Tokenization and Normalization

To segment Japanese free-text into analyzable word units, morphological analysis was performed using the *RMecab* package (version 1.1.4) [25], which interfaces with the open-source engine MeCab, using the standard Information-technology Promotion Agency dictionary (ipadic). The analysis was limited to content words (nouns, verbs, and adjectives), with verbs and adjectives being normalized to nouns. Nouns alone are typically used as morphemes in network analysis [26].

Word Frequency Analysis

Extracted and normalized words were used to calculate the word frequencies for the responses to each question at each time point (question 1 at pretest, question 1 at posttest, question 2 at pretest, and question 2 at posttest). To assess the breadth of concept adoption across the cohort and mitigate the impact of outliers (eg, a single student repeating the same word), we treated word usage as a binary variable (presence or absence) for each student. Therefore, instead of raw counts, we calculated the number of students who used each word at least once to identify the most frequently used terms.

Comparative Word Frequency and Transition Analysis

To capture changes in descriptive content, the number of students using specific words was compared among time points for each question (pretest vs posttest for question 1 and pretest vs posttest for question 2). In addition to calculating the net difference in prevalence, we performed a transition analysis to examine individual-level changes. Specifically, we identified the number of students who newly adopted a word (nonuse at

pretest to use at posttest) and those who dropped a word (use at pretest to nonuse at posttest). This allowed us to distinguish whether the observed changes in prevalence were driven by the acquisition of new concepts or the abandonment of previous ones.

Cooccurrence Network Analysis

To structurally clarify the contextual relationships among words irrespective of their frequencies, a quantitative cooccurrence network analysis was performed. In this study, cooccurrence involves capturing the semantic proximity of a pair of words appearing adjacently within the same sentence. Once text was segmented into sentence units, adjacent word pairs were extracted. Similar to the word frequency analysis, we calculated the number of students who used each word pair (binary counting) to avoid overrepresentation by single individuals, using the *widyr* package. Subsequently, a network diagram was created using the words mentioned by 5 or more students as nodes and their cooccurrence relationships as edges, with edge thickness indicating the relationship's strength.

The Jaccard coefficient was used as a similarity metric to quantify the structural changes in these networks. The Jaccard coefficient measures the similarity between 2 sets by taking their intersection and dividing it by their union [26]. In this study, the coefficient was applied to compare entire network structures and was defined as the number of shared cooccurrence pairs (edges) divided by the total number of unique pairs across both pretest and posttest networks. We applied this metric because it assesses the number of common cooccurrence pairs without being affected by each pair's frequency. Accordingly, we could determine the extent to which the network structure changed over time, which indicated a shift in conceptual relationships, rather than a simple change in discourse volume. We identified the numbers of common pairs (present at both time points) and unique pairs (present at only one time point), as well.

Program Perception Survey Data

For the optional ARCS survey data, the normality of the data distribution was assessed using the Shapiro-Wilk test. We elected to use parametric statistics by prioritizing the robustness

of these tests given the adequate sample size, which mitigates the impact of potential deviations from normality. Descriptive statistics (mean and SD) were calculated for each of the 4 components. To clarify the differences between components, repeated-measures ANOVA was conducted. Further, pairwise *t* tests with a Bonferroni correction were used for post hoc analysis.

Ethical Considerations

The study's description stated that participation was voluntary and would not affect students' grades. Only those medical students who agreed to participate accessed the questionnaire. The study was approved by the Ethics Review Committee of Jichi Medical University School of Medicine, Shimotsuke, Japan (approval number 24-137).

Informed consent was obtained electronically; students agreed to participate by accessing the questionnaire after reading the study description. To ensure privacy and confidentiality, all data were anonymized and stored securely with restricted access. No compensation was provided for participation.

Results

Participant Characteristics

In this study, a total of 118 fourth-year medical students participated. All participants provided complete data for both the MCQ and descriptive essay questions, which were the main aspects analyzed in this study. The participants were 71 (60.2%) male and 47 (39.8%) female students, all of whom were younger than 30 years.

MCQ Scores: Descriptive Statistics and Overall Effectiveness

The MCQ's internal consistency and reliability were assessed using the Cronbach α value at each time point. The α coefficients for the total score were 0.651 at pretest, 0.765 at midtest, and 0.782 at posttest. The α coefficients for the 8 subscales ranged from 0.631 to 0.760, indicating acceptable to good reliability. Table 2 depicts the detailed Cronbach α values for the total score and all subscales.

Table 2. Cronbach α coefficients for the multiple-choice questionnaire at each time point.

Factor	Pretest	Midtest	Posttest
Strengths of AI ^a	0.631	0.662	0.692
AI architecture	0.639	0.744	0.724
AI output issues	0.743	0.732	0.754
Prompts	0.657	0.707	0.696
Personal data leakage	0.76	0.708	0.77
Machine learning	0.666	0.751	0.726
Deep learning	0.704	0.729	0.757
Overfitting	0.716	0.733	0.743
Total score	0.651	0.765	0.782

^aAI: artificial intelligence.

As per the Shapiro-Wilk test, the MCQ scores at none of the 3 time points were normally distributed ($P<.001$ for all values). Therefore, descriptive statistics were presented as median (IQR)

values. The median total MCQ score increased across the time points, from 17 (IQR 15-18) at pretest to 18 (IQR 16-20) at midtest and 19 (IQR 17-21) at posttest (Table 3).

Table 3. Median (IQR) scores of the multiple-choice questionnaire (MCQ) at each time point. The total possible score range for the MCQ is 0-24.

Factor	Pretest, median (IQR)	Midtest, median (IQR)	Posttest, median (IQR)
Strengths of AI ^a	2 (2-3)	2 (2-3)	2 (2-3)
Mechanisms of AI	2 (2-3)	2 (2-3)	3 (2-3)
Challenges associated with AI use	2 (2-3)	2 (2-3)	2 (2-3)
Prompts	2 (1-2)	2 (1-3)	2 (1-3)
Personal information protection	3 (3-3)	3 (3-3)	3 (3-3)
Machine learning	2 (1-2)	2 (1-3)	2 (1-3)
Deep learning	1 (1-2)	2 (1-3)	2 (1-3)
Overfitting	2 (1-3)	2 (1-3)	2 (1-3)
Total score	17 (15-18)	18 (16-20)	19 (17-21)

^aAI: artificial intelligence.

A linear mixed-effects model was used to assess the trajectory of this increase in scores. Results (Table 4) highlight the stepwise effect of the program. For instance, a statistically significant increase of 0.75 points from the baseline score at

pretest was noted after the lecture phase at midtest ($\beta=.75$; $P=.01$). Furthermore, the end-of-program (posttest) score significantly exceeded the baseline score by 1.42 points ($\beta=1.42$; $P<.001$).

Table 4. Fixed effects from linear mixed - effects model predicting total score. Model: Score ~ Time + (1 | ID); n=354 observations from 118 students.

Effect	Estimate β (SE)	t test (df)	P value
Intercept (pretest)	16.86 (0.34)	49.55 (179)	<.001
T2 (midtest)	0.75 (0.27)	2.79 (234)	.006
T3 (posttest)	1.42 (0.27)	5.30 (234)	<.001

To address the quantitative RQ1 regarding the PBL phase's additional effect, the Wilcoxon signed-rank test was performed to compare score gains. The test found no statistically significant difference between the score gain during the lecture phase

(midtest-Pretest) and that during the PBL phase (posttest-midtest) for either the total score ($V=3084.5$; $P=.54$) or any of the 8 subscales (Table 5).

Table 5. Wilcoxon signed-rank test comparing score gains between lecture (midtest-pretest) and PBL (posttest-midtest) phases.

Factor	V statistic	P value
Strengths of AI ^a	203.0	.058
Mechanisms of AI	1733.5	.21
Challenges associated with AI use	1347.0	.82
Prompts	1601.5	.76
Personal information protection	1044.0	.55
Machine learning	981.0	.61
Deep learning	120.0	.31
Overfitting	782.0	.90
Total score	3084.5	.54

^aAI: artificial intelligence.

Qualitative RQ

To address RQ2, free-text responses to question 1 (patient explanation) and question 2 (physician precautions) were

analyzed by comparing word frequencies and cooccurrence network structures between pretest and posttests.

Responses to Question 1

After the intervention, a significant increase in the number of students using words related to AI mechanisms, limitations, and applications was noted. In particular, a high number of students used “learning” (net increase of 18 students), “input” (+17), “bias” (+15), and “utilization” (+13). Transition analysis revealed that this increase was driven by new adoption; for instance, 37 (31.4%) students newly used the term “learning” and 20 (16.9%) students newly used “bias” in the posttest,

indicating the acquisition of new perspectives. Contrastingly, the use of words describing the clinical case’s initial facts decreased. Only a few students used words such as “symptoms” (decrease of 11 students), “examination” (-11), and “explanation” (-9). Notably, transition analysis indicated that 23 (19.5%) students who used the term “symptoms” at pretest dropped it at posttest, suggesting a shift away from merely describing clinical manifestations. [Table 6](#) depicts the top 10 words with the largest changes and their transition details.

Table 6. Changes in word usage frequencies and transition analysis for “patient explanation” (question 1; N=118).

Word	Pretest, n (%)	Posttest, n (%)	Net change	New adoption ^a , n (%)	Dropped ^b , n (%)
Words with increased usage					
Learning	32 (27.1)	50 (42.4)	+18	37 (31.4)	19 (16.1)
Input	20 (16.9)	37 (31.4)	+17	29 (24.6)	12 (10.2)
Bias	10 (8.5)	25 (21.2)	+15	20 (16.9)	5 (4.2)
Utilization	12 (10.2)	25 (21.2)	+13	20 (16.9)	7 (5.9)
Misdiagnosis	7 (5.9)	20 (16.9)	+13	15 (12.7)	2 (1.7)
Information	24 (20.3)	35 (29.7)	+11	25 (21.2)	14 (11.9)
Case	31 (26.3)	40 (33.9)	+9	22 (18.6)	13 (11)
Hospital	9 (7.6)	17 (14.4)	+8	13 (11)	5 (4.2)
Convenient	4 (3.4)	10 (8.5)	+6	9 (7.6)	3 (2.5)
Judgment	32 (27.1)	38 (32.2)	+6	16 (13.6)	10 (8.5)
Words with decreased usage					
Symptoms	27 (22.9)	15 (12.7)	-12	11 (9.3)	23 (19.5)
Examination	18 (15.3)	7 (5.9)	-11	4 (3.4)	15 (12.7)
Explanation	22 (18.6)	13 (11)	-9	8 (6.8)	17 (14.4)
Presentation	12 (10.2)	4 (3.4)	-8	2 (1.7)	10 (8.5)
During	11 (9.3)	4 (3.4)	-7	2 (1.7)	9 (7.6)
Consultation	23 (19.5)	16 (13.6)	-7	8 (6.8)	15 (12.7)
View	8 (6.8)	2 (1.7)	-6	1 (0.8)	7 (5.9)
Result	29 (24.6)	24 (20.3)	-5	13 (11)	18 (15.3)
Certain	23 (19.5)	18 (15.3)	-5	8 (6.8)	13 (11)
Many	15 (12.7)	10 (8.5)	-5	5 (4.2)	10 (8.5)

^aThe term “new adoption” indicates students who did not use the word at pretest but used it at posttest, reflecting the acquisition of new concepts.

^b“Dropped” indicates students who used the word at pretest but excluded it at posttest.

Subsequently, we assessed the structural changes in students’ thinking by comparing the cooccurrence networks between pretest and posttests. The patient explanation network’s Jaccard coefficient was 0.1156, indicating very low similarity between the 2 time points. This dramatic shift was driven by a large number of unique pairs at each time point. The network had only 264 shared cooccurrence pairs; however, 1030 unique pairs existed exclusively at pretest, and 989 unique pairs emerged at

posttest. These quantitative findings are supported by a qualitative analysis of the network structures. In the pretest, the network was structured around the case’s core elements ([Figure 1](#)). However, the network became highly complex and specific at posttest, with keywords like “learning” and “data” being directly linked to “AI” and new vocabulary, such as “bias” and “blind acceptance,” explaining the limitations and risks of AI use ([Figure 2](#)).

Figure 1. The network visualizes the relationships between keywords in students' explanations to the patient before the program. Each node represents a word, and the size and color of the node indicate its degree (number of connections). The thickness and color of the edge (line) connecting 2 nodes represent the number of students who used those 2 words adjacently.

