JMIR Medical Education

Impact Factor (2024): 3.2
Volume 8 (2022), Issue 2 ISSN 2369-3762 Editor-in-Chief: Blake J. Lesselroth, MD, MBI, FACP, FAMIA

Contents

Review

Use of the CPD-REACTION Questionnaire to Evaluate Continuing Professional Development Activities for Health Professionals: Systematic Review

Gloria Ayivi-Vinz[1,2,3], BSc, MPH; Felly Bakwa Kanyinga[1,2,3], MD; Lysa Bergeron[2], BSc, MSc; Simon Décary[4], BSc, MSc, PhD; Évèhouénou Lionel Adisso[1,2,3], BSc, MSc; Hervé Tchala Vignon Zomahoun[1,5], BSc, MSc, PhD; Sam J Daniel[6], MSc, MD, FRCSC; Martin Tremblay[6], PhD; Karine V Plourde[1,2], BSc, MSc, PhD; Sabrina Guay-Bélanger[1,2], BSc, MSc, PhD; France Légaré[1,2,5,7], BSc, MSc, MD, PhD, FCMF

[1]VITAM – Centre de Recherche en Santé Durable, Centre Intégré Universitaire de Santé et de Services Sociaux de la Capitale-Nationale, Université Laval, Quebec, QC, Canada

[2]Tier 1 Canada Research Chair in Shared Decision Making and Knowledge Translation, Université Laval, Quebec, QC, Canada

[3]Department of Social and Preventive Medicine, Faculty of Medicine, Université Laval, Quebec, QC, Canada

[4]School of Rehabilitation, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada

[5]Unité de Soutien SSA Québec, Université Laval, Quebec, QC, Canada

[6]Direction du Développement Professionnel Continu, Fédération des Médecins Spécialistes du Québec, Montreal, QC, Canada

[7]Department of Family Medicine and Emergency Medicine, Faculty of Medicine, Université Laval, Quebec, QC, Canada

Corresponding Author:
France Légaré, BSc, MSc, MD, PhD, FCMF
VITAM – Centre de Recherche en Santé Durable, Centre Intégré Universitaire de Santé et de Services Sociaux de la Capitale-Nationale
Université Laval
2480 Chemin de la Canardière
Quebec, QC, G1J 2G1
Canada
Phone: 1 418 663 5919
Email: France.Legare@mfa.ulaval.ca

Abstract

Background: Continuing professional development (CPD) is essential for physicians to maintain and enhance their knowledge, competence, skills, and performance. Web-based CPD plays an essential role. However, validated theory–informed measures of their impact are lacking. The CPD-REACTION questionnaire is a validated theory–informed tool that evaluates the impact of CPD activities on clinicians' behavioral intentions.

Objective: We aimed to review the use of the CPD-REACTION questionnaire, which measures the impact of CPD activities on health professionals' intentions to change clinical behavior. We examined CPD activity characteristics, ranges of intention, mean scores, score distributions, and psychometric properties.

Methods: We conducted a systematic review informed by the Cochrane review methodology. We searched 8 databases from January 1, 2014, to April 20, 2021. Gray literature was identified using Google Scholar and Research Gate. Eligibility criteria included all health care professionals, any study design, and participants' completion of the CPD-REACTION questionnaire either before, after, or before and after a CPD activity. Study selection, data extraction, and study quality evaluation were independently performed by 2 reviewers. We extracted data on characteristics of studies, the CPD activity (eg, targeted clinical behavior and format), and CPD-REACTION use. We used the Mixed Methods Appraisal Tool to evaluate the methodological quality of the studies. Data extracted were analyzed using descriptive statistics and the Student t test (2-tailed) for bivariate analysis. The results are presented as a narrative synthesis reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.

Results: Overall, 65 citations were eligible and referred to 52 primary studies. The number of primary studies reporting the use of CPD-REACTION has increased continuously since 2014 from 1 to 16 publications per year (2021). It is available in English, French, Spanish, and Dutch. Most of the studies were conducted in Canada (30/52, 58%). Furthermore, 40 different clinical behaviors were identified. The most common CPD format was e-learning (34/52, 65%). The original version of the CPD-REACTION questionnaire was used in 31 of 52 studies, and an adapted version in 18 of 52 studies. In addition, 31% (16/52) of the studies measured both the pre- and postintervention scores. In 22 studies, CPD providers were university-based. Most studies targeted interprofessional groups of health professionals (31/52, 60%).

Conclusions: The use of CPD-REACTION has increased rapidly and across a wide range of clinical behaviors and formats, including a web-based format. Further research should investigate the most effective way to adapt the CPD-REACTION questionnaire to a variety of clinical behaviors and contexts.

Trial Registration: PROSPERO CRD42018116492; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=116492

(JMIR Med Educ 2022;8(2):e36948) doi:10.2196/36948

KEYWORDS

CPD-REACTION; behavior; intention; education medical; continuing; health care professionals; questionnaire; web-based; continuing professional development

Introduction

Continuing professional development (CPD) encompasses the multiple educational and developmental activities that health care professionals undertake to maintain and enhance their knowledge, skills, performance, and relationships in the provision of health care. The ultimate goal of CPD is to enhance the quality and safety of patient care and enhance both patient experience and health outcomes [1]. In recent years, web-based CPD has increased exponentially, and the recent COVID-19 pandemic has emphasized the need for more effective web-based CPD. Health professional behavior change (adoption or abandonment of a practice) is a long and complex process [2]. The Kirkpatrick model conceptualizes a framework for CPD assessment that measures four distinct outcome levels: satisfaction; knowledge, skills, or attitudes; transfer of learning to practice (ie, behavior); and organizational outcomes such as productivity and quality [3].

The lack of validated instruments informed by behavior change theories for assessing CPD outcomes has slowed the advancement of the CPD knowledge base [4]. In 2011, a consortium of CPD providers from the Province of Quebec, Canada, developed a tool to assess Kirkpatrick level 3 outcomes (transfer of learning to practice) based on an integrated model explaining behavior change among health professionals [5,6]. This model posits that intention is a strong predictor of behavior, and that behavioral intention, in turn, is influenced by beliefs about capabilities, beliefs about consequences, moral norms, and social influences [5]. The resulting tool, the CPD-REACTION questionnaire, is a comprehensive, theory-based, validated instrument for assessing the impact of accredited CPD activities on clinical behavioral intention [7,8]. During the past 10 years, it has been used in regular evaluations of the effects of CPD activities on behavior change by major CPD providers such as the Federation of Medical Specialists of Quebec (Fédération des Médecins Spécialistes du Québec) and to assess training for a wide variety of other health care professionals [9-12].

However, the current range of CPD-REACTION use remains unknown. Moreover, the clinical topics of CPD activities evaluated using the tool, the types of clinical behaviors sought, how often it has been used to evaluate web-based CPD, what kind of health care professionals are targeted by such CPD activities, and how the results shown by CPD-REACTION in terms of behavior change intentions are used, are also unknown. Although tool validity has been demonstrated in the Canadian context [8], other evidence on its cross-cultural validity and psychometric properties is still lacking. Therefore, we aimed to systematically review studies that have used the CPD-REACTION questionnaire.

Our research questions were as follows: (1) What are the characteristics of CPD activities in studies using CDP-REACTION? (2) What are the ranges of behavioral change intentions, mean scores, and distribution of scores across all studies that used CPD-REACTION? (3) What are the psychometric properties of CPD-REACTION?

Methods

Ethics Approval and Consent to Participate

As this research was based on published studies, ethics approval was not required for this systematic review. The protocol was registered in the PROSPERO (International Prospective Register of Systematic Reviews) registry under the number CRD42018116492 on December 4, 2018. The main change to the protocol was the inclusion of references to studies reported in a language other than English and French.

Study Design

Informed by the Cochrane review methodology [13], we conducted a systematic review and followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement [14].

Eligibility Criteria

Informed by the PICOS (Population, Intervention, Comparator, Outcomes, Study design) model [15], the inclusion criteria were as follows: (1) *Population*- the target population considered for this review included all individuals working in health fields who completed an original, translated, or adapted version of the CPD-REACTION questionnaire before, after, or before and after an activity. There was no age restriction or restriction of health care professions (eg, physician, nurse, or any other health professional). They could be working in the public or private sector, in the process of training, or have already graduated; (2)

Intervention- not specified; (3) *Comparator*- not specified; (4) *Outcomes*- The original, adapted, or translated version of the 12-item CPD-REACTION was used to assess the intention to change either a clinical practice or a health behavior. On the basis of the Godin integrated model, this tool is a questionnaire composed of 12 items that measure behavioral intention and some of its predictors; that is, beliefs about capabilities, beliefs about consequences, moral norms, and social influences [5,7]. The 5 constructs of the CPD-REACTION questionnaire have been validated, with Cronbach coefficients for the constructs varying from 0.77 to 0.85 [7,8]; (5) *Study Design*-Any study design was considered: randomized clinical trials (individual, group, or cluster, including stepped-wedge), before-and-after studies, translation studies, ecology studies, qualitative studies, or any mixed-study design (if they included the use of CPD-REACTION). Only primary studies were considered for inclusion in this systematic review. Therefore, we did not include any systematic reviews. These articles could be reported in any language.

Information Sources

The literature search was performed using eight databases: Embase, MEDLINE/PubMed, Web of Science, ERIC-EBSCO, PsycINFO-ovid, CINAHL, Social Sciences Full Text-EBSCO, and Academic Search Premier EBSCO. A temporal filter was applied from January 1, 2014, to April 20, 2021, because CPD-REACTION was published in 2014 [7]. We also performed a forward citation search using Google Scholar and Research Gate to identify studies citing the 3 main studies on the development and validation of CPD-REACTION (Multimedia Appendix 1).

Search Strategy

The first phase of developing the search strategy was carried out on PubMed and reviewed by the authors to ensure that the concepts covered all research questions. This strategy was then translated into expressions that were adapted to each database. A documentary research expert revised the search strategy and the final version was based on three key concepts: "continuing education," "CPD-REACTION questionnaire," and "questionnaires." These key concepts were searched using a combination of controlled vocabulary (MeSH [Medical Subject Headings] terms) and free-text search queries (Multimedia Appendix 1).

Selection Process

Duplicates were identified using EndNote ×9 [16] and manual checking. First, reviewers (GA-V, FBK, LB, and LS) performed an independent selection based on the title and abstract. Second, all relevant references were considered for selection by full text (GA-V, FBK, LL, LB, and LS). An internet-based system, Covidence [17], was used to complete this step. The 2 reviewers then discussed and resolved any disagreement to obtain a consensus on study selection according to the eligibility criteria and, if necessary, consulted a third author (KVP). The reasons for exclusion of articles were documented.

Data Extraction

A coding guide and corresponding extraction grid were developed and tested by the reviewers. The reviewers (GA-V, FBK, LL, LB, and LS) individually extracted data from the included studies. The reviewers discussed and resolved any disagreement.

Qualitative and quantitative data were extracted. The main groups of variables were (1) study characteristics, including author names, study design, study objectives, country, and type of CPD activity; (2) characteristics of the study participants, such as profession, setting (eg, hospital or university), average age, sex, study population (eg, single profession, mixed professions, and patients included); (3) CPD activity characteristics, such as country in which it was used, health field, duration of CPD activity, when tool was used (eg, pre- or post-CPD activity), format of CPD activity (eg, web-based), title of CPD activity, clinical behavior change targeted; (4) CPD-REACTION version used (original or adapted), adaptations to the questionnaire, eg, translations; (5) score values (mean, median, SD, minimum, and maximum) for all constructs measured, that is, behavioral intention, beliefs about capabilities, social influence, moral norm, and beliefs about consequences; (6) psychometric properties (Cronbach α, κ, or Cohen *d*).

Methodological Quality Assessment of Individual Studies

Two examiners (GA-V and FBK) assessed the quality of each identified study using the Mixed Methods Appraisal Tool (MMAT), a validated tool for evaluating the quality of qualitative, quantitative, and mixed methods studies [18]. For each type of study design, 5 criteria were evaluated and each was rated "yes," "can't tell," or "no." The tool guideline discourages the calculation of an overall score, instead suggesting presenting detailed ratings for each criterion [18] (Multimedia Appendix 2).