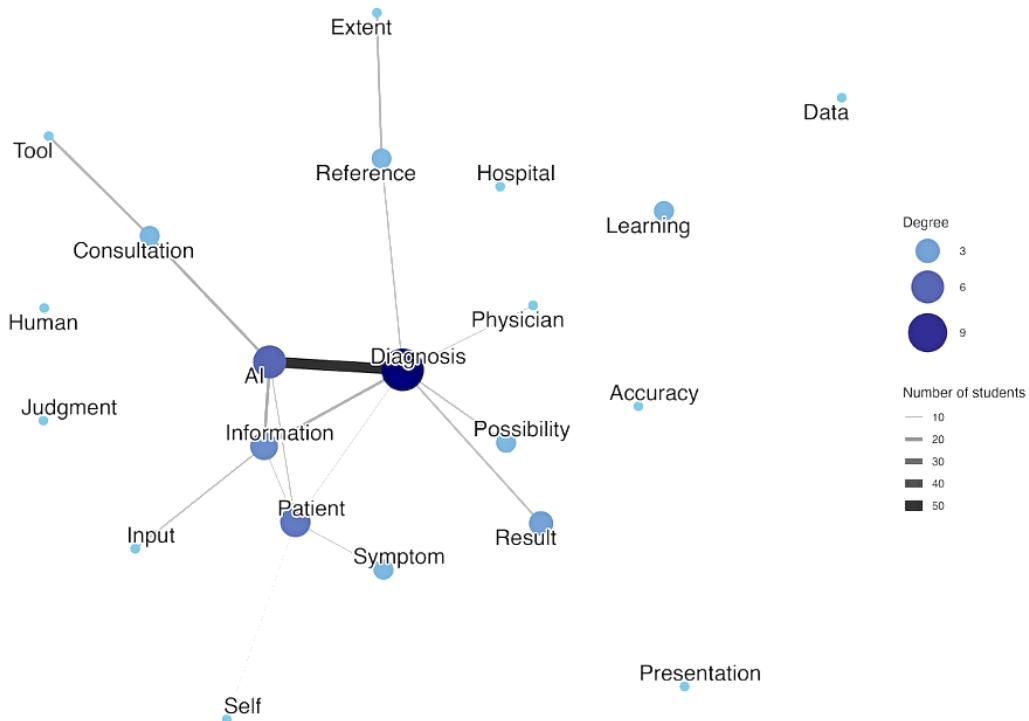
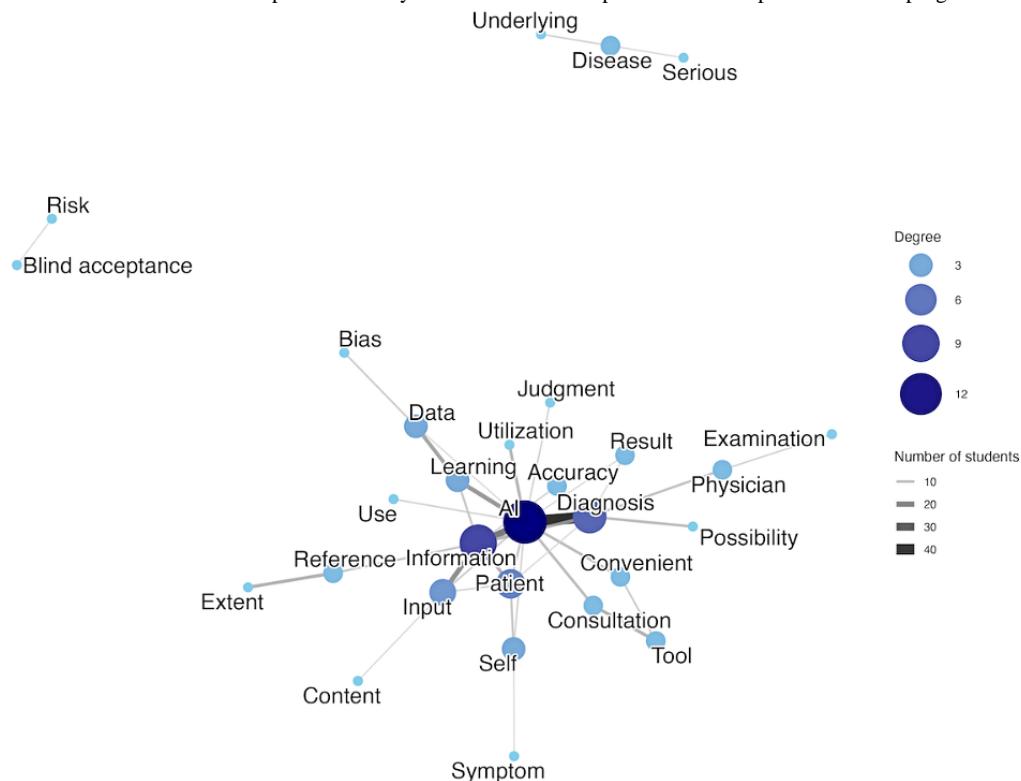


Figure 2. The network visualizes the relationships between keywords in students' explanations to the patient after the program.



Responses to Question 2

For question 2, a significant increase in the use of words related to professional and ethical responsibilities was observed. The number of students using key concepts such as "information" (net increase of 25 students), "personal info" (+14), "accuracy" (+13), and "bias" (+10) increased significantly, as well. Notably,

31 (26.3%) students newly adopted the term "personal info" in their posttest responses, suggesting a shift in their awareness of privacy issues. Contrastingly, the number of students using abstract or cautionary words, such as "possibility" (decrease of 9 students), "reference" (-5), and "keep in mind" (-5), decreased. Notably, transition analysis revealed that 30 (25.4%) students dropped the term "reference" and 24 (20.3%) students

dropped “possibility” in their posttest responses, suggesting a departure from vague expressions. **Table 7** depicts the top 10

Table 7. Changes in word usage frequencies and transition analysis for “physician precautions” (question 2; N=118).

Word	Pretest, n (%)	Posttest, n (%)	Net change	New adoption ^a , n (%)	Dropped ^b , n (%)
Words with increased usage					
Information	56 (47.5)	81 (68.6)	+25	38 (32.2)	13 (11)
AI ^c	49 (41.5)	66 (55.9)	+17	29 (24.6)	12 (10.2)
Personal info	36 (30.5)	50 (42.4)	+14	31 (26.3)	17 (14.4)
Accuracy	22 (18.6)	35 (29.7)	+13	26 (22)	13 (11)
Input	2 (1.7)	12 (10.2)	+10	12 (10.2)	2 (1.7)
Bias	5 (4.2)	15 (12.7)	+10	13 (11)	3 (2.5)
Patient	18 (15.3)	26 (22)	+8	17 (14.4)	9 (7.6)
Assistance	11 (9.3)	18 (15.3)	+7	12 (10.2)	5 (4.2)
Final	14 (11.9)	20 (16.9)	+6	15 (12.7)	9 (7.6)
View	0 (0)	5 (4.2)	+5	5 (4.2)	0 (0)
Words with decreased usage					
Possibility	37 (31.4)	28 (23.7)	-9	15 (12.7)	24 (20.3)
Reference	67 (56.8)	61 (51.7)	-6	24 (20.3)	30 (25.4)
Oneself	17 (14.4)	11 (9.3)	-6	4 (3.4)	10 (8.5)
Keep in mind	15 (12.7)	10 (8.5)	-5	7 (5.9)	12 (10.2)
Presentation	12 (10.2)	7 (5.9)	-5	5 (4.2)	10 (8.5)
Knowledge	15 (12.7)	10 (8.5)	-5	5 (4.2)	10 (8.5)
Examination	8 (6.8)	4 (3.4)	-4	3 (2.5)	7 (5.9)
Caution	54 (45.8)	50 (42.4)	-4	24 (20.3)	28 (23.7)
Output	5 (4.2)	1 (0.8)	-4	1 (0.8)	5 (4.2)
Source	17 (14.4)	13 (11)	-4	6 (5.1)	10 (8.5)

^aThe term “new adoption” indicates students who did not use the word at pretest but used it at posttest, reflecting the acquisition of new concepts.

^b“Dropped” indicates students who used the word at pretest but excluded it at posttest.

^cAI: artificial intelligence.

The cooccurrence network for physician precautions showed a significant change, as well. The Jaccard coefficient was 0.1213, confirming the occurrence of a major structural transformation. The low score was driven by a large number of unique pairs. Only 191 pairs were shared, whereas 631 unique pairs were found exclusively at pretest, and 753 new pairs appeared at posttest. The higher number of unique pairs found at posttest compared to pretest suggests that students not only abandoned

previously acquired cautionary topics but also actively explored a wider range of new professional and ethical concerns. These quantitative results were corroborated by a qualitative analysis of network structures. The pretest network comprised 3 separate, unconnected clusters (**Figure 3**), which transformed into a single, integrated network at posttest, where “AI” and “information” acted as hubs (**Figure 4**).

Figure 3. The network visualizes the relationships between keywords regarding precautions for physicians before the program.

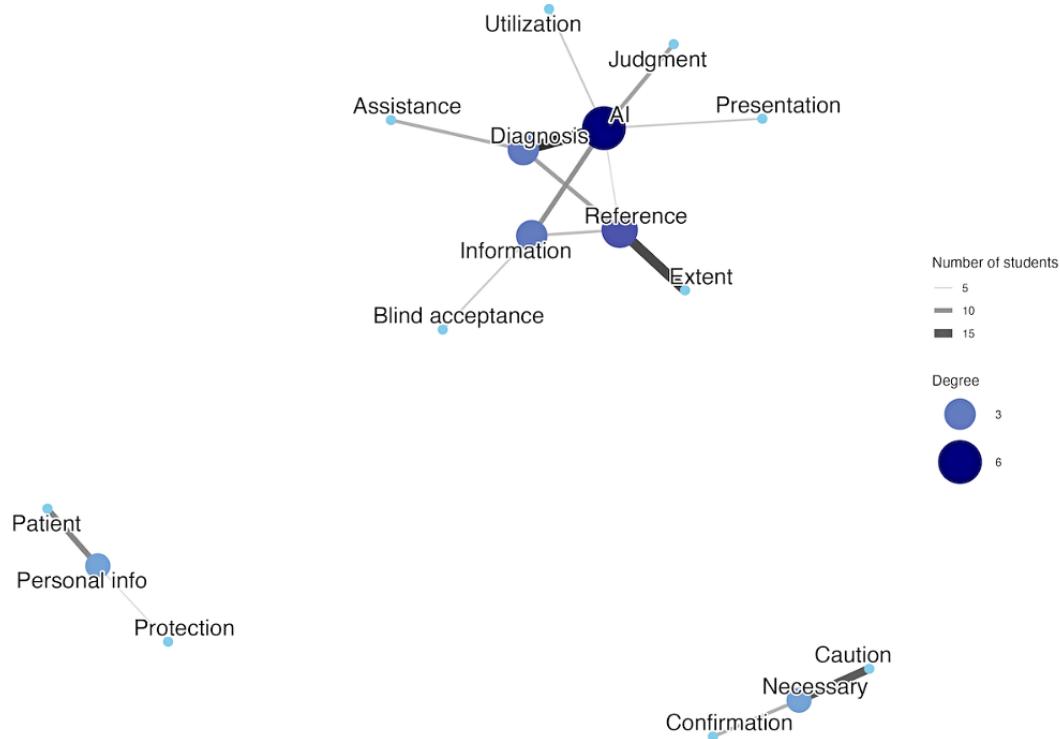
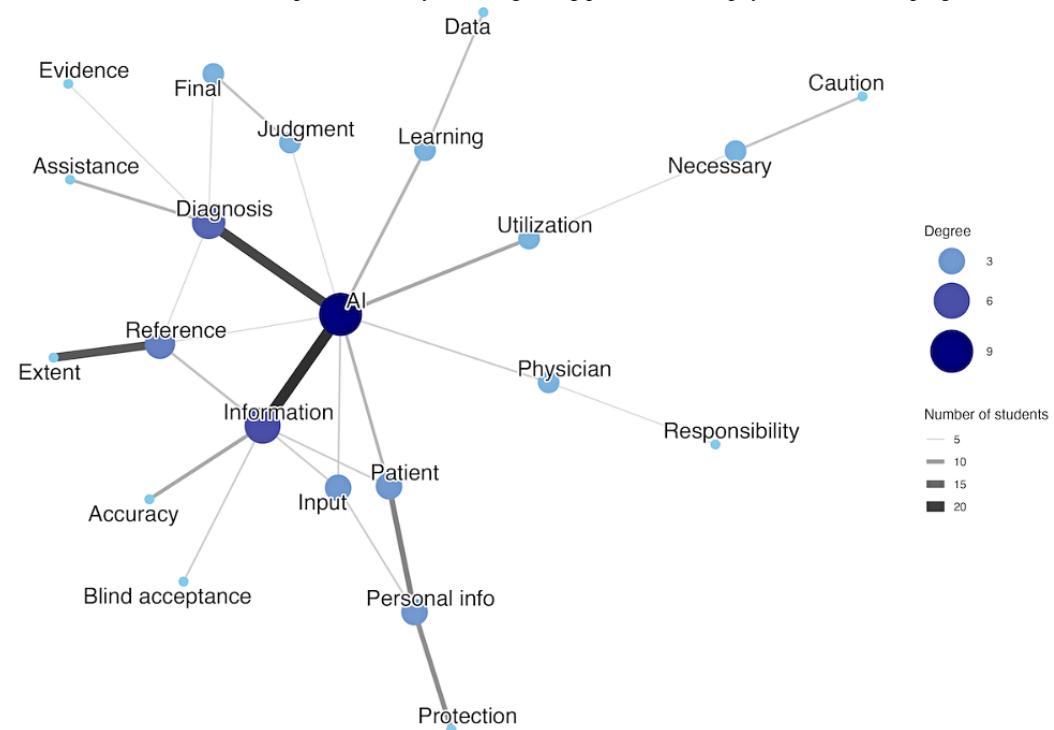


Figure 4. The network visualizes the relationships between keywords regarding precautions for physicians after the program.



Students' Perceptions of the Program

Among the 118 participants, 76 (64.4%) completed an optional postprogram survey based on the ARCS model. On a 5-point Likert scale, students rated the program highly for attention (mean 4.01, SD 0.62), relevance (mean 4.20, SD 0.71), confidence (mean 3.78, SD 0.87), and satisfaction (mean 4.03, SD 0.75). A repeated-measures ANOVA was conducted to compare and reveal a statistically significant difference between

the mean scores of the 4 ARCS components ($F_{3, 225}=13.15$; $P<.001$). Additionally, post hoc tests with Bonferroni correction revealed that the mean score for confidence was significantly lower than the scores for attention, relevance, and satisfaction. Furthermore, relevance was rated significantly higher than satisfaction.

Discussion

Principal Findings

This study's primary objective was to evaluate the effectiveness of a newly developed PBL program. Further, this study addresses the following RQs:

RQ1: The educational program improves students' objective knowledge of AI. Accordingly, we hypothesized that foundational knowledge would increase after the lecture phase (from pretest to midtest) and the subsequent PBL phase (from midtest to posttest) would produce a significant additional increase in practical and ethical knowledge.

RQ2: The educational program qualitatively transforms students' thinking regarding AI-related ethical challenges from an abstract understanding to a concrete and structured thought process based on the clinical context.

Regarding RQ1, although a significant improvement in MCQ scores was observed for the program in its entirety, no statistically significant "additional effect" of the PBL phase was observed. Regarding RQ2, text mining results suggested that after the intervention, students' thinking underwent a qualitative transformation, which was characterized by the use of specific and professional terms, such as "bias," "personal information," and "guidelines," and the development of a structured thought process.

Interpretation of Findings

Regarding RQ1, a clear additional effect of PBL was not detected in MCQ scores due to 2 main reasons. First, the knowledge measured by MCQ was the foundational content that probably had been efficiently acquired during the lecture phase. The second reason is the limitation of the MCQ format itself. As argued by van der Vleuten and Schuwirth [27] in their paper on PBL assessment, traditional standardized knowledge tests cannot adequately assess abilities such as clinical reasoning and metacognition, which are fostered by PBL. Consequently, the lack of significant MCQ score improvement during the PBL phase does not imply a lack of learning; rather, it suggests that the unique contribution of PBL lies in the qualitative structural transformation of students' thinking—specifically, a shift from abstract knowledge to context-based application—which is difficult to capture with standard testing. Although MCQs are suitable to measure objective knowledge, they may have low sensitivity in capturing changes in the higher-order practical skills fostered by PBL. Indeed, reviews of earlier research on PBL's effectiveness indicate that whereas knowledge acquisition is comparable to or better than traditional lecture-based learning, PBL is more effective in improving higher-order skills, such as communication, problem-solving, self-learning, and metacognitive competencies [28,29]. Therefore, the reason a statistically significant additional effect of PBL was not detected in MCQ scores is that PBL's true effect was reflected in the qualitative changes in students' written responses, rather than the multiple-choice questions themselves.