Data Synthesis

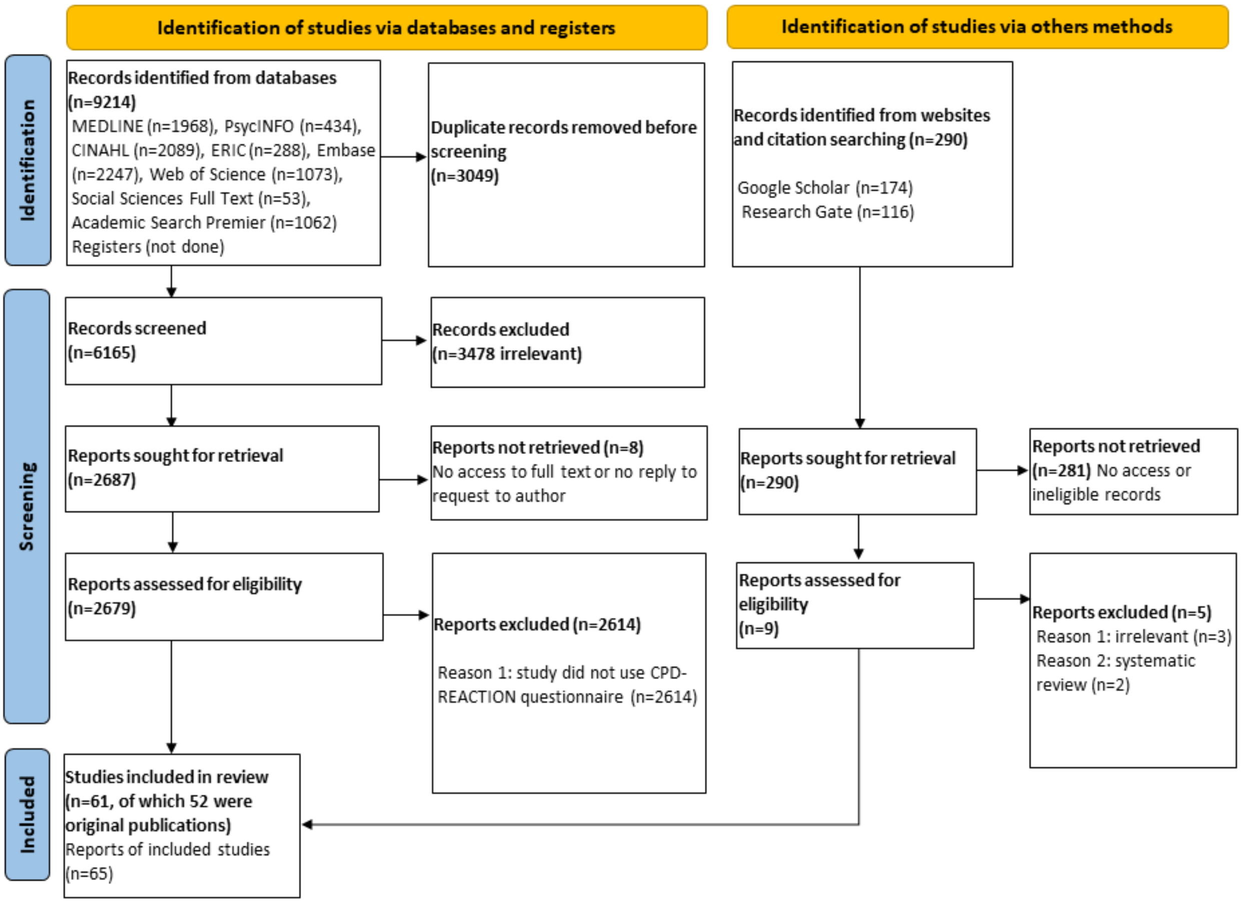
Given the large variety of behavior changes targeted by studies (clinical practice behaviors and others) and the methodological and statistical heterogeneity of studies, we performed a narrative synthesis using descriptive statistics. For the CPD-REACTION score values, we did not calculate the average scores for the construct if CPD-REACTION did not evaluate the same behavior. Instead, we summarized the construct scores based on the timing of the evaluation, that is, if it was a pre-post, only pre-evaluation, or only postevaluation. Descriptive statistics were computed using STATA (version 11; StataCorp). To summarize the target behaviors of the included studies, we performed a thematic analysis. After the analysis, we organized and summarized the main behaviors based on the emerging themes, namely, "shared decision-making," "decision aids or toolkit," and "others."

Results

Study Selection

We described the selection process in a PRISMA flowchart (Figure 1). A total of 9504 records were identified and 3330 (duplicates or ineligible) were removed. After screening, 65 records matched the eligibility criteria and referred to 61 publications and 52 unique studies [7-12,19-70] (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flowchart.



Study Characteristics

Since 2014, the number of published studies using CPD-REACTION has increased from 1 to 16 publications in 2021 (Multimedia Appendix 3). Of all the studies, 69% (36/52) were published between 2019 and 2021 [9,32-68]. Furthermore, 58% (30/52) were located in Canada [7,8,10-12,23-25,28,31,35,38-41,44-51,53,54,56,60,63,65,67] and the rest in the United States (n=6), the United Kingdom (n=4), Australia (n=2) [30,34], Iran (n=2) [55,64], Argentina (n=1) [52], Indonesia (n=1) [32], Germany (n=1) [58], Sweden (n=1) [21], the Netherlands (n=1) [59], and Burkina Faso (n=1) [37]. In addition, two multicountry studies were reported: 1 from Brazil-China-France-Japan-Mali [68] and 1 from Canada-Vietnam [19] (Table 1; Figure 2). There were no exclusive qualitative studies (Table 1). Most study designs were mixed methods (24/52, 46%) [8,9,12,18,23,26,35,42,43,45,47,50,55,56,61,64,67,71,72], followed by cross-sectional studies (9/52, 17%) [22,32,39,42,56,59].

Table 1. Study and intervention characteristics (N=52).

Study and intervention characteristics	Number of studies, n (%[a])
Study location	
Canada	30 (58)
United States	6 (12)
United Kingdom	4 (8)
Australia	2 (4)
Iran	2 (4)
Argentina	1 (2)
Burkina Faso	1 (2)
Germany	1 (2)
Indonesia	1 (2)
Netherlands	1 (2)
Sweden	1 (2)
Canada and Vietnam	1 (2)
China-Brazil-France-Mali-Canada-Japan	1 (2)
Study design	
Mixed methods study	24 (46)
Cross-sectional study	9 (17)
Baseline and follow-up or before-after or comparative study	7 (13)
Randomized trial	4 (8)
Quasi-experimental study	3 (6)
Validation study	2 (4)
Cohort study	2 (4)
Intervention study	1 (2)
Clinical setting	
Multicenter academic hospitals	3 (6)
Multicenter community hospitals	11 (21)
Multicenter both academic and community	11 (21)
Single-center academic hospital	4 (8)
Single-center community hospital	5 (10)
Not a clinical setting	13 (25)
Not reported or not applicable	5 (10)
Type of CPD[b] activities	
Course or workshop	31 (60)
Conference	1 (2)
Other[c] CPD activities	4 (8)
No activity pertaining to CPD[d]	3 (6)
Not specified or not applicable	13 (25)
CPD activity Format	
Web-based	34 (65)
In person	13 (25)
Not specified	5 (10)

Study and intervention characteristics	Number of studies, n (%[a])
Version of questionnaire used	
Adapted	18 (35)
Original	31 (60)
Not specified	3 (6)
When CPD-REACTION was used	
Preactivity	6 (12)
Postactivity	11 (21)
Pre- and postactivity	16 (31)
Not specified or not applicable	19 (37)
Delivery mode of CPD-REACTION questionnaire	
Digital platform or web-based	19 (37)
Web-based and paper	1 (2)
Paper copy	16 (31)
Not specified	16 (31)
Language of CPD questionnaire used	
Dutch	1 (2)
Spanish	1 (2)
English	28 (54)
French	14 (27)
English and French	3 (6)
Not reported	5 (10)
Type of CPD provider	
Government	1 (2)
Hospital	13 (25)
Private company	6 (12)
University	22 (42)
Not specified or not applicable	7 (19)

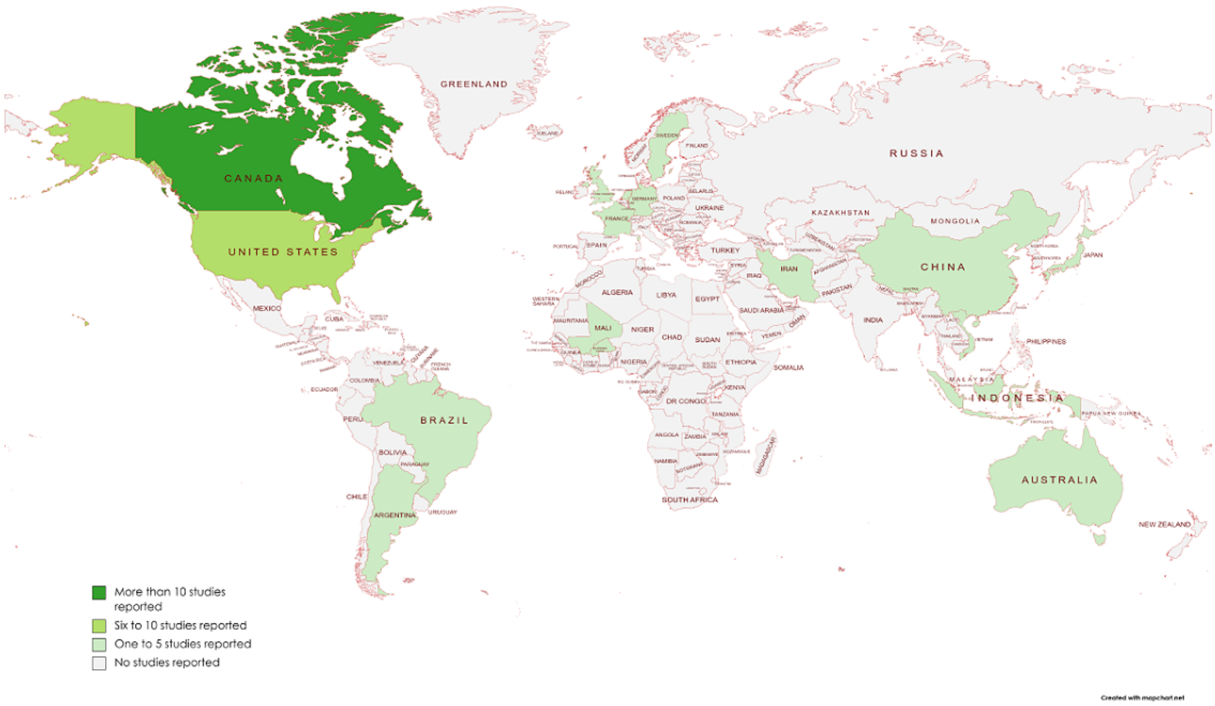
[a]All percentages may not add up to 100%.

[b]CPD: continuing professional development.

[c]Training or workshop combined with activities such as face-to-face meetings, media interviews, minutes documenting interactions, conferenced meetings, annual national collaboration meeting, and team meeting to watch video.

[d]Guidelines application, outreach sessions.

Figure 2. Distribution of published studies worldwide that used the CPD-REACTION questionnaire.



Characteristics of the Study Participants

In total, CPD-REACTION was administered to 4886 participants. Even when age was mentioned, it was not possible to properly report on age because of the heterogeneity of age ranges. The sex of the participants was not reported for all studies. The authors mostly defined participants based solely on their profession. Physicians were the most represented health profession (1843/4886, 37.72%). Furthermore, 7 studies included residents or unlicensed health professionals [8,36,38,39,46,55,64]. (Table 2). In most studies, participants in CPD activities consisted of interprofessional groups (30/52, 60%) [7,9-11,19,21,27,28,31,35,37,41,43,44,46,47,50,51,53,55-61,63,64,66-68]. Professions included nurses (5/52, 10%) [26,29,30,48,54], physicians (4/52, 8%) [38,39,49,52], social workers or other health professionals, namely occupational therapists, physiotherapists, dietitians, behavioral counselors, nutritionists, health researchers (4/52, 8%) [23-25,34,36], specialist physicians (3/52, 6%) [32,40,65], and pharmacists (2/52, 4%) [22,42]. The presence of managers or decision-makers among participants was reported in 7 (13%) out of 52 studies. The number of participants per study ranged from 8 to 489 (Figure 3).

The largest proportion of CPD providers reported was university-based (22/52, 42%), whereas others were based in hospitals (13/52, 25%), private companies (6/52, 12%), or government (1/52, 2%).

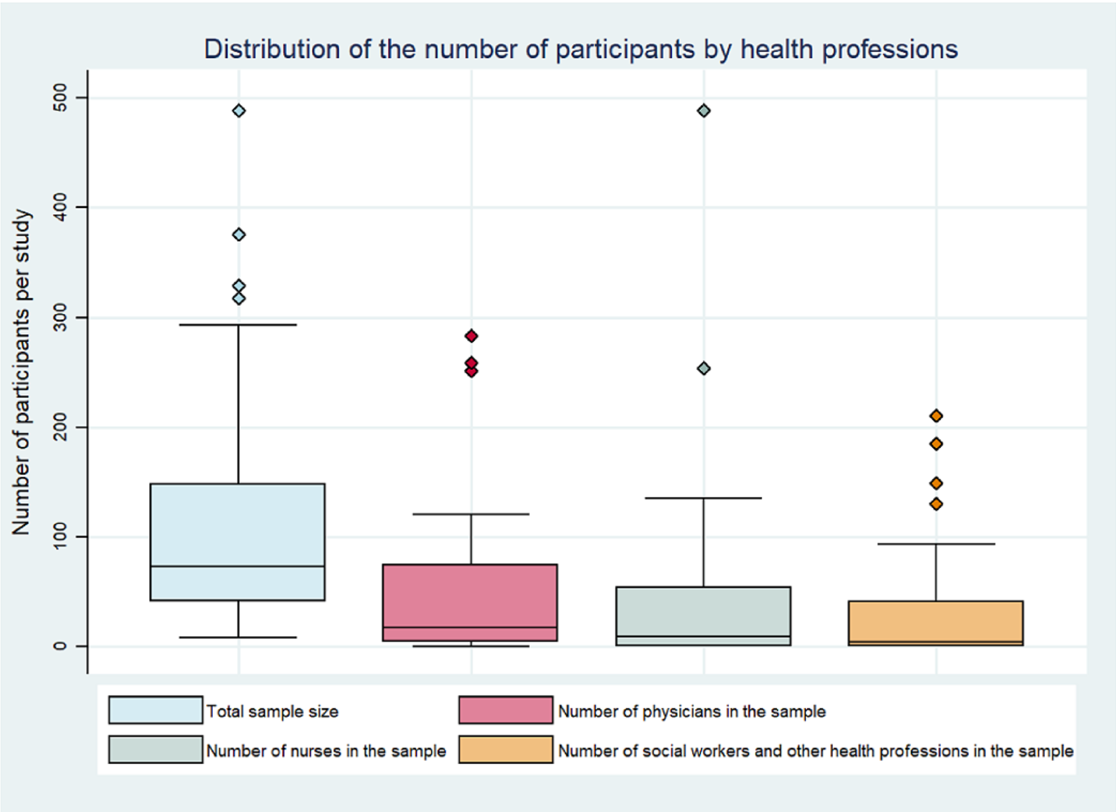
Table 2. Professional profiles of study participants.

Population characteristics	Frequency, n (%[a])
Population of interest, in the studies (n=52)	
Physicians	8 (15.4)
Interprofessional groups	31 (59.6)
Nurses	5 (9.6)
Other health professions[b]	8 (15.4)
Number of participants per professional group (n=4886)	
Interprofessional groups	1843 (37.7)
Nurses	1568 (32.1)
Social workers and other health professionals	1053 (21.6)
Not specified	422 (8.6)
Presence of managers or decision-makers among participants (n=52)	
Yes	7 (13.5)
No	45 (86.5)
Presence of residents or unlicensed health professionals among participants (n=52)	
Yes	7 (13.5)
No	45 (86.5)

[a]All percentages may not add up to 100%.

[b]Pharmacists, physical therapists, physiotherapists, providers of radiation therapy, midwives, and social workers.

Figure 3. Boxplot of number of participants by health profession present at each continuing professional development activity.



CPD Activity Characteristics

The questionnaire was administered in four languages: French (14/52, 27%) or English (28/52, 54%) [8-11,19,21-24,26-32,34-37,39-51,53-55,57,59-68], Spanish (1/52, 2%) [52] and Dutch (1/52, 2%) [58]. The median number of CPD activities targeting behavior change per study was 1 and varied between 1 and 9 CPD activities per study. One-quarter (11/52, 21%) of the studies used CPD-REACTION to measure behavioral change intention but were not linked to a specific CPD activity [21,22,29,31,38,40,45,46,51,53,59] (Table 1). The most common format for CPD activities was web-based or e-learning based (34/52, 65%). The duration of CPD activities ranged from 30 to 225 minutes, with an average of 115 (SD 67) minutes.

Targeted Clinical Behavior and Scoring of CPD-REACTION

The evaluations targeted 39 different clinical behaviors [7-12,19-70]. Thematic analysis showed that 7 (18%) out of the 39 pertained to shared decision-making [9,35,50,54,56,64,67] and 5 pertained to decision aids or toolkit implementation [53,60,61,63,66] (Table 3).

Regarding studies reporting mean scores after the intervention (n=33) [8,12,22,25,29,30,33-36,40-44,47-49,55,56,59,60,63,64,66], 9 studies (27%) reported both pre- and postactivity scores [8,22,25,29,42,49,56,60,66]. The scores were all higher after the intervention. Furthermore, in all 9 studies, the pre-post score ranges (2.5-5.7) were higher than in studies measuring prescores only (2.6-5.2) or postscores only (1.8-4.8; Table 4). The average difference between the pre- and postintention scores was 0.54 SD 0.13. Among the 5 CPD-REACTION constructs, social influence scored the lowest (43; Table 4).

Table 3. Main behaviors targeted in included studies (n=39).

	Main clinical behavior targeted in included studies	Topic theme		
		SDM[a]	Decision aids or toolkit	Others
1.	To prescribe spirometry and to interpret the result [38,39]	–[b]	–	+[c]
2.	To actively engage with and invite patients who are underserved for Medicine Use Reviews (MURS) [20,42,43]	–	–	+
3.	To adopt SDM [53]	+	–	–
4.	To engage older patients living with dementia and their caregivers in decision-making about choosing a health intervention, based on the TPB [10]	+	–	–
5.	To use Decision Box to explain to patients the benefits and harms of the options, based on the TPB [10]	–	+	–
6.	To use a decision aid in clinical practice after completing the web-based program “MyDiabetesPlan” [63]	–	+	–
7.	To implement developmental coordination disorder (DCD) best practices [23]	–	–	+
8.	To provide medical abortion [40]	–	–	+
9.	To use COSTARS (pan-Canadian Oncology Symptom Triage and Remote Support) practice guides [45]	–	–	+
10.	To use of 15 evidence-informed symptom practice guides for providing telephone or in-home nursing services to clients with cancer [45]	–	–	+
11.	To engage in IP-SDM (interprofessional shared decision-making) [50]	+	–	–
12.	To use patient decision aids [12,56,64]	–	+	–
13.	To counsel patients regarding HIV prep therapy [22]	–	–	+
14.	To use IP-SDM [50,55]	+	–	–
15.	To apply the disclosure guidelines to my practice [44]	–	–	+
16.	To apply the Situation-Background-Assessment-Recommendation (SBAR) to my practice [44]	–	–	+
17.	To apply quality improvement strategies to solve challenges in my practice [44]	–	–	+
18.	To practice the person-centered approach (PCA—MACHIP 2) in maternal health [37]	–	–	+
19.	“Utiliser l’outil d’évaluation du risque de violence” (To use the Risk of Violence evaluation tool) [31]	–	–	+
20.	To collaboratively work with and actively involve children and young people who self-harm in their care [29]	–	–	+
21.	To use the evidence of implementing FREEDOM [46]	–	–	+
22.	To implement the STEADI toolkit [61]	–	+	–
23.	To report research translation and impact on the CV[d] [51]	–	–	+
24.	To use SDM [35]	+	–	–
25.	To prescribe no pharmacological treatments [36]	–	–	+
26.	To use SDM with their next patient facing a preference-sensitive decision [56]	+	–	–
27.	To apply a systematic framework to identify and manage patients with dementia [34]	–	–	+
28.	To change and improve practice based on the interventions, that is, to order pneumococcal vaccines [41]	–	–	+
29.	To use research evidence in rheumatology [21]	–	–	+
30.	To successfully plan and implement evidence-based practice changes in health facility [27]	–	–	+
31.	To consider probiotic recommendation in infants and toddler patients [32]	–	–	+
32.	To perform SDM (action) among health professionals in any clinical setting [64]	+	–	–
33.	To use an app to decide about prenatal screening [9,54]	–	+	–
34.	To formulate a violence risk assessment and management plan [30]	–	–	+
35.	To use de-escalation techniques during escalating aggression [30]	–	–	+
36.	To use breakaway techniques when responding to a violent person [30]	–	–	+
37.	To change their practice about compassion fatigue education [57]	–	–	+
38.	To implement the 5A method training in the area of physical activity promotion [58]	–	–	+

	Main clinical behavior targeted in included studies	Topic theme		
		SDM[a]	Decision aids or toolkit	Others
39.	To care for children and young people admitted to hospital with self-harm [29]	–	–	+

[a]SDM: shared decision-making.
[b]Not related to theme.
[c]Related to theme.
[d]CV: Curriculum Vitae.

Table 4. Summary of pre- and postscores for all constructs of CPD-REACTION.

	Value, n	Pre-CPD[a] activity (range)	Post-CPD activity (range)
Interventions with pre- and post-CPD scores	9		
Intention	9	4.5-6.5	5.7-6.8
Social influence	9	2.5-5.6	3.8-5.8
Beliefs about capabilities	9	3.2-6	5.4-6.4
Moral norm	7	5.51-6.7	6.2-6.9
Beliefs about consequences	9	5.73-6.6	6.2-6.8
Interventions with only prescores			—[b]
Intention	5	2.9-6.6	
Social influence	5	2.6-6	
Beliefs about capabilities	5	2.4-6.6	
Moral norm	5	4.3-6.8	
Beliefs about consequences	4	5.2-6.7	
Interventions with only postscores		—	
Intention	19		3.4-7
Social influence	18		1.8-6.3
Beliefs about capabilities	18		3.9-6.8
Moral norm	18		4.6-6.9
Beliefs about consequences	18		4.8-4.8

[a]CPD: continuing professional development.
[b]Author did not report or measure mean scores.

CPD-REACTION Adaptations and Psychometric Properties

One-third (18/52, 35%) of the included studies reported having adapted CPD-REACTION [19,21,22,27,28,32,39,40,53,55,56,58,63,66-68,73] (Table 1). Adaptations to the questionnaire reported were reformulated items (n=3) [19,28,49], using only certain construct scales (n=3 studies) [22,53,55], adding or dropping some items without reformulating the original items (n=3) [39,56,67], reporting percentages instead of score values ranging from 1 to 7 (n=2) [23,27], translation of the questionnaire into other languages (n=2) [52,58], and using a 5-point instead of 7-point Likert scale (n=1) [40]. Furthermore, more than 80% of all studies (48/52, 92%) reported the psychometric parameters of the original version of CPD-REACTION or else stated it was a validated tool [8,10,23,25,26,29-32,34-37,40-57,60,61,64-68,74]. In addition, 4 studies reported the psychometric properties of their adapted versions [28,39,52,69], with the Cronbach α of the included constructs ranging from 0.62 to 0.91.

Risk of Bias in Studies

Although none of the studies fully met all MMAT criteria, none were rated “no” for any criteria (Multimedia Appendix 2). In the 4 quantitative randomized trials, only the criterion “randomization appropriately performed” was met by all 4 studies [10,50,54,63], and in all 17 mixed method studies, only the criterion “adequate rationale for using a mixed methods design” was met. In all the design groups, all the criteria not rated “yes” were rated “not sure.”

Discussion

Principal Findings

We found 61 publications of 52 unique studies that reported the use of the CPD-REACTION questionnaire to assess changes in behavioral intention among health professionals. Although the tool is aged <10 years (2014), we observed the most rapid increase in its use in the past 3 years, mostly in Canada, where it was developed. However, its use has spread to many other countries, including lower- and middle-income countries, and it is found in numerous languages (our finding of only 4 is an underrepresentation, as the team that produced the tool has agreed to translations into 8 languages) [75]. Since its inception, CPD-REACTION has been used by close to 5000 participants to target 39 clinical behaviors. The participants included 8 types of health professionals, with physicians and nurses being the most reported. Two-thirds of the studies included interprofessional clinical teams, including one in which 10 managers or decision-makers were CPD activity participants. The tool appeared to be mostly used for evaluating e-learning (n=34). In many cases, users adapted the questionnaire, such as using only certain construct scales or adding or dropping some items. The psychometric properties of CPD-REACTION reported in included studies showed that Cronbach α scores were very good, ranging from 0.62-0.91. However, few studies were designed to assess changes in intention (ie, scoring both pre- and postactivity), thus limiting the evidence regarding the responsiveness of the tool. Regarding behavioral intention to change, the mean difference of intention score was 0.54 SD 0.13 in the pre-post studies and the distribution of scores across all studies using CPD-REACTION ranged from 1.8-7. Although none of the studies fully met all MMAT criteria, none were rated "no" for any criteria. In all the design groups, all the criteria not rated "yes" were rated "not sure."

Significance and Comparison With Prior Work

First, the rapid adoption of CPD-REACTION across time, countries, and languages suggests that this instrument addresses the needs of CPD developers and that they seek not only validated assessment tools but also those that are informed by behavior change theories. Recent literature on this topic tends to suggest an increasing penetration of behavior change theories in the CPD developer community [2,76,77]. The use of behavior change theory has been frequently linked to effectiveness in systematic reviews of behavioral change interventions [76,77]. More recently, strategies have also focused not only on adopting new behaviors but also on abandoning low-value or harmful behaviors. However, few behavioral theories distinguish between behavior adoption and abandonment, including the theories on which CPD-REACTION is based [78]. Future research should distinguish between the two and develop theories that support both types of behavior change [79,80].

Second, physicians and nurses were the most represented health professionals. Most groups of participants engaging in CPD activities were interprofessional clinical teams, and 1 in 10 studies included managers or decision-makers among participants. This suggests that CPD designers are increasingly creating multidisciplinary training experiences to be shared with other stakeholders and professionals to enhance the relevance and impact of CPD [2,81]. Previous research has highlighted that including peer groups seems to be an effective approach to enhancing CPD activities and moving forward with professional practice change [82]. Future research should determine the effects of interprofessional participant groups or peer groups on CPD effectiveness.

Third, studies using an adapted version of CPD-REACTION reported Cronbach α ranging from .62 to .91, indicating that modified instruments perform well in terms of their psychometric properties. Other studies have reported psychometric values mentioned in the original version of CPD-REACTION. We observed that overall, the behavioral change intention scores reported ranged from 2.9 to 7. In pre-post studies, the mean difference in intention scores was 0.54 SD 0.13, and the distribution of scores across all constructs ranged from 1.8 to 7. Lower scores were observed when CPD-REACTION was used either only preactivity or only postactivity. Dissemination of the user manual will aid in the use of the tool to its best advantage. A lower score could also be because of the CPD topic being more controversial and thus less likely to be implementable. Overall, the adapted versions of CPD-REACTION reported Cronbach α values, indicating that the questionnaire had good internal consistency reliability. Furthermore, our results suggest that CPD-REACTION is adaptable to digital platforms, as two-thirds of the activities were web-based.

Fourth, using CPD-REACTION to measure construct scores, both pre- and post-CPD activity, is a helpful demonstration of the effect of CPD activities on behavioral intention and explanatory constructs. However, measuring learning outcomes for levels 3 and 4 of the Kirkpatrick model remains challenging. CPD-REACTION uses intention as a measure of behavioral intention; however, other measurement strategies are needed to directly measure behavior change. Although other outcomes such as "satisfaction of participants" were reported, the studies did not correlate these with the CPD-REACTION measures. In some studies, participants were contacted after 3 months or more to self-assess their behavior change [8]. The purpose of CPD-REACTION was not to measure its effects on patient outcomes, which is another important outcome of CPD. CPD-REACTION could be followed up by participant surveys to assess the longer-term impacts of participants' behaviors on their practices or institutions and should use patient-reported measures. Some studies suggest that CPD programs should compare self-assessments, such as CPD-REACTION, with continuous formal participant multisource assessment by peers [73].

Limitations

Our systematic review used diverse strategies to find studies that had used CPD-REACTION. However, we relied on the published results and did not contact the authors of the included studies. Thus, it is possible that we may have missed studies that were not published as well as items of interest in those we included. Owing to the large number of included studies, we had to organize the information into broad categories to increase the interpretability of the data.

Conclusions

The CPD-REACTION questionnaire is a simple, relevant, and easy-to-use tool for assessing the effectiveness of CPD activities on health professionals' behavioral intention and, as we have observed, to identify barriers and facilitators of behavior change. This tool has been used to evaluate CPD activities in a wide range of clinical topics and behaviors. However, most users do not measure intention both before and after the activity. Dissemination of a user manual will aid in the use of the tool to its best advantage. Further research should investigate the most effective way to adapt the CPD-REACTION questionnaire to various uses and contexts.

Acknowledgments

The authors wish to thank Nathalie Rheault for her dedicated assistance with the design and search strategy and Louisa Blair for language revision and editing. The authors also thank Laure Siebierski, Samy Bouderba, Angèle Musabyimana, Laura Langevin, and Yuxi Wang for their assistance with the selection process. The following bodies funded the development of “CPD-REACTION questionnaire”: The Tier 1 Canada Research Chair in Shared Decision-Making and Knowledge Translation, the *Fédération des médecins spécialistes du Québec* (FMSQ; Federation of Medical Specialists of Quebec), and the *Conseil Québécois de développement professionnel continu des médecins* (CQDPCM; Quebec Council of Continuing Professional Development for Physicians).