This is clarified by the results for RQ2. The quantitative text analysis revealed that the program structurally transformed students' thinking. Among the responses to question 1 (patient

explanation), the increased use of words related to AI's mechanisms and limitations was driven by new adoption. For instance, 37 (31.4%) students newly adopted the term "learning" and 20 (16.9%) newly adopted "bias." This suggests that students based their explanations of the bases and limitations of AI's judgments on the transparency principle (explainability) and accountability principle detailed in the AI Utilization Guideline and AI Business Guideline [19,20]. Conversely, 23 (19.5%) students dropped the term "symptoms," indicating a shift away from merely describing clinical manifestations to addressing the underlying logic of the AI. Similarly, among the responses to question 2 (physician precautions), the significant increase in the use of words related to ethical and professional responsibilities was also driven by new adoption. Notably, 31 (26.3%) students newly adopted "personal info," which can be interpreted as the result of students' efforts to learn and apply the privacy principle [19,20]. Simultaneously, we observed a decrease in abstract vocabulary. Transition analysis revealed that 30 (25.4%) students dropped the term "reference" and 24 (20.3%) dropped "possibility." This reduction in vague or passive expressions suggests that students adopted a more specific and responsible professional stance. The cooccurrence network analysis' quantitative results support the inference that the qualitative change in thinking was accompanied by the structuring of knowledge. The cooccurrence network for question 1 revealed a very low Jaccard coefficient (0.1156) between the pretest and postintervention networks. This indicates the significant transformation of the contextual relationships between words. For 264 common pairs, there were 1030 unique pairs before and 989 unique pairs after the intervention; accordingly, new knowledge and concepts were established by replacing traditional knowledge. Moreover, the Jaccard coefficient for the cooccurrence network of question 2 was low (0.1213), indicating a similar structural change. In the question 2 cooccurrence network, whereas "AI's function," "privacy," and "general caution" existed as separate clusters at pretest, a more integrated network was formed with the word "information" as a new hub, connecting "AI" and "diagnosis" with important concepts like "accuracy," at posttest. The study's quantitative and qualitative findings resonate with the argument of McCoy et al [12] that medical students must possess knowledge about AI. They argue that future physicians must be able to not only "use" AI but also "interpret" its results, recognize potential errors and biases, and "explain" the results and processes to patients and other health care professionals. In this study, the qualitative improvement in the postintervention descriptions highlight that students' abilities expanded from the "use" perspective to the "interpret and explain" perspective.

One reason for fostering this qualitative transformation of thinking is interpreted from the perspective of the 4 components of the ARCS model of motivation. First, the clinical case presented as a video at the beginning of the group work session effectively captured students' attention (mean rating 4.01, SD 0.62). A systematic review of emotions' role in medical education highlights the significance of emotional experiences in the learning process [30]. The video depicted a medical student facing an ethical dilemma and potentially fostered emotional engagement, thereby enhancing the students' motivation for subsequent learning. Second, their awareness of

the challenge involved in responding to a patient who is overreliant on AI probably enabled students to clearly understand the learning content's relevance to their future practice (mean rating 4.20, SD 0.71). This educational design aligns with the principles of "authentic learning," which involves the solving of complex, real-world contextual problems to enhance learners' competency in the practical application of knowledge [31]. Further, students perceived the case in this program as a personal and relevant task due to the effective functioning of the elements of authentic learning. On the other hand, the students' confidence in addressing a complex ethical issue resulted in a mean rating of 3.78 (SD 0.87), which was significantly lower than the values for attention, relevance, and satisfaction ($F_{3, 225}=13.15; P<.001$). Therefore, when the program prompted a qualitative transformation in the students' thinking, a single intervention was insufficient for them to acquire complete confidence in such a complex and uncertain topic as AI ethics. This program was limited to considering the explanation's content to the patient and his son and did not provide any opportunities for practical application, such as a role-play of the explanation. This lack of practical experience probably contributed to the students' low confidence in applying knowledge. An earlier study highlights the importance of maintaining a balance between confidence and humility and adopting a "not-knowing" stance as a key clinical leadership competency aligned with the Japanese health care culture [31]. The low confidence scores observed in our study were interpreted as not any failure to learn, but rather the development of the intellectual humility required by a professional, particularly those confronting complex and uncertain topics such as AI ethics.

However, the high ratings for components other than confidence, particularly attention and relevance, comprised a strong motivation for students to clearly understand the subject's importance and engage in self-directed learning [21]. The effective functioning of these elements enabled the students to complete a learning process involving deep thought, and the resulting insights probably led to the final Satisfaction (mean rating 4.03, SD 0.75) and structuring of thought observed in this study.

Research Significance and Limitations and Future

Research Scope

The primary significance of this study is that it develops, implements, and comprehensively evaluates an educational program themed on a realistic and novel clinical challenge that may be faced by physicians in this era of widespread AI use: responding to a patient who relies heavily on AI. Presenting this ethical dilemma as a video teaching material has pedagogical validity because a systematic literature review

indicates that video materials can effectively stimulate student discussion in ethics education [32]. Furthermore, the visualization of the program's educational effect as not only a change in objective knowledge but also a qualitative and structural change in thinking using text mining is a novel approach to evaluate the effects of constructivist learning on complex topics such as AI ethics.

This study has several limitations. First, this study used a single-group pretest and posttest design without a control group. Therefore, while we observed significant changes, we cannot fully rule out the influence of external factors or natural maturation. Second, because it was conducted at a single university in Japan, the generalizability of its results may be limited by the specific cultural and educational context. However, we emphasize that the core framework of this program—integrating technical AI knowledge with ethical PBL scenarios—is adaptable. Future multiinstitutional collaborative studies are needed to validate whether similar structural changes in thinking occur in different educational and cultural contexts. Third, this study considered only the program's short-term effects on learning. Therefore, the participating students must be subjected to a follow-up study when they proceed to their clinical clerkships to verify the learning effects' persistence. Clarifying how the qualitative transformation in thinking indicated by this study influences students' decision-making and behavior in clinical settings is important to evaluate the value of this type of educational programs. Fourth, as mentioned earlier, the MCQ format may not fully capture the effects of PBL. Therefore, there is an urgent need to develop and validate new assessment measures that can objectively assess the practical reasoning and communication skills that cannot be captured by applying MCQs alone.

Conclusions

The results of this study suggest that the PBL program developed and implemented in this study influenced both medical students' knowledge of AI and their thoughts on its ethical challenges. Although the program in its entirety resulted in a significant increase in objective knowledge, a statistically significant additional effect was not confirmed from the PBL phase alone. However, there was a clear change in the qualitative data revealed by text mining: Students' thinking underwent qualitative transformation from an abstract level to a concrete and structured thought process based on the clinical context. Therefore, this study indicates the potential of PBL as an effective pedagogical method to foster higher-order thinking skills in complex topics, such as AI ethics. Hence, this program's design and evaluation framework can serve as a practical model to develop and evaluate future AI-related curricula in medical education.

Acknowledgments

We extend our sincere gratitude to all the students who participated in this study. Our heartfelt thanks are also directed to Ms Akane Sumi for her cooperation in the video recording and to Ms Kazumi Kato for her professional administrative support at the graduate school.

Funding

This work was supported by research funding at the Medical Education Center of Jichi Medical University.

Data Availability

The datasets generated and/or analyzed during this study are not publicly available due to ethical constraints but are available from the corresponding author on reasonable request.

Authors' Contributions

YO was involved in all study aspects. YA contributed to the administration of the educational program and the analysis and interpretation of data. SK contributed to the creation of the evaluation instruments and the analysis and interpretation of data. YM contributed to the conceptualization and supervision of all study aspects and critically reviewed the manuscript. TK and MSY contributed significantly to the administration of problem-based learning sessions and the writing of the manuscript. All the authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Masters K. Artificial intelligence in medical education. *Med Teach* 2019;41(9):976-980. [doi: [10.1080/0142159X.2019.1595557](https://doi.org/10.1080/0142159X.2019.1595557)] [Medline: [31007106](https://pubmed.ncbi.nlm.nih.gov/31007106/)]
2. 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](https://doi.org/10.2196/16048)] [Medline: [31793895](https://pubmed.ncbi.nlm.nih.gov/31793895/)]
3. Rincón EHH, Jimenez D, Aguilar L, Flórez JMP, Tapia AER, Peñuela CLJ. Mapping the use of artificial intelligence in medical education: a scoping review. *BMC Med Educ* 2025;25(1):526 [FREE Full text] [doi: [10.1186/s12909-025-07089-8](https://doi.org/10.1186/s12909-025-07089-8)] [Medline: [40221725](https://pubmed.ncbi.nlm.nih.gov/40221725/)]
4. Car J, Ong QC, Erlikh Fox T, Leightley D, Kemp SJ, Švab I, Digital Health Systems Collaborative, et al. The digital health competencies in medical education framework: an international consensus statement based on a Delphi study. *JAMA Netw Open* 2025;8(1):e2453131 [FREE Full text] [doi: [10.1001/jamanetworkopen.2024.53131](https://doi.org/10.1001/jamanetworkopen.2024.53131)] [Medline: [39888625](https://pubmed.ncbi.nlm.nih.gov/39888625/)]
5. Longhini J, Rossettini G, Palese A. Digital Health competencies among health care professionals: systematic review. *J Med Internet Res* 2022;24(8):e36414 [FREE Full text] [doi: [10.2196/36414](https://doi.org/10.2196/36414)] [Medline: [35980735](https://pubmed.ncbi.nlm.nih.gov/35980735/)]
6. 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](https://doi.org/10.1097/ACM.0000000000004291)] [Medline: [34348374](https://pubmed.ncbi.nlm.nih.gov/34348374/)]
7. Amiri H, Peiravi S, Rezazadeh Shojaee SS, Rouhparvarzamin M, Nateghi MN, Etemadi MH, et al. Medical, dental, and nursing students' attitudes and knowledge towards artificial intelligence: a systematic review and meta-analysis. *BMC Med Educ* 2024;24(1):412 [FREE Full text] [doi: [10.1186/s12909-024-05406-1](https://doi.org/10.1186/s12909-024-05406-1)] [Medline: [38622577](https://pubmed.ncbi.nlm.nih.gov/38622577/)]
8. Çalışkan SA, Demir K, Karaca O. Artificial intelligence in medical education curriculum: an e-Delphi study for competencies. *PLoS One* 2022;17(7):e0271872 [FREE Full text] [doi: [10.1371/journal.pone.0271872](https://doi.org/10.1371/journal.pone.0271872)] [Medline: [35862401](https://pubmed.ncbi.nlm.nih.gov/35862401/)]
9. Weidener L, Fischer M. Artificial intelligence in medicine: cross-sectional study among medical students on application, education, and ethical aspects. *JMIR Med Educ* 2024;10:e51247 [FREE Full text] [doi: [10.2196/51247](https://doi.org/10.2196/51247)] [Medline: [38180787](https://pubmed.ncbi.nlm.nih.gov/38180787/)]
10. Hale J, Alexander S, Wright ST, Gilliland K. Generative AI in undergraduate medical education: a rapid review. *J Med Educ Curric Dev* 2024;11. [doi: [10.1177/23821205241266697](https://doi.org/10.1177/23821205241266697)]
11. Singla R, Pupic N, Ghaffarizadeh S, Kim C, Hu R, Forster BB, et al. Developing a Canadian artificial intelligence medical curriculum using a Delphi study. *NPJ Digit Med* 2024;7(1):323 [FREE Full text] [doi: [10.1038/s41746-024-01307-1](https://doi.org/10.1038/s41746-024-01307-1)] [Medline: [39557985](https://pubmed.ncbi.nlm.nih.gov/39557985/)]
12. 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 [FREE Full text] [doi: [10.1038/s41746-020-0294-7](https://doi.org/10.1038/s41746-020-0294-7)] [Medline: [32577533](https://pubmed.ncbi.nlm.nih.gov/32577533/)]
13. Weidener L, Fischer M. Proposing a principle-based approach for teaching AI ethics in medical education. *JMIR Med Educ* 2024;10:e55368 [FREE Full text] [doi: [10.2196/55368](https://doi.org/10.2196/55368)] [Medline: [38285931](https://pubmed.ncbi.nlm.nih.gov/38285931/)]
14. Medical Education Model Core Curriculum Expert Research Committee. The model core curriculum for medical education in Japan 2022 revision. Ministry of Education, Culture, Sports, Science and Technology. URL: https://www.mext.go.jp/content/20250411-mxt_igaku-000028108_00003-2.pdf [accessed 2025-12-23]
15. Onoue T, Asada Y, Imafuku R, Kou S, Takami H, Takahashi Y, et al. Developing competencies relating to information science and technology in Japanese undergraduate medical education. *Med Teach* 2024;46(sup1):S31-S37 [FREE Full text] [doi: [10.1080/0142159X.2024.2385199](https://doi.org/10.1080/0142159X.2024.2385199)] [Medline: [39545502](https://pubmed.ncbi.nlm.nih.gov/39545502/)]
16. Ota Y, Asada Y, Mieno M, Matsuyama Y. Competencies required to make use of information science and technology among Japanese medical students: a cross-sectional study. *BMC Med Educ* 2024;24(1):840 [FREE Full text] [doi: [10.1186/s12909-024-05786-4](https://doi.org/10.1186/s12909-024-05786-4)] [Medline: [39107733](https://pubmed.ncbi.nlm.nih.gov/39107733/)]