Data Availability

The data sets used and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

GA-V, LB, KVP, SG-B, SD, ÉLA, HTVZ, SJD, MT, and FL designed the research protocol. GA-V, FB, LB, KVP, and SD conducted study selection and data extraction. GA-V and FB performed the quality assessment of the studies. GA-V conducted data synthesis, which was revised by KVP, SG-B, MT, HTVZ, and FL. GA-V drafted the manuscript overseen by KVP, SG-B, HTVZ, LB, SD, ÉLA, SJD, MT, and FL. SJD and MT provided expertise as a continuing professional development activities developer. All authors contributed to the key intellectual content and provided consent for this version of the review to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 31 KB - mededu_v8i2e36948_app1.docx]

Multimedia Appendix 2
Mixed Methods Appraisal Tool criteria quality assessment in the included studies.
[PDF File (Adobe PDF File), 74 KB - mededu_v8i2e36948_app2.pdf]

Multimedia Appendix 3
Evolution of the number of published studies using CPD-REACTION questionnaire, 2014-2021.
[PPTX File , 68 KB - mededu_v8i2e36948_app3.pptx]

References

1. Sargeant J, Bruce D, Campbell CM. Practicing physicians' needs for assessment and feedback as part of professional development. J Contin Educ Health Prof 2013;33 Suppl 1:S54-S62. [doi: 10.1002/chp.21202] [Medline: 24347154]
2. Chichirez CM, Purcărea VL. Health marketing and behavioral change: a review of the literature. J Med Life 2018;11(1):15-19 [FREE Full text] [Medline: 29696059]
3. Kirkpatrick DL, Kirkpatrick JD. Evaluating training programs: the four levels. 3rd edition. San Francisco, CA, USA: Berrett-Koehler Publishers; 2006.
4. Légaré F, Borduas F, Jacques A, Laprise R, Voyer G, Boucher A, et al. Developing a theory-based instrument to assess the impact of continuing professional development activities on clinical practice: a study protocol. Implement Sci 2011 Mar 07;6:17 [FREE Full text] [doi: 10.1186/1748-5908-6-17] [Medline: 21385369]
5. 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 Jul 16;3:36 [FREE Full text] [doi: 10.1186/1748-5908-3-36] [Medline: 18631386]
6. Ajzen I. From intentions to actions: a theory of planned behavior. In: Kuhl J, Beckmann J, editors. Action control: from cognition to behavior. Berlin, Germany: Springer; 1985:11-39.

7. Légaré F, Borduas F, Freitas A, Jacques A, Godin G, Luconi F, CPD-KT team. 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] [Medline: 24643173]
8. Légaré F, Freitas A, Turcotte S, Borduas F, Jacques A, Luconi F, et al. 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] [Medline: 28459836]
9. Abbasgholizadeh Rahimi S, Archambault PM, Ravitsky V, Lemoine ME, Langlois S, Forest JC, et al. An analytical mobile app for shared decision making about prenatal screening: protocol for a mixed methods study. JMIR Res Protoc 2019;8(10):e13321 [FREE Full text] [doi: 10.2196/13321] [Medline: 31596249]
10. Giguere AM, Lawani MA, Fortier-Brochu É, Carmichael PH, Légaré F, Kröger E, et al. Tailoring and evaluating an intervention to improve shared decision-making among seniors with dementia, their caregivers, and healthcare providers: study protocol for a randomized controlled trial. Trials 2018 Jun 25;19(1):332 [FREE Full text] [doi: 10.1186/s13063-018-2697-1] [Medline: 29941020]
11. Légaré F, Brière N, Stacey D, Lacroix G, Desroches S, Dumont S, et al. Implementing shared decision-making in interprofessional home care teams (the IPSDM-SW study): protocol for a stepped wedge cluster randomised trial. BMJ Open 2016 Nov 24;6(11):e014023 [FREE Full text] [doi: 10.1136/bmjopen-2016-014023] [Medline: 27884857]
12. Stacey D, Taljaard M, Smylie J, Boland L, Breau RH, Carley M, et al. Implementation of a patient decision aid for men with localized prostate cancer: evaluation of patient outcomes and practice variation. Implement Sci 2016 Jul 02;11(1):87 [FREE Full text] [doi: 10.1186/s13012-016-0451-1] [Medline: 27368830]
13. Chandler J, Hopewell S. Cochrane methods--twenty years experience in developing systematic review methods. Syst Rev 2013 Sep 20;2:76 [FREE Full text] [doi: 10.1186/2046-4053-2-76] [Medline: 24050381]
14. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015 Jan 02;350:g7647 [FREE Full text] [doi: 10.1136/bmj.g7647] [Medline: 25555855]
15. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. BMC Health Serv Res 2014 Nov 21;14:579 [FREE Full text] [doi: 10.1186/s12913-014-0579-0] [Medline: 25413154]
16. EndNote. Clarivate Analytics. 2018. URL: https://endnote.com/ [accessed 2022-04-12]
17. Covidence systematic review software. Veritas Health Innovation. URL: https://www.covidence.org/ [accessed 2022-04-12]
18. Hong QN, Gonzalez-Reyes A, Pluye P. Improving the usefulness of a tool for appraising the quality of qualitative, quantitative and mixed methods studies, the Mixed Methods Appraisal Tool (MMAT). J Eval Clin Pract 2018 Jun;24(3):459-467. [doi: 10.1111/jep.12884] [Medline: 29464873]
19. Pyke SD. The perceptions of commune health workers about an education strategy received for counselling fathers about infant and breastfeeding involvement. Brock University. 2015. URL: https://dr.library.brocku.ca/bitstream/handle/10464/8684/Brock_Pyke_Stephanie_2015.pdf?sequence=1&isAllowed=y [accessed 2022-04-12]
20. Latif A, Pollock K, Anderson C, Waring J, Solomon J, Chen LC, et al. Supporting underserved patients with their medicines: a study protocol for a patient/professional coproduced education intervention for community pharmacy staff to improve the provision and delivery of Medicine Use Reviews (MURs). BMJ Open 2016 Dec 09;6(12):e013500 [FREE Full text] [doi: 10.1136/bmjopen-2016-013500] [Medline: 27940633]
21. Neher M, Stahl C, Festin K, Nilsen P. Implementation of evidence-based practice in rheumatology: what sociodemographic, social cognitive and contextual factors influence health professionals' use of research in practice? J Rheum Dis Treat 2016;2(3):1510039. [doi: 10.23937/2469-5726/1510039]
22. Unni EJ, Lian N, Kuykendall W. Understanding community pharmacist perceptions and knowledge about HIV preexposure prophylaxis (PrEP) therapy in a Mountain West state. J Am Pharm Assoc (2003) 2016;56(5):527-32.e1. [doi: 10.1016/j.japh.2016.05.004] [Medline: 27594106]
23. Camden C, Rivard LM, Hurtubise K, Héguy L, Berbari J. Can a community of practice improve physical therapists' self-perceived practice in developmental coordination disorder? Phys Ther 2017 Jul 01;97(7):746-755. [doi: 10.1093/ptj/pzx041] [Medline: 28444245]
24. Demers I, Maltais DB, Lamontagne ME, Camden C, Moffet H, Turcotte S. Développement et évaluation d'une activité de transfert de connaissance pour les physiothérapeutes visant à soutenir la participation aux activités physiques chez les enfants ayant un trouble développemental de la coordination. Quebec Provincial Adaptation-Rehabilitation Research Network. 2017. URL: http://old.repar.ca/Fichiers/Subvention/7707_Demande_de_subvention_OPPQ-REPAR_Demers2017_copiefinale.pdf [accessed 2022-04-20]
25. Ludwig C, Bennis C, Carley M, Gifford W, Kuziemsky C, Lafreniere-Davis N, et al. Managing symptoms during cancer treatments: barriers and facilitators to home care nurses using symptom practice guides. Home Health Care Manag Pract 2017 Jun 14;29(4):224-234. [doi: 10.1177/1084822317713011]
26. Manning JC, Carter T, Latif A, Horsley A, Cooper J, Armstrong M, et al. 'Our Care through Our Eyes'. Impact of a co-produced digital educational programme on nurses' knowledge, confidence and attitudes in providing care for children

and young people who have self-harmed: a mixed-methods study in the UK. BMJ Open 2017 May 04;7(4):e014750 [FREE Full text] [doi: 10.1136/bmjopen-2016-014750] [Medline: 28473515]
27. Paulo M. Implementing for success and sustainability. The University of San Francisco. 2017. URL: https://repository.usfca.edu/cgi/viewcontent.cgi?article=1110&context=dnp [accessed 2022-04-12]
28. Portocarrero L, Esther M. Identification des croyances saillantes des femmes enceintes sur l'utilisation d'un outil d'aide à la décision dans le contexte du dépistage prénatal de la trisomie 21 pour l'élaboration et validation d'un questionnaire. Université Laval. 2017. URL: https://corpus.ulaval.ca/jspui/handle/20.500.11794/27801 [accessed 2022-04-12]
29. Carter T, Latif A, Callaghan P, Manning JC. An exploration of predictors of children's nurses' attitudes, knowledge, confidence and clinical behavioural intentions towards children and young people who self-harm. J Clin Nurs 2018 Jul;27(13-14):2836-2846. [doi: 10.1111/jocn.14361] [Medline: 29569381]
30. Lamont S, Brunero S. The effect of a workplace violence training program for generalist nurses in the acute hospital setting: a quasi-experimental study. Nurse Educ Today 2018 Sep;68:45-52. [doi: 10.1016/j.nedt.2018.05.008] [Medline: 29885569]
31. Souli I, Vandyk A, Versailles D, Marcoux I, Salvador A, Peterson WE, et al. Les obstacles et les facilitateurs à l'utilisation de l'évaluation du risque de violence auprès des personnes ayant des troubles mentaux: perceptions du personnel soignant. Rech Soins Infirm 2018 Jun(133):45-57. [doi: 10.3917/rsi.133.0045] [Medline: 30066506]
32. Wagiu Basrowi R, Krisnamurti D, Wibowo Y, Vandenplas Y. Factors influencing probiotics recommendation among pediatricians in Indonesia. Integr Food Nutr Metab 2019;6(5):1-4. [doi: 10.15761/ifnm.1000265]
33. Bentley M, Kerr R, Ginger M, Karagoz J. General practice training Tasmania research note: the social influence on behavioural intentions of doctors and nurses to dementia care in general practice. Soc Influ 2019;5:8-7 [FREE Full text]
34. Bentley MW, Kerr R, Ginger M, Karagoz J. Behavioural change in primary care professionals undertaking online education in dementia care in general practice. Aust J Prim Health 2019 May 29;25(3):244. [doi: 10.1071/PY18079] [Medline: 31138397]
35. Boland L, Lawson ML, Graham ID, Légaré F, Dorrance K, Shephard A, et al. Post-training shared decision making barriers and facilitators for pediatric healthcare providers: a mixed-methods study. Acad Pediatr 2019;19(1):118-129. [doi: 10.1016/j.acap.2018.05.010] [Medline: 29860134]
36. Booth LS. Provider attitudes on using nonpharmacological methods for pain management in acute care - post an educational intervention. University of Arizona. 2019. URL: https://repository.arizona.edu/handle/10150/634386 [accessed 2022-04-12]
37. Cossou-Gbeto I. Évaluation de la formation portant sur l'approche centrée sur la personne dans le cadre du projet d'amélioration de la santé des mères et des enfants. Université Laval. 2019. URL: https://corpus.ulaval.ca/jspui/handle/20.500.11794/37536 [accessed 2019-12-12]
38. Desjardins A. Évaluation de l'intention des médecins de famille enseignants et des résidents en médecine familiale de prescrire et d'interpréter la spirométrie: une étude descriptive transversale. Université Laval. 2019. URL: https://corpus.ulaval.ca/jspui/handle/20.500.11794/33776 [accessed 2022-04-12]
39. Desjardins A, Boulay ME, Gagné M, Simon M, Boulet LP. Family medicine physician teachers and residents' intentions to prescribe and interpret spirometry: a descriptive cross-sectional study. J Asthma 2020 Feb;57(2):149-159. [doi: 10.1080/02770903.2018.1553052] [Medline: 30905218]
40. Devane C, Renner RM, Munro S, Guilbert É, Dunn S, Wagner MS, et al. Implementation of mifepristone medical abortion in Canada: pilot and feasibility testing of a survey to assess facilitators and barriers. Pilot Feasibility Stud 2019;5:126 [FREE Full text] [doi: 10.1186/s40814-019-0520-8] [Medline: 31720004]
41. Dominguez MJ. Increase in pneumococcal immunization in adults over 65 years of age in a federally qualified health center. Liberty University. 2019. URL: https://digitalcommons.liberty.edu/doctoral/1992/ [accessed 2022-04-12]
42. Latif A, Gulzar N, Abbasi N, Choudhary S. Improving provision of Medicine Use Reviews (MURs) to medically under-served groups: a baseline/follow-up comparison study investigating the impact of a co-produced digital learning resource. Int J Pharm Pract 2019 Apr;27(S2):31-32. [doi: 10.1111/ijpp.12533]
43. Latif A, Waring J, Chen LC, Pollock K, Solomon J, Gulzar N, et al. 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 Sep 17;9(9):e031548 [FREE Full text] [doi: 10.1136/bmjopen-2019-031548] [Medline: 31530620]
44. Luconi F, Boillat M, Mak S, Chartrand D, Korah N, Daly M, et al. Patient safety and quality of care are everybody's business: evaluating the impact of a continuing professional development program beyond satisfaction. MedEdPublish 2019;8:46. [doi: 10.15694/mep.2019.000046.1]
45. Ludwig C, Renaud J, Barbera L, Carley M, Henry C, Jolicoeur L, et al. Factors influencing the use by radiation therapists of cancer symptom guides: a mixed-methods study. Curr Oncol 2019 Feb;26(1):56-64 [FREE Full text] [doi: 10.3747/co.25.4198] [Medline: 30853800]
46. Miller KJ, Park JE, Ramanathan K, Abel J, Zhao Y, Mamdani A, et al. Examining coronary revascularization practice patterns for diabetics: perceived barriers, facilitators, and implications for knowledge translation. Can J Cardiol 2020 Aug;36(8):1236-1243. [doi: 10.1016/j.cjca.2019.11.026] [Medline: 32621887]

47. Norman WV, Munro S, Brooks M, Devane C, Guilbert E, Renner R, et al. Could implementation of mifepristone address Canada's urban-rural abortion access disparity: a mixed-methods implementation study protocol. BMJ Open 2019 Apr 20;9(4):e028443 [FREE Full text] [doi: 10.1136/bmjopen-2018-028443] [Medline: 31005943]
48. Stacey D, Ludwig C, Truant T, Carley M, Bennis C, Gifford W, et al. Implementing practice guides to improve cancer symptom management in homecare: a comparative case study. Home Health Care Manag Pract 2018;31(3):139-146. [doi: 10.1177/1084822318817896]
49. Understanding Pain: From Cells to Brain to Individual Perceptions. Canadian Journal of Pain 2019 Apr 02;3(1):A73-A189. [doi: 10.1080/24740527.2019.1599266]
50. Adekpedjou R, Haesebaert J, Stacey D, Brière N, Freitas A, Rivest LP, et al. Variations in factors associated with healthcare providers' intention to engage in interprofessional shared decision making in home care: results of two cross-sectional surveys. BMC Health Serv Res 2020 Mar 12;20(1):203 [FREE Full text] [doi: 10.1186/s12913-020-5064-3] [Medline: 32164669]
51. Boland L, Brosseau L, Caspar S, Graham ID, Hutchinson AM, Kothari A, et al. Reporting health research translation and impact in the curriculum vitae: a survey. Implement Sci Commun 2020;1:20 [FREE Full text] [doi: 10.1186/s43058-020-00021-9] [Medline: 32885181]
52. Fraguas L, Vietto V, Arceo D, Vázquez Peña F, Durante E. Adaptation and validation of the Argentine version of an instrument to assess continuing professional development activities. Arch Argent Pediatr 2020 Apr;118(2):125-129 [FREE Full text] [doi: 10.5546/aap.2020.eng.125] [Medline: 32199047]
53. Giguere AM, Bogza LM, Coudert L, Carmichael PH, Renaud JS, Légaré F, et al. Development of the IcanSDM scale to assess primary care clinicians' ability to adopt shared decision-making. medRxiv. Preprint posted online on July 2, 2020. [doi: 10.1101/2020.07.01.20144204]
54. Poulin Herron A, Agbadje TT, Cote M, Djade CD, Roch G, Rousseau F, et al. Web-based training for nurses on shared decision making and prenatal screening for down syndrome: protocol for a randomized controlled trial. JMIR Res Protoc 2020 Oct 29;9(10):e17878 [FREE Full text] [doi: 10.2196/17878] [Medline: 33118955]
55. Keshmiri F, Rezai M, Tavakoli N. The effect of interprofessional education on healthcare providers' intentions to engage in interprofessional shared decision-making: perspectives from the theory of planned behaviour. J Eval Clin Pract 2020 Aug;26(4):1153-1161. [doi: 10.1111/jep.13379] [Medline: 32114699]
56. Lawani MA, Côté L, Coudert L, Morin M, Witteman HO, Caron D, et al. Professional training on shared decision making with older adults living with neurocognitive disorders: a mixed-methods implementation study. BMC Med Inform Decis Mak 2020 Aug 12;20(1):189 [FREE Full text] [doi: 10.1186/s12911-020-01197-9] [Medline: 32787829]
57. McWilliams S. Addressing compassion fatigue of a family caregiver of adolescents in a mental health setting. Int J Nurs Health Care Res 2020;3:1206. [doi: 10.29011/2688-9501.101206]
58. Peters S, Pfeifer K. Gesundheitsfachleute als Anbieter bewegungsbezogener Interventionen in der gesundheitlichen Versorgung. Deutsche Digitale Bibliothek. 2020. URL: https://www.deutsche-digitale-bibliothek.de/item/NF5TOQ7AFIDA6ZW6NYGNXYCN266S7RK3 [accessed 2022-04-12]
59. Slootweg YM, Walg C, Koelewijn JM, Van Kamp IL, De Haas M. Knowledge, attitude and practices of obstetric care providers towards maternal red-blood-cell immunization during pregnancy. Vox Sang 2020 Apr;115(3):211-220 [FREE Full text] [doi: 10.1111/vox.12883] [Medline: 31884705]
60. Stacey D, Taljaard M, Breau RH, Baba N, Blackmore T, Boland L, et al. A patient decision aid for men with localized prostate cancer: a comparative case study of natural implementation approaches. Cancer Nurs 2020;43(1):E10-E21. [doi: 10.1097/NCC.0000000000000651] [Medline: 30312191]
61. 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;15:1059-1066 [FREE Full text] [doi: 10.2147/CIA.S256416] [Medline: 32753856]
62. Urban KL. Facilitating implementation of fall prevention in primary care practice. University of Arkansas for Medical Sciences. 2020. URL: https://www.proquest.com/openview/7a20fbfa6b30d5b3261d69f6d9c3bd76/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2022-04-12]
63. Yu C, Choi D, Bruno BA, Thorpe KE, Straus SE, Cantarutti P, et al. 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 Sep 30;22(9):e16984 [FREE Full text] [doi: 10.2196/16984] [Medline: 32996893]
64. Abbasgholizadeh Rahimi S, Rodriguez C, Croteau J, Sadeghpour A, Navali A, Légaré F. Continuing professional education of Iranian healthcare professionals in shared decision-making: lessons learned. BMC Health Serv Res 2021 Mar 12;21(1):225 [FREE Full text] [doi: 10.1186/s12913-021-06233-6] [Medline: 33712014]
65. Bergeron L. Identification des facteurs influençant l'intention des médecins spécialistes d'adopter un comportement clinique ciblé par des formations de développement professionnel continu. Université Laval. 2021. URL: https://corpus.ulaval.ca/jspui/handle/20.500.11794/67901?locale=fr [accessed 2022-04-12]
66. Kepper MM, Walsh-Bailey C, Brownson RC, Kwan BM, Morrato EH, Garbutt J, et al. Development of a health information technology tool for behavior change to address obesity and prevent chronic disease among adolescents: designing for

dissemination and sustainment using the ORBIT model. Front Digit Health 2021;3:648777 [FREE Full text] [doi: 10.3389/fdgth.2021.648777] [Medline: 34713122]
67. Lawani MA, Turgeon Y, Côté L, Légaré F, Witteman HO, Morin M, et al. User-centered and theory-based design of a professional training program on shared decision-making with older adults living with neurocognitive disorders: a mixed-methods study. BMC Med Inform Decis Mak 2021 Feb 17;21(1):59 [FREE Full text] [doi: 10.1186/s12911-021-01396-y] [Medline: 33596874]
68. Ridde V, Gautier L, Dagenais C, Chabrol F, Hou R, Bonnet E, et al. Learning from public health and hospital resilience to the SARS-CoV-2 pandemic: protocol for a multiple case study (Brazil, Canada, China, France, Japan, and Mali). Health Res Policy Syst 2021 May 06;19(1):76 [FREE Full text] [doi: 10.1186/s12961-021-00707-z] [Medline: 33957954]
69. Fraguas L, Vietto V, Arceo D, Vázquez Peña F, Durante E. Adaptación y validación de la versión argentina de un instrumento para evaluar actividades de desarrollo profesional continuo. Arch Argent Pediatr 2020;118(2):125-129. [doi: 10.5546/aap.2020.125]
70. Bentley MW, Kerr R, Ginger M, Karagoz J. DCTEP Research Note_June2019.pdf. ResearchGate. 2019 Jun. URL: https://www.researchgate.net/publication/333678470_DCTEP_Research_Note_June2019pdf [accessed 2022-04-20]
71. Martinengo L, Yeo NJ, Markandran KD, Olsson M, Kyaw BM, Car LT. Digital health professions education on chronic wound management: a systematic review. Int J Nurs Stud 2020 Apr;104:103512. [doi: 10.1016/j.ijnurstu.2019.103512] [Medline: 32086027]
72. Ngenzi JL, Scott RE, Mars M. Information and communication technology to enhance continuing professional development (CPD) and continuing medical education (CME) for Rwanda: a scoping review of reviews. BMC Med Educ 2021 Apr 29;21(1):245 [FREE Full text] [doi: 10.1186/s12909-021-02607-w] [Medline: 33926419]
73. Ramani S, McMahon GT, Armstrong EG. Continuing professional development to foster behaviour change: from principles to practice in health professions education. Med Teach 2019 Sep;41(9):1045-1052. [doi: 10.1080/0142159X.2019.1615608] [Medline: 31131672]
74. Sinclair PM, Kable A, Levett-Jones T, Booth D. The effectiveness of Internet-based e-learning on clinician behaviour and patient outcomes: a systematic review. Int J Nurs Stud 2016 May;57:70-81. [doi: 10.1016/j.ijnurstu.2016.01.011] [Medline: 27045566]
75. Légaré F. Le développement professionnel continu en partenariat pour l'application des connaissances. Université Laval. 2017. URL: http://www.decision.chaire.fmed.ulaval.ca/recherche/5d8a3c8fb4956133050a2bad [accessed 2022-04-12]
76. Colquhoun HL, Squires JE, Kolehmainen N, Fraser C, Grimshaw JM. Methods for designing interventions to change healthcare professionals' behaviour: a systematic review. Implement Sci 2017 Mar 04;12(1):30 [FREE Full text] [doi: 10.1186/s13012-017-0560-5] [Medline: 28259168]
77. Dalgetty R, Miller CB, Dombrowski SU. Examining the theory-effectiveness hypothesis: a systematic review of systematic reviews. Br J Health Psychol 2019 May;24(2):334-356. [doi: 10.1111/bjhp.12356] [Medline: 30793445]
78. Patey AM, Hurt CS, Grimshaw JM, Francis JJ. Changing behaviour 'more or less'-do theories of behaviour inform strategies for implementation and de-implementation? A critical interpretive synthesis. Implement Sci 2018 Oct 29;13(1):134 [FREE Full text] [doi: 10.1186/s13012-018-0826-6] [Medline: 30373635]
79. Park YS, Zaidi X, O'Brien BC. RIME foreword: what constitutes science in educational research? Applying rigor in our research approaches. Acad Med 2020 Nov;95(11S Association of American Medical Colleges Learn Serve Lead: Proceedings of the 59th Annual Research in Medical Education Presentations):Si-Sv. [doi: 10.1097/ACM.0000000000003636] [Medline: 32769461]
80. Chen W, Berry A, Drowos J, Lama A, Kleinheksel AJ. Improving the evaluation of faculty development programs. Acad Med 2021 Oct 01;96(10):1496. [doi: 10.1097/ACM.0000000000004151] [Medline: 33951678]
81. Hilty DM, Liu HY, Stubbe D, Teshima J. Defining professional development in medicine, psychiatry, and allied fields. Psychiatr Clin North Am 2019 Sep;42(3):337-356. [doi: 10.1016/j.psc.2019.04.001] [Medline: 31358116]
82. François P, Philibert AC, Esturillo G, Sellier E. [Peer groups: a model for the continuous professional development in general practice]. Presse Med 2013 Jan;42(1):e21-e27. [doi: 10.1016/j.lpm.2012.04.017] [Medline: 22721631]

Abbreviations

CPD: continuing professional development
MeSH: Medical Subject Headings
MMAT: Mixed Methods Appraisal Tool
PICOS: Population, Intervention, Comparator, Outcomes, Study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews

Edited by G Eysenbach, T Leung; submitted 31.01.22; peer-reviewed by M Yordy, L McNamara, S Pesälä; comments to author 25.02.22; revised version received 11.03.22; accepted 21.03.22; published 02.05.22.

Please cite as:
Ayivi-Vinz G, Bakwa Kanyinga F, Bergeron L, Décary S, Adisso ÉL, Zomahoun HTV, Daniel SJ, Tremblay M, Plourde KV, Guay-Bélanger S, Légaré F
Use of the CPD-REACTION Questionnaire to Evaluate Continuing Professional Development Activities for Health Professionals: Systematic Review
JMIR Med Educ 2022;8(2):e36948
URL: https://mededu.jmir.org/2022/2/e36948
doi:10.2196/36948
PMID:35318188

Review

The Opportunities and Challenges of Digital Anatomy for Medical Sciences: Narrative Review

Nilmini Wickramasinghe[1,2], MBA, PhD; Bruce R Thompson[1,3,4], BAppSci, PhD, ; Junhua Xiao[1,5], MBBS, PhD

[1]School of Health Sciences, Swinburne University of Technology, Victoria, Australia
[2]Epworth Healthcare, Melbourne, Australia
[3]Alfred Health, Melbourne, Australia
[4]School of Health Sciences, University of Melbourne, Parkville, Australia
[5]School of Allied Health, La Trobe University, Bundoora, Australia

Corresponding Author:
Junhua Xiao, MBBS, PhD
School of Health Sciences
Swinburne University of Technology
John Street
Hawthorn
Victoria, 3022
Australia
Phone: 61 3 92145042
Email: jxiao@swin.edu.au

Abstract

Background: Anatomy has been the cornerstone of medical education for centuries. However, given the advances in the Internet of Things, this landscape has been augmented in the past decade, shifting toward a greater focus on adopting digital technologies. Digital anatomy is emerging as a new discipline that represents an opportunity to embrace advances in digital health technologies and apply them to the domain of modern medical sciences. Notably, the use of augmented or mixed and virtual reality as well as mobile and platforms and 3D printing in modern anatomy has dramatically increased in the last 5 years.

Objective: This review aims to outline the emerging area of digital anatomy and summarize opportunities and challenges for incorporating digital anatomy in medical science education and practices.

Methods: Literature searches were performed using the PubMed, Embase, and MEDLINE bibliographic databases for research articles published between January 2005 and June 2021 (inclusive). Out of the 4650 articles, 651 (14%) were advanced to full-text screening and 77 (1.7%) were eligible for inclusion in the narrative review. We performed a Strength, Weakness, Opportunity, and Threat (SWOT) analysis to evaluate the role that digital anatomy plays in both the learning and teaching of medicine and health sciences as well as its practice.

Results: Digital anatomy has not only revolutionized undergraduate anatomy education via 3D reconstruction of the human body but is shifting the paradigm of pre- and vocational training for medical professionals via digital simulation, advancing health care. Importantly, it was noted that digital anatomy not only benefits in situ real time clinical practice but also has many advantages for learning and teaching clinicians at multiple levels. Using the SWOT analysis, we described strengths and opportunities that together serve to underscore the benefits of embracing digital anatomy, in particular the areas for collaboration and medical advances. The SWOT analysis also identified a few weaknesses associated with digital anatomy, which are primarily related to the fact that the current reach and range of applications for digital anatomy are very limited owing to its nascent nature. Furthermore, threats are limited to technical aspects such as hardware and software issues.