17. Taylor D, Miflin B. Problem-based learning: where are we now? *Med Teach* 2008;30(8):742-763. [doi: [10.1080/01421590802217199](https://doi.org/10.1080/01421590802217199)] [Medline: [18946818](https://pubmed.ncbi.nlm.nih.gov/18946818/)]
18. Ota Y, Asada Y, Matsuyama Y. Proposal of a rubric for 'Competency in Utilizing IT: Information Science and Technology' education (in Japanese). *Igaku Kyoiku* 2025;56(1):40-41. [doi: [10.11307/mededjapan.56.1_40](https://doi.org/10.11307/mededjapan.56.1_40)]
19. Ministry of Internal Affairs and Communications (MIC), Japan. AI utilization guidelines: practical reference for AI utilization. 2019 Presented at: The Conference toward AI Network Society; August 9, 2019; Mita Kyoyo Kaigisho URL: https://www.soumu.go.jp/main_content/000658284.pdf
20. Ministry of Internal Affairs and Communications, Ministry of Economy, Trade and Industry. AI guidelines for business (ver1.1). 2025. URL: https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20240419_14.pdf [accessed 2025-12-23]
21. Ota Y, Asada Y, Matsuyama Y. Implementation of a PBL program on the credibility of AI-based clinical judgments: insights from the ARCS model (in Japanese). *Igaku Kyoiku* 2025;56(3):181-188. [doi: [10.11307/mededjapan.56.3_181](https://doi.org/10.11307/mededjapan.56.3_181)]
22. Keller J. Motivational Design for Learning and Performance: The ARCS Model Approach. London: Springer; 2010.
23. R Core Team. R: a language and environment for statistical computing. 2023. URL: <https://www.R-project.org/> [accessed 2025-12-23]
24. Silge J, Robinson D. tidytext: text mining and analysis using tidy data principles in R. *J Open Source Softw* 2016;1(3):37. [doi: [10.21105/joss.00037](https://doi.org/10.21105/joss.00037)]
25. Ishida M. RMeCab: an R interface to MeCab, a Japanese morphological analysis engine. GitHub. 2008. URL: <https://github.com/IshidaMotohiro/RMeCab> [accessed 2025-12-23]
26. Kang K, Han SJ, Chun J, Kim H. Research trends related to childhood and adolescent cancer survivors in South Korea using word co-occurrence network analysis. *Child Health Nurs Res* 2021;27(3):201-210 [FREE Full text] [doi: [10.4094/chnr.2021.27.3.201](https://doi.org/10.4094/chnr.2021.27.3.201)] [Medline: [35004510](https://pubmed.ncbi.nlm.nih.gov/35004510/)]
27. van der Vleuten CPM, Schuwirth LWT. Assessment in the context of problem-based learning. *Adv Health Sci Educ Theory Pract* 2019;24(5):903-914 [FREE Full text] [doi: [10.1007/s10459-019-09909-1](https://doi.org/10.1007/s10459-019-09909-1)] [Medline: [31578642](https://pubmed.ncbi.nlm.nih.gov/31578642/)]
28. 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;22(1):104 [FREE Full text] [doi: [10.1186/s12909-022-03154-8](https://doi.org/10.1186/s12909-022-03154-8)] [Medline: [35177063](https://pubmed.ncbi.nlm.nih.gov/35177063/)]
29. Stentoft D. Problem-based projects in medical education: extending PBL practices and broadening learning perspectives. *Adv Health Sci Educ Theory Pract* 2019;24(5):959-969. [doi: [10.1007/s10459-019-09917-1](https://doi.org/10.1007/s10459-019-09917-1)] [Medline: [31641941](https://pubmed.ncbi.nlm.nih.gov/31641941/)]
30. Lee J, Campbell S, Choi M, Bae J. Authentic learning in healthcare education: a systematic review. *Nurse Educ Today* 2022;119:105596. [doi: [10.1016/j.nedt.2022.105596](https://doi.org/10.1016/j.nedt.2022.105596)] [Medline: [36283196](https://pubmed.ncbi.nlm.nih.gov/36283196/)]
31. Shikama Y, Oikawa S, Stanyon M, Yasuda M, Otani K. Culturally-aligned clinical leadership competencies for effective teamwork in Japanese healthcare. *BMC Med Educ* 2024;24(1):1315 [FREE Full text] [doi: [10.1186/s12909-024-06272-7](https://doi.org/10.1186/s12909-024-06272-7)] [Medline: [39548540](https://pubmed.ncbi.nlm.nih.gov/39548540/)]
32. Rattani A, Kaakour D, Syed RH, Kaakour A. Changing the channel on medical ethics education: systematic review and qualitative analysis of didactic-icebreakers in medical ethics and professionalism teaching. *Monash Bioeth Rev* 2021;39(1):125-140. [doi: [10.1007/s40592-020-00120-2](https://doi.org/10.1007/s40592-020-00120-2)] [Medline: [33070300](https://pubmed.ncbi.nlm.nih.gov/33070300/)]

Abbreviations

AI: artificial intelligence

ARCS: attention, relevance, confidence, and satisfaction

IST: information science and technology

MCQ: multiple-choice questionnaire

PBL: problem-based learning

RQ: research question

Edited by D Chartash; submitted 22.Sep.2025; peer-reviewed by N Bievre, A Marušić; comments to author 07.Dec.2025; accepted 21.Dec.2025; published 14.Jan.2026.

Please cite as:

Ota Y, Asada Y, Kubo S, Kanno T, Yagi MS, Matsuyama Y

Evaluation of a Problem-Based Learning Program's Effect on Artificial Intelligence Ethics Among Japanese Medical Students: Mixed Methods Study

JMIR Med Educ 2026;12:e84535

URL: <https://mededu.jmir.org/2026/1/e84535>

doi: [10.2196/84535](https://doi.org/10.2196/84535)

PMID:

©Yuma Ota, Yoshikazu Asada, Saori Kubo, Takeshi Kanno, Machiko Saeki Yagi, Yasushi Matsuyama. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 14.Jan.2026. 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.

Original Paper

AI-Driven Objective Structured Clinical Examination Generation in Digital Health Education: Comparative Analysis of Three GPT-4o Configurations

Zineb Zouakia^{1,2}, MD; Emmanuel Logak^{1,2}, MD; Alan Szymczak^{1,2}, MD; Jean-Philippe Jais^{3,4}, MD, PhD; Anita Burgun^{1,2}, MD, PhD; Rosy Tsopra^{1,2,5}, MD, PhD

¹Clinical Bioinformatics Laboratory, Imagine Institute, Université Paris Cité, INSERM UMR1163, Paris, France

²Department of Medical Informatics, Hôpital Européen Georges Pompidou, Hôpital Necker Enfants Malades, AP-HP, Paris, France

³Imagine Institute, Université Paris Cité, Paris, France

⁴Biostatistic Unit, Necker University Hospital, AP-HP, Paris, France

⁵UPPERS US 007 F-75006, Université Paris Cité, Paris, France

Corresponding Author:

Zineb Zouakia, MD

Clinical Bioinformatics Laboratory

Imagine Institute

Université Paris Cité, INSERM UMR1163

24 Bd du Montparnasse, 75015 Paris

Paris, F-75006

France

Phone: 33 668169844

Email: zineb.zouakia@gmail.com

Abstract

Background: Objective Structured Clinical Examinations (OSCEs) are used as an evaluation method in medical education, but require significant pedagogical expertise and investment, especially in emerging fields like digital health. Large language models (LLMs), such as ChatGPT (OpenAI), have shown potential in automating educational content generation. However, OSCE generation using LLMs remains underexplored.

Objective: This study aims to evaluate 3 GPT-4o configurations for generating OSCE stations in digital health: (1) standard GPT with a simple prompt and OSCE guidelines; (2) personalized GPT with a simple prompt, OSCE guidelines, and a reference book in digital health; and (3) simulated-agents GPT with a structured prompt simulating specialized OSCE agents and the digital health reference book.

Methods: Overall, 24 OSCE stations were generated across 8 digital health topics with each GPT-4o configuration. Format compliance was evaluated by one expert, while educational content was assessed independently by 2 digital health experts, blind to GPT-4o configurations, using a comprehensive assessment grid. Statistical analyses were performed using Kruskal-Wallis tests.

Results: Simulated-agents GPT performed best in format compliance and most content quality criteria, including accuracy (mean 4.47/5, SD 0.28; $P=.01$) and clarity (mean 4.46/5, SD 0.52; $P=.004$). It also had 88% (14/16) for usability without major revisions and first-place preference ranking, outperforming the other configurations. Personalized GPT showed the lowest format compliance, while standard GPT scored lowest for clarity and educational value.

Conclusions: Structured prompting strategies, particularly agents' simulation, enhance the reliability and usability of LLM-generated OSCE content. These results support the use of artificial intelligence in medical education, while confirming the need for expert validation.

(*JMIR Med Educ* 2026;12:e82116) doi:[10.2196/82116](https://doi.org/10.2196/82116)

KEYWORDS

large language models; generative artificial intelligence; ChatGPT; GPT-4o; medical education; objective structured clinical examination; digital health; digital health education; medical informatics; prompt design

Introduction

Since ChatGPT was launched by OpenAI in 2022, large language models (LLMs) have undergone rapid development, leading to growing interest in medical education [1-6]. Their ability to produce diverse and context-specific texts from simple prompts offers new opportunities to create and improve educational resources.

In medical education, several studies explored the ability of LLMs to answer medical examination questions [7-12], generate multiple-choice questions [13-16], or simulate patients [17-26], with results considered promising by both educators and learners. However, using LLMs to generate clinical scenarios is still new and has only been explored in a few studies, across three situations:

1. Generation of clinical cases or vignettes (ie, scenarios given to students without interaction): LLMs demonstrated their ability to generate cases considered accurate and appropriate in pharmacology [27], but also useful [28] and usable (with moderate or minor edits) in general medicine [29]. However, the authors noticed some issues with realism [28] and precision [28,29].
2. Generation of clinical simulations (ie, scenarios given to students with interaction, such as simulated patients): LLMs demonstrated their ability to generate accurate cases in asthma [30], as well as realistic cases in the nursing field [31]. They also offered potential time savings, reducing development time by 2.6 hours compared to experts [32]. However, limitations remained, such as the omission of critical patient characteristics [31].
3. Generation of Objective Structured Clinical Examination (OSCE) stations: These scenarios, introduced by Harden and Gleeson in 1979 [33], are key assessment tools in medical education. OSCEs consist of simulated clinical situations conducted within a controlled and reproducible environment [34,35], where students interact with a standardized participant (SP), playing the role of a patient or health care professional. Students are observed and evaluated by an examiner using a standardized scoring rubric, based on predefined criteria. Previous studies explored the potential of LLMs to simulate standardized participants [36-38], examiners [39-43], or students [44-46]. Conceptual work has also discussed their potential to enhance OSCE development [47-49]. However, to our

knowledge, no study has yet generated OSCE stations using LLMs.

Designing OSCE stations is a complex and resource-intensive process [50]. It requires considerable clinical and educational expertise to produce realistic, standardized scenarios that accurately assess targeted competencies [51]. The need for precision, clarity, and reproducibility, combined with time and human resource constraints in educational teams, makes OSCE development challenging and burdensome [52,53]. In domains like digital health, this challenge is amplified by a shortage of available experts. In this context, LLMs could help educators by assisting them in creating OSCE stations.

To explore this hypothesis, we aimed to design and assess 3 configurations for generating OSCE stations using ChatGPT in the area of digital health, an emerging field in medical education.

Methods

Background

The study follows the recommendations of the METRICS (Model, Evaluation, Timing, Range/Randomization, Individual factors, Count, Specificity) checklist, for designing and reporting studies involving generative language models in medical education [54] (Multimedia Appendix 1).

To analyze the ability of ChatGPT to generate OSCE stations in digital health, we tested 3 configurations of the ChatGPT-4o model (Table 1). These differed by prompt type (simple or advanced) and the documents provided to the LLM (reference books for OSCE or digital health or both).

The configurations of the ChatGPT-4o model are listed as follows:

1. Configuration 1: standard GPT, used the free version of ChatGPT-4o with a simple prompt and the reference book for OSCE construction (named “OSCE Vadémécum”).
2. Configuration 2: personalized GPT, used the paid version of custom GPT, with a simple prompt, the OSCE Vadémécum, and the reference book on digital health from Université Paris Cité (named “UPCité reference book”).
3. Configuration 3: simulated-agents GPT, used the paid version of custom GPT, with a prompt simulating a fictional multiagent system specialized in OSCE construction (aligned with the OSCE Vadémécum), and the UPCité reference book.

Table 1. Description of the 3 GPT configurations analyzed.

Characteristics	Standard GPT	Personalized GPT	Simulated-agents GPT
Objective	Simulate minimal and spontaneous use of GPT	Simulate personalized use of GPT with a knowledge base specific to the medical field	Simulate advanced use of GPT with specialized assistants and a knowledge base specific to the medical field
Model	GPT-4o	GPT-4o	GPT-4o
Account type	Free	Paid	Paid
Prompt ^a	Simple	Simple	Simulated multiagent system
OSCE ^b Vadémécum ^c	PDF file	PDF file	Instructions embedded in the prompt
UPCité reference book ^d (knowledge base) ^{—e}		5 Microsoft Word documents	5 Microsoft Word documents

^aIn French ([Multimedia Appendix 2](#)).

^bOSCE: Objective Structured Clinical Examination.

^cNational reference document for designing OSCEs (67-page PDF).

^dThe UPCité reference book was written by 25 educators from 7 health disciplines at Université Paris Cité. It is the reference book for all health students (medicine, pharmacy, nursing, rehabilitation, midwifery, and dentistry). It is divided into 5 booklets: health data, cybersecurity, communication, digital tools in health care, and telehealth (5 Word documents, a total of 222 pages).

^eNot applicable.

To allow comparison, we provided the same sets of digital health competencies to all 3 configurations, following the process listed in [Textbox 1](#).

Textbox 1. Digital health competencies for configurations.

- 8 digital health topics were first selected: 5 were single-theme (“Health data, Communication, Cybersecurity, Digital tools, Telehealth”), and 3 were multiple-theme (“Communication and Digital tools”; “Communication, Digital tools, and Cybersecurity”; and “Telehealth, Health data, and Communication”).
- For each selected topic (eg, cybersecurity), 3 digital health competencies were then provided to each configuration (eg, responding to a cyberattack). Each configuration then generated an OSCE station related to these predefined competencies.
- In total, 24 OSCE stations were generated (8 topics × 3 configurations), all on the same day, June 1, 2025, during a single 4-hour session.
- Each generated OSCE was in French and included 4 components:
 - The primary and secondary learning domains,
 - The case vignette for the student,
 - The standardized observation checklist for examiners,
 - The script for standardized patients (SP script).