Conclusions: This review highlights the advances in digital health and Health 4.0 in key areas of digital anatomy analytics. The continuous evolution of digital technologies will increase their ability to reinforce anatomy knowledge and advance clinical practice. However, digital anatomy education should not be viewed as a simple technical conversion and needs an explicit pedagogical framework. This review will be a valuable asset for educators and researchers to incorporate digital anatomy into the learning and teaching of medical sciences and their practice.

(JMIR Med Educ 2022;8(2):e34687) doi:10.2196/34687

KEYWORDS

digital anatomy; digital health; virtual reality; augmented reality; medical education

Introduction

Background

For over 10 years, the relatively nascent domain of digital health [1,2] and the application and embracement of the tools, techniques, and technologies of Industry 4.0 into health care delivery have advanced and matured [3,4]. Much of this advancement has been because of the breakthroughs in the technologies making up the Internet of Things, including mobile and platforms, virtual reality (VR), mixed reality (MR) and augmented reality (AR), 3D printing, analytics, and sensors [3]. Taken together, these technologies are serving to affect the digital transformation of health care delivery [4]; that is, making the interaction with technology, be it mobile solutions, reliance on analytics, sensors, AR, VR, or MR, an integral part of patient and clinician activity in either the receiving or delivery of care. This is being done to support a health care value proposition of better access to care and a higher quality of care and to ensure that a high value of care ensues [5].

In this context, it is natural to see a similar transformation in various aspects of the medical field [6,7]. To date, notable advances in medicine which uses these technologies include orthopedics where 3D printing is now being used to replace or repair body parts such as a broken jaw; robotics is used to facilitate minimal invasive surgery; and analytics visual, imaging, or text in particular is being used extensively in cancer care [8]. One area that has been slow to embrace technological advances has been anatomy, a field that is of critical importance to medicine and the delivery of health care. Today, in most medical schools, anatomy is taught in the traditional fashion with cadavers [9-11], and only a few leading medical schools are venturing into the domain of digital anatomy, where the technologies of the Internet of Things, especially VR, AR, and MR, are used to recreate human structures to support the study of the human body.

Specifically, digital anatomy, or what is often defined as computer-based 3D modeling of the human body [12], is an area of growing importance and significance. Digital anatomy is advancing and becoming more sophisticated given the progress and sophistication in computer modeling, VR, MR and AR. Moreover, the benefits of digital anatomy appear to be far-reaching and provide assistance to students, clinicians, patients, and other stakeholders [12]. In addition, digital anatomy provides a cost-effective approach to realizing high-quality outcomes [13]. This study contends that digital anatomy represents an opportunity to embrace advances in digital health technologies and apply them to the domain of anatomy to enhance and modernize this core area of medical science. This review aims to outline the emerging area of digital anatomy in both teaching and research and importantly summarize opportunities and challenges for incorporating digital anatomy in medical science education and practices, which is particularly pertinent to future medical and health science education. The following then serves to answer the following research questions: "What are the barriers and facilitators for the implementation of the digital technology in anatomy education and research," "How can we embrace digital health technologies to advance the domain of anatomy," and "What are the strengths, weaknesses, opportunities, and threats in which may arise when digital health technologies are incorporated to advance the domain of anatomy?"

Background of Digital Anatomy

Anatomy has over 2000 years of history [14-16] with the first documented scientific description of human structures by Hippocrates of Cos (V-IV centuries BC) [17]. Although *ancient*, this scientific discipline is *born* to be highly adaptive and has undergone checkered changes from its beginning in ancient Greece, Renaissance, to the 19th and now the 21st century [16,18-20]. Modern anatomy teaching is now embracing new innovative modalities. Although human cadavers are viewed as the gold standard for best practices in anatomical education, there is a clear shift from traditional, cadaver-based anatomy teaching toward digital tools–based or mixed (digital tools plus cadavers) curriculum [12,21]. This shift is inevitable, not only driven by an increase in student numbers and financial and ethical constraints on cadaver use but also by the rapid development of medical technology. Digital anatomy is emerging as a new discipline [12,22], representing an intersection of converging disciplines, including medical imaging, 3D reconstruction and printing, AR, and artificial intelligence and robotics.

Digital Anatomy Education Is Fast Developing

An array of digital anatomy tools is currently available with supplementary features [23] and provides curriculum developers with more opportunities to achieve the desired learning outcomes. VR, AR, MR, 3D printing, and tablet-based programs are digital approaches commonly used worldwide for teaching gross and regional anatomy [12,23] either on-site or on the web (web- or cloud-based). When combined to teach the same structures, VR, AR, and tablet anatomy apps have been found to increase learner immersion and engagement [24], and learners develop a deeper understanding of surface anatomy and internal structures relative to their surroundings [25], the latter of which is a desired anatomy learning outcome.

One genuine concern, however, is whether digital anatomy is adequately sufficient for medical education. It is important to appreciate that digital anatomy tools have become more sophisticated, and their fidelity has improved compared with what they were 10 or 5 years ago [12]. Recent digital anatomy resources provide comparative learning outcomes such as understanding of disease and pathology to students, including medical students, compared with cadaver-based education [26-29]. Over 75% of anatomy pedagogy research (126 out of 164 studies) conducted between 2007 and 2017 found that digital technologies enhance anatomical education across multiple disciplines, including medicine, surgery, dentistry, and allied health professions [30]. Virtual or digital dissection tables have been effectively implemented in medical education, providing an alternative experience for understanding human body complexity and layers of internal structures. Junior medical students perceive the use of virtual dissection as a valuable tool for learning anatomy and radiology [27]. Evaluation of the effectiveness of anatomical education using digital technologies, including AR, VR, and MR, supports several pedagogy measurements, including student experience and satisfaction [23,24,28,30,31], learning performance and outcome [25-27,29,31,32], problem-solving skills and clinical reasoning [32,33], and postintervention knowledge and skills outcomes [34-36], with or without comparison with traditional teaching. Overall, these studies support the premise of applying digital anatomy as a means of curriculum development for more surgical-oriented training. The development of digital anatomy cannot replace cadaver-based anatomy teaching but offers unique and sustainable learning experiences and outcomes. Hence, the question of whether digital technology can be used in anatomical education has passed and curriculum developers have now reached a new phase of how to effectively apply digital modalities to learner-centered education that best suit course design and learning outcomes.

Today, digital approaches are combined with blended teaching modes in both face-to-face and remote learning and case- and group-based studies, fostering interdisciplinary collaboration. However, effective digital anatomy education requires curriculum redesign. Digital anatomy education should not be viewed as a simple technical conversion from cadaveric to digital approaches. When designing a digital anatomy curriculum, regardless of the discipline, content delivery needs to be tightly linked with the chosen digital tools. Careful alignment between learning tasks and performance measures using digital tools is required [37]. Different digital anatomy tools come with supplementary features; hence, their selection needs to best meet the graduates' attributes. Moreover, anatomy education, either virtual or classical, requires trained personnel and time. However, the decline in trained anatomy instructors is an ongoing challenge in this discipline [14]. When moving into the digital phase, the competence and attitude of the staff toward using digital tools in teaching and research is another key factor in the success of digital anatomy course delivery and development.

Digital Anatomy Research Is Moving Into a New Dimension

Anatomy is far more than *landmarks*. Many medical advances have been made in anatomical research. Specialized medical imaging techniques such as computed tomography scans and magnetic resonance imaging have revolutionized health care service quality through better visualization of patients' anatomy, including pathology. Together with physicians, surgeons, radiologists, and computational scientists, anatomists can now adopt the latest digital technologies to promulgate a questioning scientific spirit. The human body is a typical example of organism diversity and variation, but 2D slices are no longer sufficient to provide a full picture that can guide clinical management and regimes. In this context, digital anatomy research driven by clinical problems has established a new research dimension in this discipline. 3D-printed anatomy models not only function as a training tool for both health professionals and patients [38,39] but also advance the development of tissue engineering [40-42] and patient-specific medical devices [43,44], which have lifesaving potential for complex cases. Moreover, the digital processing of anatomical data provides a precise representation of the patient's anatomy. Indeed, the digitization of patient-specific anatomy has gained momentum in recent years [45-47] and has moved beyond 3D printing [46,48]. It is envisaged that patient-specific digital anatomy could be a catalyst for changing and bridging the quality of personalized care. AR merges virtual patient-specific anatomy into a real surgical view. By adopting 3D modeling and preoperative virtual planning, anatomists can generate 3D reconstructed organs for individual patients before surgery [45,46] and enable statistical modeling that can accurately predict posttraumatic or postsurgical conditions [47]. Therefore, patient-specific anatomy will allow for more accurate risk evaluation of invasive surgery [47,49] and provide advanced personalized patient care. In this context, digital anatomy will revolutionize our understanding of anatomical variations and effectively apply them in clinical settings. From a training perspective, this new field of digital anatomy research may also enable the rehearsal of virtual surgeries, such as virtual transplantation. In the future, patient-specific digital anatomy research could be successfully integrated with big data analytics and deep learning [8]. In so doing, the outcomes of digital anatomy research will be transformational, not only applying to personalized surgeries for many organs and digital health but also providing unprecedented insights into health care advancement.

Digital Anatomy in Clinical Practice: A Shifting Paradigm for Future Health Care Education

The rapid development of digital anatomy will not only revolutionize undergraduate anatomy education but also shift the paradigm of pre- and vocational training for health care professionals. New clinical skills that require digital technology have evolved for advanced treatments [50]. However, restricted operative opportunities for medical and surgical trainees remain a challenging issue in clinical training around the world because of financial, resource, and other logistical constraints and the recent COVID-19 pandemic [51]. Furthermore, traditional cadaver-based clinical training does not address the need for growing health care advancements. Thus, efforts have been made to provide digital-based operative education outside the theater and cadaver laboratory, while reducing the variability in trainees' operative experience and skills.

Digital-based clinical simulation has been delivered for new and advanced practices, with consistent positive effects on knowledge, skills, and professional behaviors. Digital simulations with improved fidelity such as haptic features are being applied at all levels of learners to a wide range of surgical training, including ear, nose, or throat [52,53]; orthopedic [54]; vascular [55]; ophthalmic [56-58], and neurosurgeries [59-63], involving endoscopy [61,64-68], laparoscopy [69], and robotics [8]. The AR or VR anatomy teaching of health professionals shows clear benefits of improving surgical confidence [70], performance [63], and postintervention knowledge and skills [34-36], particularly in complex conditions such as neurological and cardiovascular diseases [33]. These studies demonstrate that digital simulation for surgical anatomy training is feasible and provides depth perception of surgical procedures, which is particularly pertinent for training novice or inexperienced surgical residents [63,71]

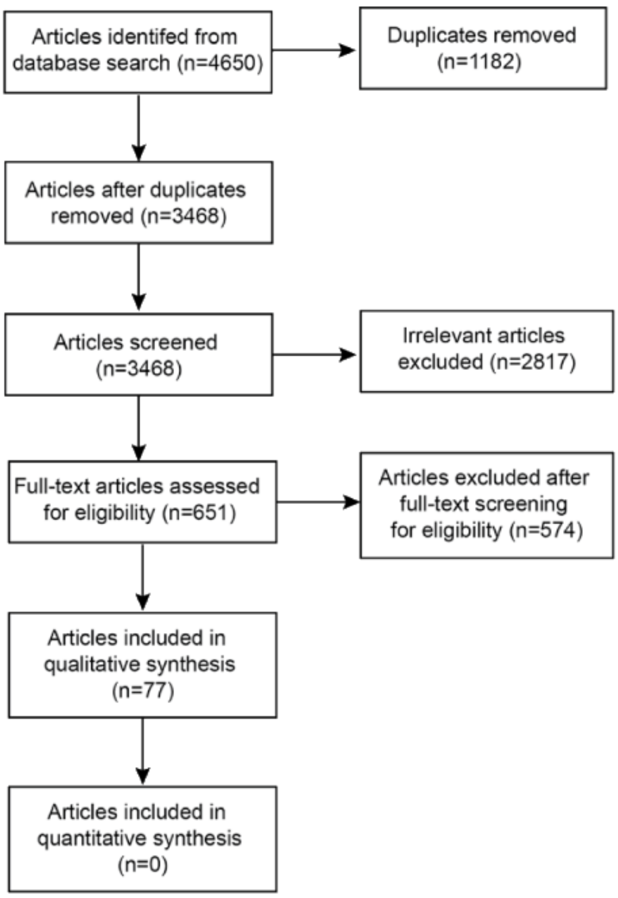
Moreover, a virtual surgical simulator has been developed for cleft repair surgical education [72,73]. Virtual technology has been found to increase nursing students' clinical skills without harming patients and help prepare nurses for new practices such as robotic surgery [74]. Together, digital-based clinical skill training has been established as a new training model for health professionals, improving postintervention knowledge and skill outcomes via a virtual and immersive environment when compared with traditional education [36].

Modern health care practitioners, from nurses to physicians and allied health professionals, are required to work effectively in teams and rely on each other's expertise to provide holistic and optimal patient-centered management. Digital health has substantially improved the way health professionals handle clinical phases. Recent advances in the digital acquisition of patient-specific anatomy data can provide substantial information to the clinician; hence, complex surgeries and less invasive local therapies can be easily planned [45-47,49]. Thus, digital anatomy–based clinical training and practice will continue to address restrictions and reduce disparities in surgical training. This is expected to have a significant impact on future precision surgery and development of collaborative global curricula in surgical education. It is possible that digital anatomy laboratories may become part of the operation in future hospitals, and health professionals in different disciplines, including medical and surgical residents and fellows, could review regions pertinent to their specialties.