Description of the 3 Configurations Used to Generate OSCE Stations

Configuration 1: Standard GPT

This configuration aimed to assess ChatGPT’s ability to generate digital health OSCE stations in a minimal configuration, involving a simple prompt and the OSCE Vadémécum. It simulated a minimalist and spontaneous use case, akin to an educator working without technical assistance or a reference book.

The configuration description is as follows: (1) Model type: the free version of ChatGPT 4o [[55](#)]. (2) Prompt design: a simple prompt relied on a role-prompting strategy. The model was instructed to act as a “digital health expert and instructional designer” and create an OSCE based on 3 given digital health

competencies. It was invited to review the OSCE Vadémécum to choose learning domains, and then produced the OSCE’s components. No sequential structure or examples were provided, making this a basic, minimalist prompt ([Multimedia Appendix 2](#)). (3) Documents used: only the OSCE Vadémécum (in PDF format) was provided and served as the reference framework for structuring the OSCE station. (4) Technical configuration: ChatGPT interface, without customization and advanced features.

Configuration 2: Personalized GPT

This configuration aimed to assess whether providing a domain-specific knowledge base (here, the reference book in digital health) could improve the quality of the generated OSCE stations. It simulated the use of a personalized GPT by an educator equipped with a reference book.

The configuration description is as follows: (1) Model type: the paid version custom GPT-4o (cGPT) [56], which allows configuring GPT with a specific role using tailored instructions, example queries, and integrated knowledge bases. (2) Prompt design: the simple prompt was similar to Configuration 1, but added an instruction to review the UPCité reference book on digital health ([Multimedia Appendix 2](#)). (3) Documents used: both the OSCE Vademecum (in PDF format) and UPCité reference book (in Word format) were provided to cGPT. To optimize readability for the model, the UPCité digital health reference book was preprocessed by removing noneducational pages, simplifying formatting, and converting tables and figures to plain text. (4) Technical configuration:

1. assigning an icon and name to the cGPT
2. inserting instructions in the form of a prompt
3. adding a conversation starter with a user query, here: “I would like to design an OSCE station for assessment in digital health”
4. integrating a knowledge base, here the OSCE Vademecum and UPCité reference book.

None of the optional Custom GPT features (eg, web browsing, code interpreter, data analysis, and image generation) was used.

Configuration 3: Simulated-Agents GPT

This configuration aimed to assess whether a rigorous procedural structure, via the simulation of step-by-step simulated agents' reasoning, could improve the quality of the generated OSCE stations. It simulated the use of specialized assistants at each stage of OSCE construction by an educator equipped with a reference book.

Configuration description: (1) Model type: the same as in Configuration 2 (ie, the paid version of cGPT). (2) Prompt design: the prompt was built from instructions extracted from the OSCE Vademecum and structured to simulate a multiagent system. It followed a supervised, sequential, and specialized architecture, functioning like a processing chain in which each agent had a distinct pedagogical role in OSCE generation. The supervisor agent acted as the central coordinator. It collected the 3 competencies provided by the educator, determined the sequence of agent activation, passed contextual information between them, and ensured the overall coherence, without interfering with pedagogical content. After each step, it asked, “Would you like me to proceed to the next agent?” simulating a controlled, step-by-step process.

The following specialized agents were used: (1) The learning domain agent selected a primary and secondary domain from the predefined list included in the OSCE Vademecum, aligned with the learning objectives of the graduate medical curriculum (eg, “Education and Prevention”). (2) The vignette agent generated the case vignette for students, conforming to standardized formatting. (3) The checklist agent created the observation checklist for examiners, following the standardized format. (4) The SP script agent drafted the script for the standardized patient, following the standardized format.

Agents were activated sequentially, each receiving the outputs of the previous steps, and generated a formatted output

according to OSCE guidelines and UPCité reference book ([Multimedia Appendix 2](#)).

The prompt design combined multiple prompting strategies [48,57]:

1. Role prompting by assigning a specific pedagogical role to each agent;
2. Instruction-based prompting by providing precise, structured instructions drawn from the OSCE Vademecum;
3. Chain prompting by structuring tasks into a sequenced, logical workflow;
4. Few-shot prompting by including illustrations of expected formats to improve consistency and reproducibility.

(3) Documents used: the UPCité reference book on digital health was provided in the cGPT's knowledge base. (4) Technical configuration: similar setup to Configuration 2 (provision of a name, icon, and prompt, same conversation starter, knowledge base, and no optional features).

Evaluation of the Quality of OSCE Stations Generated by the 3 Configurations

To compare the 3 configurations, OSCE stations were generated and assessed by experts. The evaluation was conducted blindly of the configuration used, with a comprehensive assessment grid based on a literature review [27-32,58-62].

Evaluation of the Format Compliance of the Generated OSCEs

A fifth-year medical student, with a background in computer science and trained in OSCE methodology, measured how well the generated OSCE adhered to the OSCE Vademecum in terms of format compliance.

The compliance checklist included 27 criteria covering (1) the validity of the learning domains, (2) the vignette format, (3) the structure of the observation checklist, and (4) the completeness of the SP script. Each criterion was assessed using a binary scale (Yes or No), focusing only on format (independently of pedagogical quality).

Evaluation of the Educational Quality of the Generated OSCEs

A duo of digital health experts, involving the head of the medical informatics ward of “Hôpital Européen Georges Pompidou and Hôpital Necker Enfants Malades AP-HP, and a resident in digital health with a computer science background, conducted the evaluation of the content quality, independently and blindly of the configuration used (each OSCE station was anonymized and randomized).

For each OSCE, 9 evaluation dimensions were assessed:

1. “Relevance of the learning domains,” “Clarity,” “Pedagogical validity,” “Realism,” “Feasibility,” “Educational value,” and “Originality,” which were rated using a 5-point Likert scale of agreement (1: strongly disagree, 2: somewhat disagree, 3: neither agree nor disagree, 4: somewhat agree, and 5: strongly agree).

2. “Information accuracy”, which was rated on a severity scale (1: major issue compromising usability, 2: major issue, 3: moderate issue, 4: minor issue, and 5: no issues identified).
3. “Overall usability,” which was rated on a usability scale (1: not usable, 2: usable with major revisions, 3: usable with minor revisions, and 4: usable as is, without modification).

The criteria used were derived from the literature [27-32,58-61] and from the international OSCE guidelines [62], in order to align with international standards.

At the end of the evaluation process, OSCEs were grouped by similar topic (eg, Cybersecurity). For each group of 3 OSCEs (one per configuration), evaluators were asked to rank them, from first to third, without knowing which configuration had generated each one.

Statistical Analysis

For format-related criteria, binary variables (Yes or No) were described using frequencies and the percentage of “Yes” responses.

For educational criteria, expert ratings were averaged per OSCE and summarized using means and SD values (n=8 per configuration). Overall usability and ranking were categorical variables, analyzed without averaging expert ratings (n=16 per configuration), and described using frequencies and percentages. Due to the nonparametric distribution of data and the ordinal nature of variables, the Kruskal-Wallis test was used to compare the 3 configurations. For variables comprising multiple sub-criteria, an overall mean score per dimension was calculated and used in the Kruskal-Wallis test to determine statistical significance.

All statistical analyses were performed using the R statistical environment (R version 4.3.3) with RStudio GUI (version

2023.12.1+402; Posit PBC). The threshold for statistical significance was set at $P<.05$. Given the exploratory nature of the study, no correction for multiple testing was performed, with an emphasis on observed trends and effect sizes.

Ethical Considerations

This study did not involve human subjects. All generated outputs were fictional and intended for educational use. Therefore, in accordance with French Law n°2012-300 [63], no Institutional Review Board approval was required.

Results

Format Compliance of the OSCEs Generated

Overall, the simulated-agents GPT showed the highest format compliance, followed by the standard GPT, while the personalized GPT exhibited the lowest performance (Table 2).

Standard GPT OSCEs were fully compliant for the “learning domain” and “vignette” components. The “checklist” showed less compliance (between 3/8, 38% and 4/8, 50%) for the advanced criteria (“distinct items”, “assessment of skill only”, and “clear validation criteria for each item”). The “SP script” component was incomplete for the contextual clinical data of all generated cases, with information, such as symptoms, medications, and socioprofessional background, never present.

Personalized GPT OSCEs were poorly compliant for all components. The compliance was full only for 3 criteria of the “vignette” component (eg, “tasks to perform”).

Simulated-agents GPT OSCEs were fully compliant for the “learning domain,” “vignette,” and “SP script” components. Only 4 items of the “checklist” (“observable items” and the 3 advanced criteria) were partially compliant, ranging from 12% (1/8) to 75% (6/8).

Table 2. Evaluation of the format compliance of the Objective Structured Clinical Examination generated by the configuration type.

OSCE ^a components and criteria	Standard GPT (n=8), n (%)	Personalized GPT (n=8), n (%)	Simulated-agents GPT (n=8), n (%)
Domain			
The selected primary learning domain is included in the list of the 11 domains related to the OSCE assessment	8 (100)	0 (0)	8 (100)
The selected secondary learning domain is included in the list of the 11 domains related to the OSCE assessment	8 (100)	0 (0)	8 (100)
Vignette			
States the candidate's role	8 (100)	6 (75)	8 (100)
Describes the setting	8 (100)	8 (100)	8 (100)
Introduces the SP's identity (name, gender, and age)	8 (100)	6 (75)	8 (100)
Presents the issue or reason for consultation	8 (100)	8 (100)	8 (100)
Provides tasks to be performed by the student	8 (100)	8 (100)	8 (100)
Specifies actions the student should not take	8 (100)	0 (0)	8 (100)
Checklist			
Contains 10 to 15 items	8 (100)	2 (25)	8 (100)
Each item begins with an action verb	8 (100)	5 (62)	8 (100)
Items are observable	7 (88)	6 (75)	6 (75)
Items are dichotomous (Yes or No)	8 (100)	5 (62)	8 (100)
Items are distinct (or grouped if necessary, with precise scoring instructions)	3 (38)	1 (12)	1 (12)
Assesses only skills (not attitudes or communication)	4 (50)	1 (12)	2 (25)
Provides clear validation criteria for each item	3 (38)	1 (12)	3 (38)
SP script components included			
Scenario summary	4 (50)	4 (50)	8 (100)
SP's mindset or behavior	6 (75)	6 (75)	8 (100)
Additional data	0 (0)	1 (12)	8 (100)
Opening sentence	2 (25)	0 (0)	8 (100)
Identity	7 (88)	6 (75)	8 (100)
Socioprofessional background – Hobbies	0 (0)	2 (25)	8 (100)
Personal medical history	1 (12)	1 (12)	8 (100)
Familial medical history	0 (0)	0 (0)	8 (100)
Current medications	0 (0)	0 (0)	8 (100)
Symptoms	0 (0)	0 (0)	8 (100)
Conditional disclosure of information	8 (100)	7 (88)	8 (100)
Answers to all the items present in the "checklist" are provided	6 (75)	2 (25)	8 (100)

^aOSCE: Objective Structured Clinical Examination.

Educational Quality of the OSCEs Generated***Learning Domain Relevance***

Primary domain relevance was satisfactory for all configurations, highest for the simulated-agents GPT (mean

4.38, SD 1.38). Secondary domain relevance was more variable, with the standard GPT scoring lowest (mean 2.62, SD 1.36). Differences were not statistically significant ($P=.16$; [Tables 3](#) and [4](#); [Figure 1](#)).

Table 3. Evaluation of the content quality of the generated Objective Structured Clinical Examinations, by configuration type.

Evaluation domain, items, and section	Standard GPT (n=8), mean (SD)	Personalized GPT (n=8), mean (SD)	Simulated-agents GPT (n=8), mean (SD)
Learning domain relevance^a			
Primary domain	3.50 (1.13)	3.94 (1.82)	4.38 (1.38)
Secondary domain	2.62 (1.36)	3.56 (1.76)	3.56 (1.18)
Information accuracy^b			
No errors or inexact information			
Vignette	3.88 (0.88)	4.38 (0.88)	4.75 (0.46)
Checklist	4.06 (0.86)	4.19 (0.75)	4.88 (0.35)
SP ^c script	4.25 (0.85)	3.62 (0.69)	4.19 (0.92)
No missing information			
Vignette	4.19 (1.13)	4 (0.60)	4.81 (0.37)
Checklist	4.06 (0.82)	3.88 (0.52)	4.62 (0.44)
SP script	3.94 (1.02)	3.44 (0.86)	4.19 (0.92)
No irrelevant or unnecessary information			
Vignette	4.56 (0.73)	4.62 (0.58)	4.69 (0.37)
Checklist	4 (1)	4.31 (0.65)	4.81 (0.26)
SP script	4.62 (0.69)	4.38 (0.88)	4.56 (0.73)
No direct cues, leading formulations, or implicit answers			
Vignette	3 (1.51)	2.38 (1.75)	2.69 (1.19)
SP script	4.56 (0.82)	4.88 (0.35)	5 (0)
Clarity^a			
Clear and comprehensible writing			
Vignette	3 (0.96)	3.62 (1.38)	4.44 (0.78)
Checklist	3.75 (0.65)	3.62 (0.64)	4.56 (0.56)
SP script	3 (0.85)	3.56 (0.98)	4.38 (0.88)
Pedagogical validity^a			
Alignment with the competencies to be assessed			
All	3.88 (0.64)	4.06 (0.82)	4.56 (0.42)
Alignment with the expectations of a professional trained in digital health			
All	4.12 (0.52)	4.44 (0.68)	4.50 (0.46)
Realism^a			
Realistic situation in a medical professional's practice			
All	3.62 (1.13)	3.25 (0.89)	4.12 (0.69)
Feasibility^a			
Ease of implementation			
All	4.19 (0.80)	3.56 (1.35)	4.06 (0.98)
Feasible within an 8-minute timeframe			
All	4.19 (0.70)	4.06 (0.78)	4 (0.46)
Ease of recruiting an SP			
All	4.25 (0.85)	4.62 (0.88)	4.69 (0.59)
Educational value^a			

Evaluation domain, items, and section	Standard GPT (n=8), mean (SD)	Personalized GPT (n=8), mean (SD)	Simulated-agents GPT (n=8), mean (SD)
Useful for promoting digital health learning			
All	3.62 (0.74)	3.94 (0.86)	4.12 (0.88)
Encourages analysis, reflection, and decision-making			
All	3.62 (0.88)	3.81 (1.07)	3.44 (0.73)
Originality^a			
Creative potential			
All	4.06 (0.50)	3.88 (0.23)	3.50 (0.76)

^aCriteria assessed with a Likert scale from 1: strongly disagree to 5: strongly agree.

^bCriteria assessed with a severity scale from 1: major issues to 5: no issues.

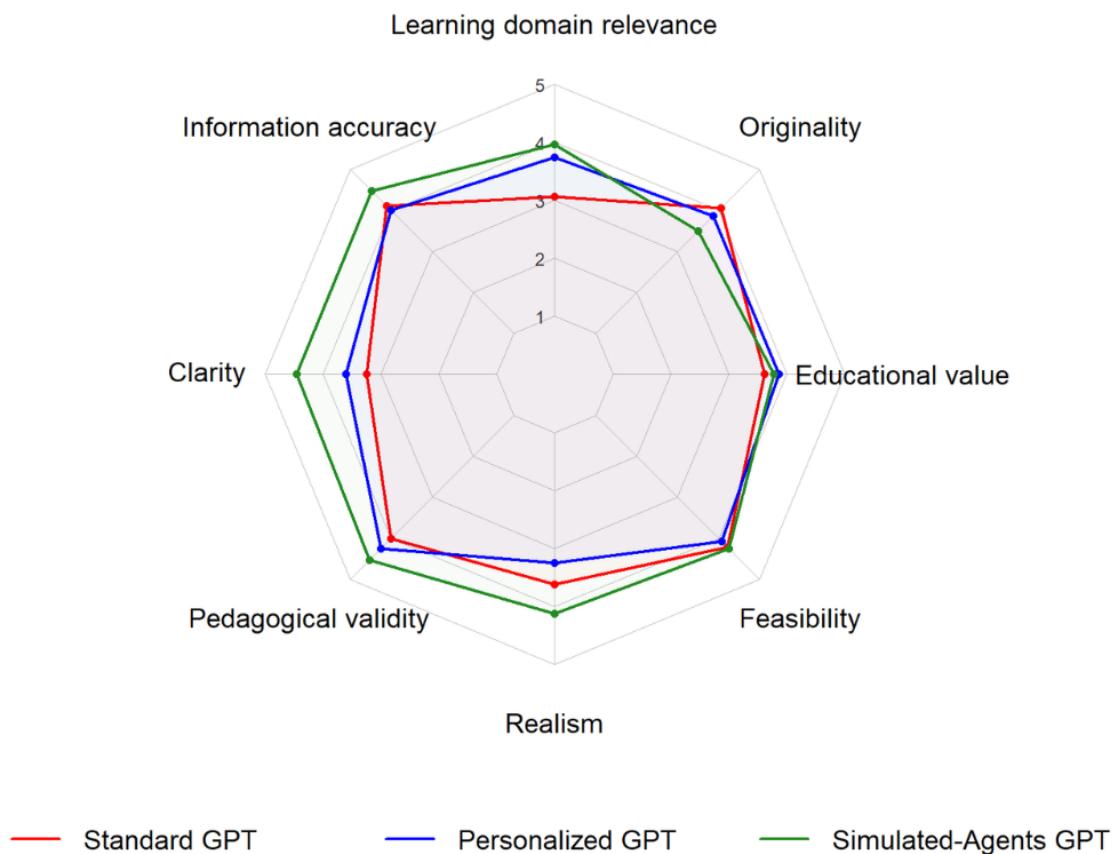
^cSP: standardized participant.

Table 4. Kruskal-Wallis test comparisons of the 3 GPT configurations by evaluation domain. To perform the statistical test, the mean score was calculated for each evaluation domain and configuration. Test results are presented as P values from the Kruskal-Wallis test.

Evaluation domain	Standard GPT, mean (SD)	Personalized GPT, mean (SD)	Simulated-agents GPT, mean (SD)	P value ^a
Learning domain relevance	3.06 (1.02)	3.75 (1.75)	3.97 (1.06)	.16
Information accuracy	4.10 (0.36)	4.01 (0.21)	4.47 (0.28)	.01 ^a
Clarity	3.25 (0.45)	3.60 (0.83)	4.46 (0.52)	.004 ^a
Pedagogical validity	4 (0.5)	4.25 (0.65)	4.53 (0.34)	.12
Realism	3.63 (1.13)	3.25 (0.89)	4.13 (0.70)	.19
Feasibility	4.21 (0.70)	4.08 (0.80)	4.25 (0.51)	.92
Educational value	3.63 (0.65)	3.88 (0.93)	3.78 (0.76)	.75
Originality	4.06 (0.49)	3.88 (0.23)	3.50 (0.76)	.32

^aP<.05 (statistically significant).

Figure 1. Radar plot of evaluation scores (5-point scales) across the 3 GPT configurations. Agreement-based dimensions (eg, clarity, realism) were rated from 1 (strongly disagree) to 5 (strongly agree); information accuracy used a severity scale from 1 (major issues) to 5 (no issues). The simulated-agents GPT tended to score higher in information accuracy, clarity, realism, pedagogical validity, and learning domain relevance, while it lagged in terms of originality. All 3 configurations showed similar performance for feasibility and educational value.



Information Accuracy

Overall, there was a significant difference in information accuracy ($P=.01$), with the simulated-agents GPT providing the most accurate and comprehensive content (Tables 3 and 4; Figure 1). Specifically,

1. “Errors or inaccurate information” were low across configurations (mean>3.88), with a favorable trend toward the simulated-agents GPT, particularly in the “vignette” (mean 4.75, SD 0.46) and the “checklist” (mean 4.88, SD 0.35) components.
2. “Missing information” was rare in the simulated-agents GPT (mean \geq 4.19 across all sections), but more common in the personalized GPT, especially in the “script” (mean 3.44, SD 0.86).
3. “Irrelevant or unnecessary information” was rarely noted, with similarly high averages across configurations (mean \geq 4).
4. “The presence of direct cues or leading formulations” was frequent in the “vignette” (2.38< mean< 3), but nearly absent in the “script”, especially for the simulated-agents GPT (mean 5, SD 0).

Clarity

Clarity was significantly different among the 3 configurations ($P=.004$), with a trend in favor of the simulated-agents GPT,

getting better results for all components (mean>4.38). The personalized GPT showed intermediate results, while the standard GPT scored lower, particularly for the “vignette” and the “script” (mean 3 for both, SD 0.96 and 0.85, respectively; Tables 3 and 4; Figure 1).

Pedagogical Validity

Pedagogical validity was high across all configurations (mean>3.88). Simulated-agents GPT scored highest (mean>4.5), followed by the personalized GPT (mean>4.06), and then the standard GPT (mean>3.88). Although differences were not statistically significant ($P=.12$), a favorable trend for simulated-agents GPT emerged (Tables 3 and 4; Figure 1).

Realism and Feasibility

OSCE’s realism was satisfactory across configurations, with no statistically significant difference ($P=.19$). The simulated-agents GPT scored the highest (mean 4.12, SD 0.69), followed by the standard GPT (mean 3.62, SD 1.13) and the personalized GPT (mean 3.25, SD 0.89).

No significant difference was observed for the feasibility ($P=.92$). The standard GPT tended to propose OSCEs that were easiest to implement and more feasible to perform within an 8-minute timeframe (mean 4.19, SD 0.70). Ease of recruiting an SP was rated high for all, with a slight advantage for the

simulated-agents GPT (mean 4.69, SD 0.59; **Tables 3** and **4**; **Figure 1**).

Educational Value and Originality

Educational value was comparable across configurations, with no statistically significant differences ($P=.75$). The simulated-agents GPT had the highest score for the use to promote digital health learning (mean 4.12, SD 0.88), while it had a slightly lower score than personalized GPT for encouraging analysis and decision-making (mean 3.44 SD 0.73 vs mean 3.81 SD 1.07 for personalized GPT; **Tables 3** and **4**; **Figure 1**).

Regarding originality, there was no significant difference observed ($P=.32$). The standard GPT had the highest average (mean 4.06, SD 0.50), followed by the personalized GPT (mean 3.88, SD 0.23) and the simulated-agents GPT (mean 3.50, SD 0.76; **Tables 3** and **4**; **Figure 1**).

Table 5. Overall assessment of the OSCEs generated by the type of configuration. Test results are presented as P values from the Kruskal-Wallis test.

Assessment domain and items	Standard GPT (n=16), n (%)	Personalized GPT (n=16), n (%)	Simulated-agents GPT (n=16), n (%)	P value ^a
Usability				.02 ^a
Not usable	2 (12)	2 (12)	1 (6)	
Usable with major revisions	5 (31)	5 (31)	1 (6)	
Usable with minor revisions	7 (44)	6 (38)	9 (56)	
Usable as is, without modification	2 (12)	3 (19)	5 (31)	
Ranking				<.001 ^a
First	0 (0)	2 (12)	14 (88)	
Second	7 (44)	7 (44)	2 (12)	
Third	9 (56)	7 (44)	0 (0)	

^a $P<.05$ (statistically significant).

Qualitative Feedback

In addition to quantitative ratings, qualitative feedback from experts revealed consistent issues across configurations. All tended to include excessive vignette details, often with leading cues that overly guided students (eg, “Identify and articulate the main risks associated with the use of telemedicine (technical, organizational, legal)”). Inconsistencies were noted between the vignette and SP script, with details missing or contradictory (eg, in the vignette: “Mrs. D [...] has also received a message on a secure messaging platform”; in the script: “You do not own a smartphone [...] You do not have a computer at home.”). Finally, even when highly rated, OSCEs were often considered too theoretical, likely due to reliance on the UPCité digital health booklet, reducing their practical, hands-on applicability for OSCEs.

Discussion

Principal Findings

This study compared 3 configurations of GPT for generating OSCE stations in the field of digital health: the standard GPT (simple prompt, OSCE Vademeum); the personalized GPT

Overall Assessment

Overall, usability differed significantly across configurations ($P=.02$). For simulated-agents GPT, 88% (14/16) of OSCEs were usable without major revisions, while for standard and personalized GPT, 31% (5/16) of OSCEs required major revisions, and 12% (2/16) were unusable (**Table 5**).

Overall, generated case rankings differed significantly as well ($P<.001$). The simulated-agents GPT stood out, with 88% (14/16) of OSCEs ranked first and none ranked third. Conversely, the standard GPT ranked third (9/16, 56%) or second (7/16, 44%) but never ranked first. The personalized GPT received a few first-place rankings (2/16, 12%) and was mainly split between second and third places (7/16, 44% each; **Table 5**).

(simple prompt, OSCE Vademeum, a reference book in digital health), and the simulated-agents GPT (a prompt simulating a fictional multiagent system specialized in OSCEs, a reference book in digital health). Regarding format, the simulated-agents GPT achieved the highest compliance, with minor gaps in advanced checklist criteria. The standard GPT ranked second, with weaker checklist performance and consistently incomplete SP scripts, while the personalized GPT showed the lowest compliance across all components. Regarding educational content, the simulated-agents GPT achieved significantly higher ratings for information accuracy ($P=.01$) and clarity ($P=.004$). It also outperformed in educational validity and realism, although these differences weren't statistically significant. Additionally, it also had significantly better usability, with 88% (14/16) of the OSCEs usable ($P=.02$). Overall, the simulated-agents GPT significantly outperformed the others, ranking first in 88% (14/16) of the OSCEs and never placing third.

Interpretation of Results

Differences between the 3 configurations may be explained by several reasons. Regarding format compliance, the differences might reflect how each configuration processed and prioritized

information. The standard GPT, which used only the OSCE Vadémécum, followed the guidelines more closely than the personalized GPT that also included the UPCité reference book. This may be explained by cognitive overload, as adding an extra knowledge source increased retrieval complexity and reduced effective prioritization of key instructions [64]. These issues are amplified by the reliance on a proprietary Retrieval-Augmented Generation (RAG) system that offers only limited transparency and user control [65,66]. The simulated-agents GPT, which used a structured prompt with explicit OSCE guidelines, appeared to improve format adherence. Although it did not fully meet all checklist-format requirements, this suggests that carefully designed prompts are crucial for ensuring format accuracy.

Regarding educational content. The simulated-agents GPT, despite the 8000-character limit, produced high-quality outputs, showing that structured, modular prompts work well even with technical constraints [67]. Interestingly, the standard and personalized GPTs were perceived as more original and more effective at fostering analytical thinking. This suggests a trade-off: while structured approaches improve consistency and adherence to standards, more flexible, broader configurations can boost creativity and originality, albeit with reduced accuracy and lower consistency and control.

Finally, these findings should be interpreted with caution because technology is rapidly evolving [68,69]. The evaluation was based on a dataset generated at a specific time, and since generative models can produce different outputs with the same prompts, reproducibility may be affected. Additionally, the algorithms powering systems like OpenAI's GPT are constantly improving, which could impact future results. However, these considerations do not undermine the validity of our findings. The main goal of the study was to compare 3 OSCE generation strategies, offering insights into how generative models can support medical education rather than advocating for a fixed implementation.

Limitations

Regarding the LLM used. We focused on ChatGPT, whereas other LLMs have already been used to generate educational resources. For example, Google Bard (now called Gemini Google DeepMind) and Microsoft Bing were used to create multiple-choice questions in medical physiology [70], and LLaMA (Meta AI) was used in radiology [71]. However, prior to designing this study, we conducted preliminary tests with open-source models such as Mixtral-8x7B (Mistral AI) and Llama-3.3-70B combined with custom RAG pipelines. These initial tests were quite disappointing as they had suboptimal performances, particularly for OSCE-style tasks. Notably, recent comparative studies confirm GPT-4's superior performance on RAG-enhanced clinical reasoning tasks [72,73], which reinforces our choice of model despite the broader range of available LLMs.

Regarding the evaluation. It was conducted by 2 experts, which may introduce subjective bias and limit the diversity of evaluative perspectives. However, this limitation is mitigated by their good expertise in digital health and the rigorous evaluation process. A total of 24 OSCEs were reviewed,

covering many topics, providing a diverse and valuable dataset. Each OSCE was evaluated with detailed criteria for both format and content, making the results more reliable. A key strength of this study is its strict use of the METRICS checklist [54], a recent guide for studies involving generative models in medical education. This makes the results clearer, easier to reproduce, and comparable. Additionally, while the focus on digital health allowed for targeted exploration in an emerging educational field, it may limit the generalizability of the results to other medical specialties. However, since digital health is growing across many areas, the insight from this work, especially about prompt design and model behavior, is likely useful in other competency-based assessments too.