Methods

Literature searches were performed using the PubMed, Embase, and MEDLINE bibliographic databases for research articles published between January 2005 and June 2021 (inclusive; see Figure 1 for the search flow diagram). Combinations of the following search terms and subheadings were considered appropriate for this investigation: *anatomy, digital anatomy, virtual reality, augmented reality, mixed reality, apps, teaching, education, anatomy,* and *training.* The publications chosen were restricted to those written in English that described human anatomical education or training within the health and medical sciences. Articles published in the fields of nonhuman anatomy education and microscopic anatomy were excluded from the study. The primary aim of this study was to identify the barriers to and facilitators for the implementation of digital anatomy education and research. No ethical clearance was required for this study because all selected studies had previously received ethics approval from local institutional review boards. A Strength, Weakness, Opportunity, and Threat (SWOT) analysis (descriptive analysis) was performed to evaluate the roles that digital anatomy plays in both the learning and teaching of medicine and health sciences as well as its practice. A formal meta-analysis was not performed owing to the heterogeneity of the retrieved data.

Figure 1. A summary of selected articles shown in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.



Results

Overview

A move toward digital anatomy is inevitable. As with the development of any new discipline, there are both strengths and weaknesses associated with digital anatomy, accompanied by potential threats and opportunities, as shown by the SWOT analysis (Table 1).

Table 1. Matrix analysis of digital anatomy.

Strategy	Strength	Weakness	Threat	Opportunity
Cost	Cost-effective and low maintenance over the long term (personnel and resources) [13]	Considerable initial setup cost such as hardware devices	—[a]	Digital anatomy resources more accessible
Education setting	Flexible; time-efficient; low maintenance required for manual handling and occupational safety	Virtual; digital tools not always compatible with existing infrastructure or teaching settings in the cadaver laboratory	Hardware and software upgrade; the cost of changing; conservative thinkers	New technical development and software upgrade available; compatible with increase in student numbers and the demand of remote learning; new options for future clinical skill laboratory in the hospital
Learner experience	Combine surface and regional anatomy [23]; consistent learner satisfaction [23,24,28,30,31]; better visualizing deeper structures incorporating virtual dissection; integrating anatomy, physiology, and pathology; integrating gross and microscopic anatomy with medical imaging in one setting	Currently limited on showing anatomical variations; current virtual dissection has lack of tactile information; shortfall in learner-centered digital technologies in health care education [32]	Variable digital competencies of users (instructor and student) [30]; limited exposure to human body variability	Augmented reality and virtual reality resources more sophisticated [12] with supplementary features [23]; digital and haptic technologies are being integrated for surgical anatomy training [75,76]; embrace anatomy learning with new medical technology; enable discipline-specific learning [64,65]
Learning outcome	Enable streamed group-based study on the same anatomical structure (not possible on a single cadaver or model) [23]; cognitive skills and memory retention; postintervention knowledge and skills outcomes [36]; unique attributes to safe clinical practice [74]; improve clinical reasoning [32,33,77]	Currently lacked explicit pedagogical framework	Limited education opportunity for learners' feelings about death; potential lack of traditional surgical skills training; impact of new digital anatomy curricula on future surgical competencies unclear	Allow vertical integration of surgical anatomy through advanced curricula; enable training for new and advanced practices [74,78]; advance patient-specific anatomy for personalized health care and training [46]; learner-centered health care education [32]
Collaboration and medical advances	Accessible for users' self-revision; enable flexible and rapid curriculum change; address restrictions and reduce disparities in surgical training; improve informed patient consent and education for surgical planning [38,39]; enable and catalyze resource sharing and collaboration at all levels of training and practice	—	—	Enable sophisticated preoperative study [45-47,49]; trigger curriculum redesign; foster new collaborative graduate courses [79]; catalyze new specialties and medical advances; advance personalized patient health care [43,45-47]; integrating into digital health

[a]No data available.

Cost and Class Setting

From an economic perspective, the implementation of digital anatomy has a clear and cost-effective impact. The use of digital techniques would dramatically reduce the ongoing cost spent on classes, books, mannequins, cadavers (procurement, preparation, and disposal), and ventilation systems compared with traditional anatomy teachings, estimated at US $390,000 every 5 years [13]. Moreover, the number of digital technology options continues to increase. Hence, the cost of such equipment is likely to decrease over time because of the rise in competition. Although there is considerable initial setup cost, particularly for hardware devices, the overall maintenance is low over the long term [13]. Moreover, digital anatomy laboratories require minimum manual handling and ventilation monitoring because of occupational fixative exposure, both of which are ongoing work-related safety hazards in cadaver laboratories. Thus, the class setting and turnover of the digital anatomy laboratory are flexible, time-efficient, and readily accessible for both learners and instructors. In addition, many digital anatomy resources allow web- or cloud-based access, thereby allowing the development and delivery of new courses for remote learning without significant administrative investment. In light of the lengthy global pandemic, digital anatomy represents unique advantages for class setting change, resource development, and new course development during both the COVID and post-COVID periods. The rapid development of digital technology together with patient-specific anatomy research will make virtual surgical planning and rehearsal room possible in the future [46]. Overall, digital anatomy can provide cost-effective education at a time when the demands on training and health care services continue to increase.

Medical Education: Learning Experience and Outcome

Digital technology–based education has demonstrated consistent learner satisfaction, outperforming the traditional approaches [23,24,28,30,31]. However, representations of human body variations are limited to the current digital resource capacities. Thus, there is a need to expand the repertoire of digitized cadaveric resources, particularly for the appreciation of human body variations and diversity. New digital resources are being rapidly developed and have shown a good correlation between cadaveric materials and digital resources, such as for coronary artery distribution [80]. When combining AR with virtual dissection, digital anatomy has clear advantages for visualizing internal deeper structures and integrating anatomy, physiology, pathology, and medical imaging for the same structure at both gross and microscopic levels (Table 1). This ultimately enables a streamed group-based study of the same anatomical structure, which is not possible with a single cadaver or model [23]. When implemented for simulation classes, blended digital anatomy teaching enhances postintervention knowledge and skills [36].

Competent clinicians, particularly surgeons, need a deep understanding of anatomy for safe clinical procedures. In this context, digital anatomy has unique attributes for clinical training, such as patient safety, postintervention knowledge and skill outcomes, and real-life conditions without time limitations and patient discomfort [36,74]. However, there is a concern that medical students without exposure to cadaver-based dissection could be less competent in surgical skills and have limited opportunities for learning about death; hence, they are poorly prepared when entering clerkships and residency programs. The cadaver-based dissection class, limited by the number of bodies, means to preserve them, and associated logistics, is an ongoing and worldwide challenge in medical and surgical education. In this context, virtual or digital dissection provides an alternative and valuable learning experience for medical students [27] when classic dissection classes are not accessible. Moreover, the ongoing and rapid development of sophisticated digital tools with high fidelity, such as digital anatomy education incorporating haptic technologies [76], will significantly advance new clinical practices such as robotic surgery for health professionals.

That said, the impact of new digital anatomy curriculum reforms on the retention of future surgical competencies is currently unclear (Table 1). Therefore, future research is required to evaluate students' perceptions and effectiveness of digital simulation in the satisfactory development of surgical skills. In addition, proficiency in surgical care is complex, as it not only involves knowledge of instrumentation and surgical procedures but also a comprehensive integration of anatomy and physiology for the organ being operated. In this context, digital anatomy resources provide sophisticated tools for preoperative study, particularly complex surgical planning [45-47,49] (Table 1). Although digital anatomy has unique attributes that can improve learning outcomes when compared with traditional learning methods [32], there is currently a shortfall in learner-centered implementation of digital technologies in health care education, where these technologies have the capacity to cause a paradigm shift (Table 1).

The digital competencies of users (both instructors and students) and their attitude toward technology is another key element that heavily influences curriculum effectiveness (Table 1). Researchers might be overconfident in the use of digital technologies [30], highlighting the importance of adequate technical training for instructors. VR and AR technologies display supplementary features and suit different purposes of anatomy teaching and research (Table 1). VR was found to be the most prevalent and influential digital technology, followed by web-based and computer-aided resources [30]. Although AR and VR are relatively mature technologies suitable for surface and regional applications, MR is still a developing technology and is not necessarily consumer-ready at this point of time [23]. However, research in the educational setting shows great promise in its potential to allow multiple users to visualize the same structure if the headsets communicate with each other [23].

Discussion

Using SWOT analysis (Table 1), we described strengths and opportunities that together serve to underscore the benefits of embracing digital anatomy, particularly in the areas of medical education and advances. The SWOT analysis also identified a few weaknesses associated with digital anatomy, which are primarily related to the fact that the current reach and range of applications for digital anatomy are very limited because of its nascent nature. Furthermore, threats are limited to technical aspects such as hardware and software issues. This finding highlights the advances in digital health and Health 4.0 in key areas of digital anatomy analytics.

Digital health is still quite nascent; however, the benefits of the application of advances in technology and the development of technologies that make up the Internet of Things to health care are significant and difficult to quantify. Many aspects of digital health, such as telehealth, have recently become very important given the recent COVID pandemic but before this, although recognized as having many advantages, had yet not received universal appeal [4]. It is in this context that applying such advances to the area of anatomy, the domain that has for the most part remained quite traditional in not only the way it is taught to medical students but also in how clinicians typically refer to anatomy in clinical consultations or in the operating theater is considered. Specifically, the broadening of anatomy is recommended to include consideration of digital anatomy. From a technical perspective, digital anatomy is not a major challenge given that the technologies of AR, VR and MR are well developed and easily transferable to this context as are other techniques and technologies around analytics and even 3D printing. However, digital anatomy education should not be viewed as a simple technical conversion and needs an explicit pedagogical framework. It is more challenging to apply these technological advances to relatively traditional domains. This requires changes in the processes and mindsets of key stakeholders, policymakers, regulatory bodies, and advocate groups. We assert that it is necessary to embrace digital advances in anatomy and incorporate digital anatomy into various contexts such as education and clinical practice so that it will be possible to realize the benefits, strengths, and opportunities identified in Table 1 that digital anatomy affords. This will serve not only to advance digital health but also to provide a better educational experience and understanding of anatomy for clinicians, which in turn will translate into better patient outcomes.

The preceding section outlined the emerging area of digital anatomy. This was done by first highlighting the advances in digital health and Health 4.0 in key areas of analytics and AR, MR and VR as well as mobile and platforms and 3D printing. From this, the impact of analytics and AR, MR and VR in particular can be harnessed to enable the vision of digital anatomy to be realized. Importantly, it was noted that digital anatomy cannot only benefit in situ real time clinical practice but also has many advantages for learning and teaching clinicians at multiple levels.

To unpack this further, we presented a SWOT analysis of opportunities for incorporating digital anatomy. It is apparent from Table 1 that there are very few weaknesses with digital anatomy, most of which are related to the fact that the current reach and range of applications for digital anatomy are very limited because of its nascent nature. Furthermore, the threats are limited to technical aspects such as hardware and software issues; given that hardware and software costs not only continue to decrease over time but also witness significant improvements and advances, we are of the opinion that this threat, although minimal, will diminish further in the fullness of time. The continuous evolution of digital technologies will increase their ability to reinforce anatomy knowledge and advance clinical practice. Similarly, we are confident that digital literacy and proficiency of students and clinicians will increase over time. Moreover, we are confident that expectations from students and clinicians regarding digital anatomy will increase over time. What is particularly pleasing is that Table 1 clearly highlights many and multiple strengths and opportunities that taken together serve to underscore the benefits of embracing digital anatomy. Future research should evaluate the long-term effectiveness of digital anatomy and its impact on multiple domains, such as changes in learners' practice, behavior, and skills. There are also some limitations associated with this study. Qualitative analysis was performed on the included studies; however, a meta-analysis was not performed because of the heterogeneity of the retrieved data.

We set out to create a case for digital anatomy. Table 1 lists its merits and benefits. However, that is in reality only the beginning. For digital anatomy to be fully embraced and realized, it is necessary for policy to be affected to encourage and support its adoption. Our future work will focus on this in detail, but areas of consideration include support for digital anatomy from health care advocate bodies, medical education bodies, and leading health care societies such as the Health Informatics Management Systems Society or the Australasian Institute of Digital Health. In addition, careful attention should be paid to the development of appropriate standards and quality for digital anatomy construction, design, and development. We believe that this is an exciting, emerging area that holds the promise of significant advances for both the learning and teaching of medicine as well as the practice of medicine and close by calling for more research in this key area.

Conflicts of Interest

None declared.

References

1. Nikolic A, Wickramasinghe N, Claydon-Platt D, Balakrishnan V, Smart P. The use of communication apps by medical staff in the Australian health care system: survey study on prevalence and use. JMIR Med Inform 2018 Feb 09;6(1):e9. [doi: 10.2196/medinform.9526] [Medline: 29426813]
2. Wickramasinghe N, John B, George J, Vogel D. Achieving value-based care in chronic disease management: intervention study. JMIR Diabetes 2019 May 03;4(2):e10368 [FREE Full text] [doi: 10.2196/10368] [Medline: 31066699]
3. Chen CE, Harrington RA, Desai SA, Mahaffey KW, Turakhia MP. Characteristics of digital health studies registered in ClinicalTrials.gov. JAMA Intern Med 2019 Jun 01;179(6):838-840 [FREE Full text] [doi: 10.1001/jamainternmed.2018.7235] [Medline: 30801617]
4. Kansal V, Ranjan R, Sinha S, Tiwari R, Wickramasinghe N. Healthcare and Knowledge Management for Society 5.0: Trends, Issues, and Innovations. Boca Raton, FL, USA: CRC Press; 2021.
5. Wickramasinghe N. Handbook of Research on Optimizing Healthcare Management Techniques. Hershey, PA, USA: IGI Global; 2020.
6. Sadoughi F, Behmanesh A, Sayfouri N. Internet of things in medicine: a systematic mapping study. J Biomed Inform 2020 Mar;103:103383 [FREE Full text] [doi: 10.1016/j.jbi.2020.103383] [Medline: 32044417]
7. Sriram RD, Subrahmanian E. Transforming health care through digital revolutions. J Indian Inst Sci 2020;100(4):753-772 [FREE Full text] [doi: 10.1007/s41745-020-00195-0] [Medline: 33132546]
8. Bhandari M, Zeffiro T, Reddiboina M. Artificial intelligence and robotic surgery: current perspective and future directions. Curr Opin Urol 2020 Jan;30(1):48-54. [doi: 10.1097/MOU.0000000000000692] [Medline: 31724999]