Comparison With Other Studies

In the context of OSCEs, other studies used LLMs for tasks other than the generation of OSCE stations:

1. To simulate standardized patients [36-38]. For example, Yamamoto et al [36] used artificial intelligence (AI)-powered chatbots to simulate standardized patients, helping students improve their clinical interview skills. Students working with the chatbots scored significantly higher than those who did not.
2. To automate the OSCE assessment [39-43]. For example, Jamieson et al [41] developed an AI-based grading system to grade post-encounter notes, matching expert scores with 90% agreement while reducing manual grading effort by 91%.
3. To simulate OSCE candidates [44-46]. For instance, Huang et al [44] found that ChatGPT-4.0 performed as well as or better than junior emergency residents in history-taking and record-writing, though humans still demonstrated higher overall consultation quality.

Beyond the context of OSCEs, LLMs have been used to create clinical cases or vignettes for medical education. Like the simulated-agents GPT, studies showed high clarity of the generated content. For example, Coşkun et al [59] reported a 4.11/5 rating for the comprehensibility of the generated scenarios. Additionally, similar to our results, studies also showed strong usability, with Yanagita et al [29] indicating a 97% usability rate for GPT-generated general medicine cases. However, issues with “accuracy” and “missing information” were also reported. For example, Takahashi et al [28] reported 68% accuracy, Yanagita et al [29] reported only 58%, and Scherr et al [30] achieved 100% accuracy, but their evaluation was limited to a single and narrow area (acute asthma). Vaughn et al [31] found that 88% of nursing simulations lacked critical clinical details. In contrast, our simulated-agents GPT consistently yielded complete and context-rich OSCE stations. However, these comparisons should be made carefully, as methodology, design, and metrics differ considerably between studies.

In summary, previous works have largely focused on the use of LLMs to simulate patients, students, or evaluators, or to generate isolated clinical cases. In contrast, to our knowledge, our study is the first to use LLMs to generate OSCE stations, including the vignette, checklist, and SP script.

Future Directions

This work is part of the DigiHealth Paris Cité project [74], which aims to integrate digital health training into health curricula through innovative methods such as simulation and immersive learning. Here, the objective was to support educators in designing reliable, standardized OSCE stations for student assessment. Based on our findings, the simulated-agents GPT configuration appeared best suited as an assistant for educators, helping to quickly create structured OSCEs that still require validation before introduction into official evaluation sessions.

Regarding technical and pedagogical improvements. The simulated-agents GPT configuration could be further optimized through the refinement of the prompting strategy to enhance output quality and alignment with educational objectives. One effective approach will be iterative development, where expert feedback is continuously integrated into the LLM's generation cycle. For instance, Coşkun et al [59] used expert prompts to revise and improve 15 clinical cases directly in ChatGPT without manual editing. Similarly, Yanagita et al [29] demonstrated that feeding expert-reviewed vignettes back into the system can help guide and standardize future generations. Therefore, our next steps will include developing an interactive OSCE-generation assistant for educators using the simulated-agents GPT framework, integrating continuous expert feedback to improve content quality, and creating an internal repository of validated AI-generated OSCEs for educators at our faculty, promoting transparency, reproducibility, and shared pedagogical development.

Acknowledgments

This work was supported by a French government grant managed by the Agence Nationale de la Recherche under the France 2030 program, reference DigiHealth Paris Cité ANR-23-CMAS-0034 (coordinator: RT).

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Data Availability

The prompts used to generate OSCEs are available in [Multimedia Appendix 2](#). The OSCEs generated during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

METRICS checklist.

[[DOCX File , 16 KB - mededu_v12i1e82116_app1.docx](#)]

Multimedia Appendix 2

Prompts used to generate the OSCEs for each configuration.

[[DOCX File , 21 KB - mededu_v12i1e82116_app2.docx](#)]

References

1. Preiksaitis C, Rose C. Opportunities, Challenges, and Future Directions of Generative Artificial Intelligence in Medical Education: Scoping Review. JMIR Med Educ 2023 Oct 20;9:e48785 [FREE Full text] [doi: [10.2196/48785](https://doi.org/10.2196/48785)] [Medline: [37862079](https://pubmed.ncbi.nlm.nih.gov/37862079/)]
2. Aster A, Laupichler MC, Rockwell-Kollmann T, Masala G, Bala E, Raupach T. ChatGPT and other large language models in medical education - scoping literature review. Med Sci Educ 2025;35(1):555-567. [doi: [10.1007/s40670-024-02206-6](https://doi.org/10.1007/s40670-024-02206-6)] [Medline: [40144083](https://pubmed.ncbi.nlm.nih.gov/40144083/)]
3. Gordon M, Daniel M, Ajiboye A, Uraiby H, Xu N, Bartlett R, et al. A scoping review of artificial intelligence in medical education: BEME guide No. 84. Med Teach 2024;46(4):446-470 [FREE Full text] [doi: [10.1080/0142159X.2024.2314198](https://doi.org/10.1080/0142159X.2024.2314198)] [Medline: [38423127](https://pubmed.ncbi.nlm.nih.gov/38423127/)]
4. Lucas HC, Upperman JS, Robinson JR. A systematic review of large language models and their implications in medical education. Med Educ 2024;58(11):1276-1285. [doi: [10.1111/medu.15402](https://doi.org/10.1111/medu.15402)] [Medline: [38639098](https://pubmed.ncbi.nlm.nih.gov/38639098/)]
5. 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;10:e52483 [FREE Full text] [doi: [10.2196/52483](https://doi.org/10.2196/52483)] [Medline: [38598263](https://pubmed.ncbi.nlm.nih.gov/38598263/)]
6. Zhui L, Yhap N, Liping L, Zhengjie W, Zhonghao X, Xiaoshu Y, et al. Impact of large language models on medical education and teaching adaptations. JMIR Med Inform 2024;12:e55933 [FREE Full text] [doi: [10.2196/55933](https://doi.org/10.2196/55933)] [Medline: [39087590](https://pubmed.ncbi.nlm.nih.gov/39087590/)]
7. Liu M, Okuhara T, Chang X, Shirabe R, Nishie Y, Okada H, et al. Performance of ChatGPT across different versions in medical licensing examinations worldwide: systematic review and meta-analysis. J Med Internet Res 2024;26:e60807 [FREE Full text] [doi: [10.2196/60807](https://doi.org/10.2196/60807)] [Medline: [39052324](https://pubmed.ncbi.nlm.nih.gov/39052324/)]
8. 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](https://doi.org/10.2196/45312)] [Medline: [36753318](https://pubmed.ncbi.nlm.nih.gov/36753318/)]
9. 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;2(2):e0000198 [FREE Full text] [doi: [10.1371/journal.pdig.0000198](https://doi.org/10.1371/journal.pdig.0000198)] [Medline: [36812645](https://pubmed.ncbi.nlm.nih.gov/36812645/)]
10. Nakao T, Miki S, Nakamura Y, Kikuchi T, Nomura Y, Hanaoka S, et al. Capability of GPT-4V(ision) in the Japanese national medical licensing examination: evaluation study. JMIR Med Educ 2024;10:e54393 [FREE Full text] [doi: [10.2196/54393](https://doi.org/10.2196/54393)] [Medline: [38470459](https://pubmed.ncbi.nlm.nih.gov/38470459/)]
11. Zong H, Wu R, Cha J, Wang J, Wu E, Li J, et al. Large language models in worldwide medical exams: platform development and comprehensive analysis. J Med Internet Res 2024;26:e66114. [doi: [10.2196/66114](https://doi.org/10.2196/66114)] [Medline: [39729356](https://pubmed.ncbi.nlm.nih.gov/39729356/)]
12. Jaleel A, Aziz U, Farid G, Zahid Bashir M, Mirza TR, Khizar Abbas SM, et al. Evaluating the potential and accuracy of ChatGPT-3.5 and 4.0 in medical licensing and in-training examinations: systematic review and meta-analysis. JMIR Med Educ 2025;11:e68070 [FREE Full text] [doi: [10.2196/68070](https://doi.org/10.2196/68070)] [Medline: [40973108](https://pubmed.ncbi.nlm.nih.gov/40973108/)]
13. Artsi Y, Sorin V, Konen E, Glicksberg BS, Nadkarni G, Klang E. Large language models for generating medical examinations: systematic review. BMC Med Educ 2024;24(1):354 [FREE Full text] [doi: [10.1186/s12909-024-05239-y](https://doi.org/10.1186/s12909-024-05239-y)] [Medline: [38553693](https://pubmed.ncbi.nlm.nih.gov/38553693/)]
14. Kiyak YS, Emekli E. ChatGPT prompts for generating multiple-choice questions in medical education and evidence on their validity: a literature review. Postgrad Med J 2024;100(1189):858-865. [doi: [10.1093/postmj/qgaa065](https://doi.org/10.1093/postmj/qgaa065)] [Medline: [38840505](https://pubmed.ncbi.nlm.nih.gov/38840505/)]
15. Cheung BHH, Lau GKK, Wong GTC, Lee EYP, Kulkarni D, Seow CS, 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 [FREE Full text] [doi: [10.1371/journal.pone.0290691](https://doi.org/10.1371/journal.pone.0290691)] [Medline: [37643186](https://pubmed.ncbi.nlm.nih.gov/37643186/)]
16. 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 [FREE Full text] [doi: [10.2196/69521](https://doi.org/10.2196/69521)] [Medline: [40446330](https://pubmed.ncbi.nlm.nih.gov/40446330/)]
17. Holderried F, Stegemann-Philipps C, Herschbach L, Moldt J, Nevins A, Griewatz J, 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;10:e53961 [FREE Full text] [doi: [10.2196/53961](https://doi.org/10.2196/53961)] [Medline: [38227363](https://pubmed.ncbi.nlm.nih.gov/38227363/)]
18. Holderried F, Stegemann-Philipps C, Herrmann-Werner A, Festl-Wietek T, Holderried M, Eickhoff C, et al. A language model-powered simulated patient with automated feedback for history taking: prospective study. JMIR Med Educ 2024;10:e59213 [FREE Full text] [doi: [10.2196/59213](https://doi.org/10.2196/59213)] [Medline: [39150749](https://pubmed.ncbi.nlm.nih.gov/39150749/)]
19. Öncü S, Torun F, Ülkü HH. AI-powered standardised patients: evaluating ChatGPT-4o's impact on clinical case management in intern physicians. BMC Med Educ 2025;25(1):278 [FREE Full text] [doi: [10.1186/s12909-025-06877-6](https://doi.org/10.1186/s12909-025-06877-6)] [Medline: [39979969](https://pubmed.ncbi.nlm.nih.gov/39979969/)]
20. Yu H, Zhou J, Li L, Chen S, Gallifant J, Shi A, et al. AIPatient: simulating patients with EHRs and LLM powered agentic workflow. : arXiv [FREE Full text] Preprint posted online on 27 Nov 2025

21. Wang C, Li S, Lin N, Zhang X, Han Y, Wang X. Application of large language models in medical training evaluation-using ChatGPT as a standardized patient: multimetric assessment. *J Med Internet Res* 2025;27:e59435 [FREE Full text] [doi: [10.2196/59435](https://doi.org/10.2196/59435)] [Medline: [39742453](#)]
22. 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;25:e43311 [FREE Full text] [doi: [10.2196/43311](https://doi.org/10.2196/43311)] [Medline: [37335593](#)]
23. Li Y, Zeng C, Zhong J, Zhang R, Zhang M, Zou L. Leveraging large language model as simulated patients for clinical education. : arXiv [FREE Full text] Preprint posted online on 25 Apr 2024
24. Liu Y, Shi C, Wu L, Lin X, Chen X, Zhu Y, et al. Development and validation of a large language model-based system for medical history-taking training: prospective multicase study on evaluation stability, human-ai consistency, and transparency. *JMIR Med Educ* 2025;11:e73419 [FREE Full text] [doi: [10.2196/73419](https://doi.org/10.2196/73419)] [Medline: [40882613](#)]
25. Cook DA, Overgaard J, Pankratz VS, Del Fiol G, Aakre CA. Virtual patients using large language models: scalable, contextualized simulation of clinician-patient dialogue with feedback. *J Med Internet Res* 2025;27:e68486. [doi: [10.2196/68486](https://doi.org/10.2196/68486)] [Medline: [39854611](#)]
26. Cross J, Kayalackakom T, Robinson RE, Vaughans A, Sebastian R, Hood R, et al. Assessing ChatGPT's capability as a new age standardized patient: qualitative study. *JMIR Med Educ* 2025;11:e63353 [FREE Full text] [doi: [10.2196/63353](https://doi.org/10.2196/63353)] [Medline: [40393017](#)]
27. Sridharan K, Sequeira RP. *BMC Med Educ* 2024 Apr 22;24(1):431 [FREE Full text] [doi: [10.1186/s12909-024-05365-7](https://doi.org/10.1186/s12909-024-05365-7)] [Medline: [38649959](#)]
28. Takahashi H, Shikino K, Kondo T, Komori A, Yamada Y, Saita M, et al. Educational utility of clinical vignettes generated in Japanese by chatgpt-4: mixed methods study. *JMIR Med Educ* 2024;10:e59133 [FREE Full text] [doi: [10.2196/59133](https://doi.org/10.2196/59133)] [Medline: [39137031](#)]
29. Yanagita Y, Yokokawa D, Uchida S, Li Y, Uehara T, Ikusaka M. Can AI-generated clinical vignettes in Japanese be used medically and linguistically? *J Gen Intern Med* 2024;39(16):3282-3289. [doi: [10.1007/s11606-024-09031-y](https://doi.org/10.1007/s11606-024-09031-y)] [Medline: [39313665](#)]
30. Scherr R, Spina A, Dao A, Andalib S, Halaseh FF, Blair S, et al. Novel evaluation metric and quantified performance of ChatGPT-4 patient management simulations for early clinical education: experimental study. *JMIR Form Res* 2025;9:e66478 [FREE Full text] [doi: [10.2196/66478](https://doi.org/10.2196/66478)] [Medline: [40013991](#)]
31. Vaughn J, Ford SH, Scott M, Jones C, Lewinski A. Enhancing healthcare education: leveraging chatgpt for innovative simulation scenarios. *Clinical Simulation in Nursing* 2024;87:101487. [doi: [10.1016/j.ecns.2023.101487](https://doi.org/10.1016/j.ecns.2023.101487)]
32. Violato E, Corbett C, Rose B, Rauschning B, Witschen B. The effectiveness and efficiency of using ChatGPT for writing health care simulations. *ijohs* 2023. [doi: [10.54531/wjgb5594](https://doi.org/10.54531/wjgb5594)]
33. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979;13(1):41-54. [Medline: [763183](#)]
34. Patrício MF, Julião M, Fareleira F, Carneiro AV. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? *Med Teach* 2013;35(6):503-514. [doi: [10.3109/0142159X.2013.774330](https://doi.org/10.3109/0142159X.2013.774330)] [Medline: [23521582](#)]
35. Majumder MAA, Kumar A, Krishnamurthy K, Ojeh N, Adams OP, Sa B. An evaluative study of objective structured clinical examination (OSCE): students and examiners perspectives. *Adv Med Educ Pract* 2019;10:387-397 [FREE Full text] [doi: [10.2147/AMEPS197275](https://doi.org/10.2147/AMEPS197275)] [Medline: [31239801](#)]
36. Yamamoto A, Koda M, Ogawa H, Miyoshi T, Maeda Y, Otsuka F, et al. Enhancing medical interview skills through ai-simulated patient interactions: nonrandomized controlled trial. *JMIR Med Educ* 2024;10:e58753 [FREE Full text] [doi: [10.2196/58753](https://doi.org/10.2196/58753)] [Medline: [39312284](#)]
37. Pereira DSM, Falcão F, Nunes A, Santos N, Costa P, Pêgo JM. Designing and building OSCEBot ® for virtual OSCE - performance evaluation. *Med Educ Online* 2023;28(1):2228550 [FREE Full text] [doi: [10.1080/10872981.2023.2228550](https://doi.org/10.1080/10872981.2023.2228550)] [Medline: [37347808](#)]
38. Hirosawa T, Yokose M, Sakamoto T, Harada Y, Tokumasu K, Mizuta K, et al. Utility of Generative Artificial Intelligence for Japanese Medical Interview Training: Randomized Crossover Pilot Study. *JMIR Med Educ* 2025 Aug 01;11:e77332 [FREE Full text] [doi: [10.2196/77332](https://doi.org/10.2196/77332)] [Medline: [40749190](#)]
39. Luordo D, Torres Arrese M, Tristán Calvo C, Shani Shani KD, Rodríguez Cruz LM, García Sánchez FJ, et al. Application of Artificial Intelligence as an Aid for the Correction of the Objective Structured Clinical Examination (OSCE). *Applied Sciences* 2025 Jan 23;15(3):1153. [doi: [10.3390/app15031153](https://doi.org/10.3390/app15031153)]
40. Geathers J, Hicke Y, Chan C, Rajashekhar N, Sewell J, Cornes S, et al. Benchmarking generative AI for scoring medical student interviews in objective structured clinical examinations (OSCEs). : ArXiv [FREE Full text] Preprint posted online on May 15, 2025
41. Jamieson AR, Holcomb MJ, Dalton TO, Campbell KK, Vedovato S, Shakur AH, et al. Rubrics to Prompts: Assessing Medical Student Post-Encounter Notes with AI. *NEJM AI* 2024 Nov 27;1(12). [doi: [10.1056/aics2400631](https://doi.org/10.1056/aics2400631)]
42. Shakur AH, Holcomb MJ, Hein D, Kang S, Dalton TO, Campbell KK, et al. Large language models for medical OSCE assessment: a novel approach to transcript analysis. : ArXiv [FREE Full text] Preprint posted online on 11 Oct 2024

43. Tekin M, Yurdal MO, Toraman Ç, Korkmaz G, Uysal İ. Is AI the future of evaluation in medical education?? AI vs. human evaluation in objective structured clinical examination. *BMC Med Educ* 2025;25(1):641 [FREE Full text] [doi: [10.1186/s12909-025-07241-4](https://doi.org/10.1186/s12909-025-07241-4)] [Medline: [40312328](#)]

44. Huang T, Hsieh PH, Chang Y. Performance comparison of junior residents and ChatGPT in the objective structured clinical examination (OSCE) for medical history taking and documentation of medical records: development and usability study. *JMIR Med Educ* 2024;10:e59902 [FREE Full text] [doi: [10.2196/59902](https://doi.org/10.2196/59902)] [Medline: [39622713](#)]

45. Li SW, Kemp MW, Logan SJS, Dimri PS, Singh N, Mattar CNZ, et al. ChatGPT outscored human candidates in a virtual objective structured clinical examination in obstetrics and gynecology. *Am J Obstet Gynecol* 2023;229(2):172.e1-172.e12 [FREE Full text] [doi: [10.1016/j.ajog.2023.04.020](https://doi.org/10.1016/j.ajog.2023.04.020)] [Medline: [37088277](#)]

46. Aqvabil-Jahromi S, Eftekhari M, Akbari H, Aligholi-Zahraie M. Evaluation of correctness and reliability of GPT, bard, and bing chatbots' responses in basic life support scenarios. *Sci Rep* 2025;15(1):11429 [FREE Full text] [doi: [10.1038/s41598-024-82948-w](https://doi.org/10.1038/s41598-024-82948-w)] [Medline: [40180985](#)]

47. Tsang R. Practical applications of ChatGPT in undergraduate medical education. *J Med Educ Curric Dev* 2023;10:23821205231178449 [FREE Full text] [doi: [10.1177/23821205231178449](https://doi.org/10.1177/23821205231178449)] [Medline: [37255525](#)]

48. Maaz S, Palaganas JC, Palaganas G, Bajwa M. A guide to prompt design: foundations and applications for healthcare simulationists. *Front Med (Lausanne)* 2024;11:1504532 [FREE Full text] [doi: [10.3389/fmed.2024.1504532](https://doi.org/10.3389/fmed.2024.1504532)] [Medline: [39980724](#)]

49. Misra SM, Suresh S. Artificial intelligence and objective structured clinical examinations: using ChatGPT to revolutionize clinical skills assessment in medical education. *J Med Educ Curric Dev* 2024;11:23821205241263475 [FREE Full text] [doi: [10.1177/23821205241263475](https://doi.org/10.1177/23821205241263475)] [Medline: [39070287](#)]

50. Bearman M, Ajjawi R, Bennett S, Boud D. The hidden labours of designing the objective structured clinical examination: a practice theory study. *Adv Health Sci Educ Theory Pract* 2021;26(2):637-651. [doi: [10.1007/s10459-020-10015-w](https://doi.org/10.1007/s10459-020-10015-w)] [Medline: [33196956](#)]

51. Khan KZ, Ramachandran S, Gaunt K, Pushkar P. The objective structured clinical examination (OSCE): AMEE Guide No. 81. Part I: an historical and theoretical perspective. *Med Teach* 2013;35(9):e1437-e1446. [doi: [10.3109/0142159X.2013.818634](https://doi.org/10.3109/0142159X.2013.818634)] [Medline: [23968323](#)]

52. Armijo-Rivera S, Fuenzalida-Muñoz B, Vicencio-Clarke S, Elbers-Arce A, Bozzo-Navarrete S, Kunakov N, et al. Advancing the assessment of clinical competence in Latin America: a scoping review of OSCE implementation and challenges in resource-limited settings. *BMC Med Educ* 2025;25(1):587 [FREE Full text] [doi: [10.1186/s12909-025-07151-5](https://doi.org/10.1186/s12909-025-07151-5)] [Medline: [40259375](#)]

53. Ataro G, Worku S, Asaminew T. Experience and challenges of objective structured clinical examination (OSCE): perspective of students and examiners in a clinical department of ethiopian university. *Ethiop J Health Sci* 2020;30(3):417-426 [FREE Full text] [doi: [10.4314/ejhs.v30i3.13](https://doi.org/10.4314/ejhs.v30i3.13)] [Medline: [32874085](#)]

54. Sallam M, Barakat M, Sallam M. A preliminary checklist (METRICS) to standardize the design and reporting of studies on generative artificial intelligence-based models in health care education and practice: development study involving a literature review. *Interact J Med Res* 2024;13:e54704 [FREE Full text] [doi: [10.2196/54704](https://doi.org/10.2196/54704)] [Medline: [38276872](#)]

55. ChatGPT. OpenAI. 2025. URL: <https://chatgpt.com> [accessed 2025-06-29]

56. ChatGPT: explore GPTs. OpenAI. 2025. URL: <https://chatgpt.com/gpts> [accessed 2025-06-29]

57. Meskó B. Prompt engineering as an important emerging skill for medical professionals: tutorial. *J Med Internet Res* 2023;25:e50638 [FREE Full text] [doi: [10.2196/50638](https://doi.org/10.2196/50638)] [Medline: [37792434](#)]

58. Başaranoğlu M, Akbay E, Erdem E. AI-generated questions for urological competency assessment: a prospective educational study. *BMC Med Educ* 2025;25(1):611 [FREE Full text] [doi: [10.1186/s12909-025-07202-x](https://doi.org/10.1186/s12909-025-07202-x)] [Medline: [40281621](#)]

59. Coşkun Ö, Kiyak YS, Budakoğlu I. ChatGPT to generate clinical vignettes for teaching and multiple-choice questions for assessment: a randomized controlled experiment. *Med Teach* 2025;47(2):268-274. [doi: [10.1080/0142159X.2024.2327477](https://doi.org/10.1080/0142159X.2024.2327477)] [Medline: [38478902](#)]

60. Ahmed A, Kerr E, O'Malley A. Quality assurance and validity of AI-generated single best answer questions. *BMC Med Educ* 2025;25(1):300 [FREE Full text] [doi: [10.1186/s12909-025-06881-w](https://doi.org/10.1186/s12909-025-06881-w)] [Medline: [40001164](#)]

61. Arain SA, Akhund SA, Barakzai MA, Meo SA. Transforming medical education: leveraging large language models to enhance PBL-a proof-of-concept study. *Adv Physiol Educ* 2025;49(2):398-404 [FREE Full text] [doi: [10.1152/advan.00209.2024](https://doi.org/10.1152/advan.00209.2024)] [Medline: [39918742](#)]

62. Khan KZ, Gaunt K, Ramachandran S, Pushkar P. *Med Teach* 2013 Sep;35(9):e1447-e1463. [doi: [10.3109/0142159X.2013.818635](https://doi.org/10.3109/0142159X.2013.818635)] [Medline: [23968324](#)]

63. Titre II : Recherches impliquant la personne humaine (Articles L1121-1 à L1128-12) [Webpage in French]. Légifrance. 2025. URL: https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000006072665/LEGISCTA000006154978/#LEGISCTA000025457387 [accessed 2025-08-06]

64. Upadhayay B, Behzadan V, Karbasi A. Cognitive overload attack:prompt injection for long context. : ArXiv [FREE Full text] Preprint posted online on 15 Oct 2024

65. Knowledge in GPTs. OpenAI Help Center. URL: <https://help.openai.com/en/articles/8843948-knowledge-in-gpts> [accessed 2025-06-29]

66. Retrieval augmented generation (RAG) and semantic search for GPTs. OpenAI Help Center. 2025. URL: <https://help.openai.com/en/articles/8868588-retrieval-augmented-generation-rag-and-semantic-search-for-gpts> [accessed 2025-06-29]
67. Key guidelines for writing instructions for custom GPTs. OpenAI Help Center. URL: <https://help.openai.com/en/articles/9358033-key-guidelines-for-writing-instructions-for-custom-gpts> [accessed 2025-06-29]
68. Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt J, Laleh NG, et al. The future landscape of large language models in medicine. *Commun Med (Lond)* 2023;3(1):141 [FREE Full text] [doi: [10.1038/s43856-023-00370-1](https://doi.org/10.1038/s43856-023-00370-1)] [Medline: [37816837](https://pubmed.ncbi.nlm.nih.gov/37816837/)]
69. Wang Z, Chu Z, Doan TV, Ni S, Yang M, Zhang W. History, development, and principles of large language models: an introductory survey. *AI Ethics* 2024;5(3):1955-1971. [doi: [10.1007/s43681-024-00583-7](https://doi.org/10.1007/s43681-024-00583-7)]
70. 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;15(6):e40977 [FREE Full text] [doi: [10.7759/cureus.40977](https://doi.org/10.7759/cureus.40977)] [Medline: [37519497](https://pubmed.ncbi.nlm.nih.gov/37519497/)]
71. 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;31(9):3872-3878 [FREE Full text] [doi: [10.1016/j.acra.2024.06.046](https://doi.org/10.1016/j.acra.2024.06.046)] [Medline: [39013736](https://pubmed.ncbi.nlm.nih.gov/39013736/)]
72. Vaid A, Lampert J, Lee J, Sawant A, Apakama D, Sakuja A, et al. Natural language programming in medicine: administering evidence based clinical workflows with autonomous agents powered by generative large language models. : ArXiv [FREE Full text]
73. Ke YH, Jin L, Elangovan K, Abdullah HR, Liu N, Sia ATH, et al. Retrieval augmented generation for 10 large language models and its generalizability in assessing medical fitness. *NPJ Digit Med* 2025;8(1):187 [FREE Full text] [doi: [10.1038/s41746-025-01519-z](https://doi.org/10.1038/s41746-025-01519-z)] [Medline: [40185842](https://pubmed.ncbi.nlm.nih.gov/40185842/)]
74. DigiHealth Paris Cité : la révolution de l'enseignement de la santé numérique [webpage in French]. Faculté de Santé, Université Paris Cité. 2025. URL: <https://u-paris.fr/sante/digihealth/> [accessed 2025-10-13]

Abbreviations

AI: artificial intelligence

cGPT: custom GPT-4o

LLM: large language model

METRICS: Model, Evaluation, Timing, Range/Randomization, Individual factors, Count, Specificity

OSCE: Objective Structured Clinical Examination

RAG: retrieval-augmented generation

SP: standardized participant

Edited by R Pellegrino; submitted 10.Aug.2025; peer-reviewed by E Emekli, A Hidki, J Wang; comments to author 09.Oct.2025; revised version received 15.Oct.2025; accepted 13.Dec.2025; published 15.Jan.2026.

Please cite as:

Zouakia Z, Logak E, Szymczak A, Jais JP, Burgun A, Tsopra R

AI-Driven Objective Structured Clinical Examination Generation in Digital Health Education: Comparative Analysis of Three GPT-4o Configurations

JMIR Med Educ 2026;12:e82116

URL: <https://mededu.jmir.org/2026/1/e82116>

doi: [10.2196/82116](https://doi.org/10.2196/82116)

PMID:

©Zineb Zouakia, Emmanuel Logak, Alan Szymczak, Jean-Philippe Jais, Anita Burgun, Rosy Tsopra. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 15.Jan.2026. 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.

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

<https://www.jmirpublications.com/>