9. Naidoo N, Akhras A, Banerjee Y. Confronting the challenges of anatomy education in a competency-based medical curriculum during normal and unprecedented times (COVID-19 pandemic): pedagogical framework development and implementation. JMIR Med Educ 2020 Oct 07;6(2):e21701 [FREE Full text] [doi: 10.2196/21701] [Medline: 32873536]
10. Pan S, Chan LK, Yan Y, Yang X. Survey of gross anatomy education in China: the past and the present. Anat Sci Educ 2020 May;13(3):390-400. [doi: 10.1002/ase.1952] [Medline: 32107879]
11. Smith CF, Freeman SK, Heylings D, Finn GM, Davies DC. Anatomy education for medical students in the United Kingdom and Republic of Ireland in 2019: a 20-year follow-up. Anat Sci Educ (forthcoming) 2021 Jul 27. [doi: 10.1002/ase.2126] [Medline: 34314569]
12. Uhl JF, Jorge J, Lopes DS, Campos PF. Digital Anatomy: Applications of Virtual, Mixed and Augmented Reality. Cham, Switzerland: Springer; 2021.
13. Duarte ML, Santos LR, Guimarães Júnior JB, Peccin MS. Learning anatomy by virtual reality and augmented reality. A scope review. Morphologie 2020 Dec;104(347):254-266. [doi: 10.1016/j.morpho.2020.08.004] [Medline: 32972816]
14. Sugand K, Abrahams P, Khurana A. The anatomy of anatomy: a review for its modernization. Anat Sci Educ 2010;3(2):83-93. [doi: 10.1002/ase.139] [Medline: 20205265]
15. Alghamdi MA, Ziermann JM, Diogo R. An untold story: the important contributions of Muslim scholars for the understanding of human anatomy. Anat Rec (Hoboken) 2017 Jun;300(6):986-1008 [FREE Full text] [doi: 10.1002/ar.23523] [Medline: 27875640]
16. Conti AA, Paternostro F. Anatomical study in the Western world before the Middle Ages: historical evidence. Acta Biomed 2019 Dec 23;90(4):523-525 [FREE Full text] [doi: 10.23750/abm.v90i4.8738] [Medline: 31910179]
17. Smith WD. The Hippocratic Tradition. Ithaca, NY, USA: Cornell University Press; 1979.
18. Laurenza D. Leonardo's contributions to human anatomy. Lancet 2019 Apr 06;393(10179):1473-1476. [doi: 10.1016/S0140-6736(19)30716-0] [Medline: 30967218]
19. Shaffer K. Teaching anatomy in the digital world. N Engl J Med 2004 Sep 23;351(13):1279-1281. [doi: 10.1056/NEJMp048100] [Medline: 15385652]
20. Delmas V, Uhl JF, Campos PF, Lopes DS, Jorge J. From anatomical to digital dissection: a historical perspective since antiquity towards the twenty-first century. In: Uhl JF, Jorge J, Lopes DS, Campos PF, editors. Digital Anatomy: Applications of Virtual, Mixed and Augmented Reality. Cham, Switzerland: Springer; 2021:11-39.
21. Drake RL, McBride JM, Lachman N, Pawlina W. Medical education in the anatomical sciences: the winds of change continue to blow. Anat Sci Educ 2009;2(6):253-259. [doi: 10.1002/ase.117] [Medline: 19890982]
22. UNESCO Digital Anatomy Chair. 2015. URL: https://anatomieunescoorg/ [accessed 2022-05-12]
23. Stirling AC, Moro C. The use of virtual, augmented and mixed reality in anatomy education. In: Chan LK, Pawlina W, editors. Teaching Anatomy: A Practical Guide. Cham, Switzerland: Springer; 2020:359-366.
24. Moro C, Štromberga Z, Raikos A, Stirling A. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. Anat Sci Educ 2017 Nov;10(6):549-559. [doi: 10.1002/ase.1696] [Medline: 28419750]
25. Deng X, Zhou G, Xiao B, Zhao Z, He Y, Chen C. Effectiveness evaluation of digital virtual simulation application in teaching of gross anatomy. Ann Anat 2018 Jul;218:276-282. [doi: 10.1016/j.aanat.2018.02.014] [Medline: 29679721]
26. 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 Oct 07;19(1):366 [FREE Full text] [doi: 10.1186/s12909-019-1806-5] [Medline: 31590672]
27. Darras KE, Forster BB, Spouge R, de Bruin AB, Arnold A, Nicolaou S, et al. Virtual dissection with clinical radiology cases provides educational value to first year medical students. Acad Radiol 2020 Nov;27(11):1633-1640. [doi: 10.1016/j.acra.2019.09.031] [Medline: 31786075]
28. Moro C, Birt J, Stromberga Z, Phelps C, Clark J, Glasziou P, et al. Virtual and augmented reality enhancements to medical and science student physiology and anatomy test performance: a systematic review and meta-analysis. Anat Sci Educ 2021 May;14(3):368-376. [doi: 10.1002/ase.2049] [Medline: 33378557]
29. Zhao J, Xu X, Jiang H, Ding Y. The effectiveness of virtual reality-based technology on anatomy teaching: a meta-analysis of randomized controlled studies. BMC Med Educ 2020 Apr 25;20(1):127 [FREE Full text] [doi: 10.1186/s12909-020-1994-z] [Medline: 32334594]
30. Pringle Z, Rea PM. Do digital technologies enhance anatomical education? Pract Evid Scholar Teach Learn High Educ 2018 May;13(1):2-27. [doi: 10.4324/9780429020629-3]
31. Triepels CP, Smeets CF, Notten KJ, Kruitwagen RF, Futterer JJ, Vergeldt TF, et al. Does three-dimensional anatomy improve student understanding? Clin Anat 2020 Jan;33(1):25-33 [FREE Full text] [doi: 10.1002/ca.23405] [Medline: 31087400]
32. Logeswaran A, Munsch C, Chong YJ, Ralph N, McCrossnan J. The role of extended reality technology in healthcare education: towards a learner-centred approach. Future Healthc J 2021 Mar;8(1):e79-e84 [FREE Full text] [doi: 10.7861/fhj.2020-0112] [Medline: 33791482]
33. Stromberga Z, Phelps C, Smith J, Moro C. Teaching with disruptive technology: the use of augmented, virtual, and mixed reality (HoloLens) for disease education. Adv Exp Med Biol 2021;1317:147-162. [doi: 10.1007/978-3-030-61125-5_8] [Medline: 33945136]

34. Gerup J, Soerensen CB, Dieckmann P. Augmented reality and mixed reality for healthcare education beyond surgery: an integrative review. Int J Med Educ 2020 Jan 18;11:1-18 [FREE Full text] [doi: 10.5116/ijme.5e01.eb1a] [Medline: 31955150]
35. Zhu E, Hadadgar A, Masiello I, Zary N. Augmented reality in healthcare education: an integrative review. PeerJ 2014 Jul 8;2:e469 [FREE Full text] [doi: 10.7717/peerj.469] [Medline: 25071992]
36. 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] [Medline: 30668519]
37. Saltarelli AJ, Roseth CJ, Saltarelli WA. Human cadavers vs. multimedia simulation: a study of student learning in anatomy. Anat Sci Educ 2014;7(5):331-339. [doi: 10.1002/ase.1429] [Medline: 24415563]
38. Eisenmenger LB, Wiggins 3rd RH, Fults 3rd DW, Huo EJ. Application of 3-dimensional printing in a case of osteogenesis imperfecta for patient education, anatomic understanding, preoperative planning, and intraoperative evaluation. World Neurosurg 2017 Nov;107:1049.e1-1049.e7. [doi: 10.1016/j.wneu.2017.08.026] [Medline: 28823657]
39. Zhuang YD, Zhou MC, Liu SC, Wu JF, Wang R, Chen CM. Effectiveness of personalized 3D printed models for patient education in degenerative lumbar disease. Patient Educ Couns 2019 Oct;102(10):1875-1881. [doi: 10.1016/j.pec.2019.05.006] [Medline: 31113688]
40. Ostrovidov S, Salehi S, Costantini M, Suthiwanich K, Ebrahimi M, Sadeghian RB, et al. 3D bioprinting in skeletal muscle tissue engineering. Small 2019 Jun;15(24):e1805530 [FREE Full text] [doi: 10.1002/smll.201805530] [Medline: 31012262]
41. Jamróz W, Szafraniec J, Kurek M, Jachowicz R. 3D printing in pharmaceutical and medical applications - recent achievements and challenges. Pharm Res 2018 Jul 11;35(9):176 [FREE Full text] [doi: 10.1007/s11095-018-2454-x] [Medline: 29998405]
42. Di Gesù R, Acharya AP, Jacobs I, Gottardi R. 3D printing for tissue engineering in otolaryngology. Connect Tissue Res 2020 Mar;61(2):117-136. [doi: 10.1080/03008207.2019.1663837] [Medline: 31524001]
43. Thayaparan GK, Owbridge MG, Thompson RG, D'Urso PS. Designing patient-specific 3D printed devices for posterior atlantoaxial transarticular fixation surgery. J Clin Neurosci 2018 Oct;56:192-198. [doi: 10.1016/j.jocn.2018.06.038] [Medline: 30017594]
44. Noecker AM, Chen JF, Zhou Q, White RD, Kopcak MW, Arruda MJ, et al. Development of patient-specific three-dimensional pediatric cardiac models. ASAIO J 2006;52(3):349-353. [doi: 10.1097/01.mat.0000217962.98619.ab] [Medline: 16760727]
45. Soler L, Mutter D, Pessaux P, Marescaux J. Patient specific anatomy: the new area of anatomy based on computer science illustrated on liver. J Vis Surg 2015 Nov 23;1:21 [FREE Full text] [doi: 10.3978/j.issn.2221-2965.2015.11.06] [Medline: 29075611]
46. Soler L, Mutter D, Marescaux J. Patient-specific anatomy: the new area of anatomy based on 3D modelling. In: Uhl JF, Jorge J, Lopes DS, Campos PF, editors. Digital Anatomy: Applications of Virtual, Mixed and Augmented Reality. Cham, Switzerland: Springer; 2021:285-297.
47. Vlachopoulos L, Lüthi M, Carrillo F, Gerber C, Székely G, Fürnstahl P. Restoration of the patient-specific anatomy of the proximal and distal parts of the humerus: statistical shape modeling versus contralateral registration method. J Bone Joint Surg Am 2018 Apr 18;100(8):e50. [doi: 10.2106/JBJS.17.00829] [Medline: 29664855]
48. Kaschwich M, Horn M, Matthiensen S, Stahlberg E, Behrendt CA, Matysiak F, et al. Accuracy evaluation of patient-specific 3D-printed aortic anatomy. Ann Anat 2021 Mar;234:151629. [doi: 10.1016/j.aanat.2020.151629] [Medline: 33137459]
49. Schieferbein V, Bredemann J, Schmitt R, Stenin I, Klenzner T, Schipper J, et al. Influence of patient-specific anatomy on medical computed tomography and risk evaluation of minimally invasive surgery at the otobasis. Eur Arch Otorhinolaryngol 2019 Feb;276(2):375-382. [doi: 10.1007/s00405-018-5249-0] [Medline: 30554360]
50. Konttila J, Siira H, Kyngäs H, Lahtinen M, Elo S, Kääriäinen M, et al. Healthcare professionals' competence in digitalisation: a systematic review. J Clin Nurs 2019 Mar;28(5-6):745-761. [doi: 10.1111/jocn.14710] [Medline: 30376199]
51. Montanari S. Body of Knowledge. The Slate Group. 2020 Jun 23. URL: https://slate.com/technology/2020/06/med-school-cadaver-dissection-virtual-reality.html2020 [accessed 2022-05-12]
52. Piromchai P, Avery A, Laopaiboon M, Kennedy G, O'Leary S. Virtual reality training for improving the skills needed for performing surgery of the ear, nose or throat. Cochrane Database Syst Rev 2015 Sep 09(9):CD010198. [doi: 10.1002/14651858.CD010198.pub2] [Medline: 26352008]
53. Huang C, Cheng H, Bureau Y, Agrawal SK, Ladak HM. Face and content validity of a virtual-reality simulator for myringotomy with tube placement. J Otolaryngol Head Neck Surg 2015 Oct 20;44:40 [FREE Full text] [doi: 10.1186/s40463-015-0094-2] [Medline: 26481401]
54. Brouwers L, Pull Ter Gunne AF, de Jongh MA, Maal TJ, Vreeken R, van der Heijden FH, et al. What is the value of 3D virtual reality in understanding acetabular fractures? Eur J Orthop Surg Traumatol 2020 Jan;30(1):109-116. [doi: 10.1007/s00590-019-02537-w] [Medline: 31531739]
55. Agarwal N, Schmitt PJ, Sukul V, Prestigiacomo CJ. Surgical approaches to complex vascular lesions: the use of virtual reality and stereoscopic analysis as a tool for resident and student education. BMJ Case Rep 2012 Aug 01;2012:bcr0220125859 [FREE Full text] [doi: 10.1136/bcr.02.2012.5859] [Medline: 22859380]
56. Chhaya N, Helmy O, Piri N, Palacio A, Schaal S. Comparison of 2D and 3D video displays for teaching vitreoretinal surgery. Retina 2018 Aug;38(8):1556-1561. [doi: 10.1097/IAE.0000000000001743] [Medline: 28700420]

57. Famery N, Abdelmassih Y, El-Khoury S, Guindolet D, Cochereau I, Gabison EE. Artificial chamber and 3D printed iris: a new wet lab model for teaching Descemet's membrane endothelial keratoplasty. Acta Ophthalmol 2019 Mar;97(2):e179-e183 [FREE Full text] [doi: 10.1111/aos.13852] [Medline: 30259685]
58. Pahuta MA, Schemitsch EH, Backstein D, Papp S, Gofton W. Virtual fracture carving improves understanding of a complex fracture: a randomized controlled study. J Bone Joint Surg Am 2012 Dec 19;94(24):e182. [doi: 10.2106/JBJS.K.00996] [Medline: 23318621]
59. Clifton W, Damon A, Stein R, Pichelmann M, Nottmeier E. Biomimetic 3-dimensional-printed posterior cervical laminectomy and fusion simulation: advancements in education tools for trainee instruction. World Neurosurg 2020 Mar;135:308. [doi: 10.1016/j.wneu.2019.12.134] [Medline: 31899386]
60. Coelho G, Rabelo NN, Vieira E, Mendes K, Zagatto G, Santos de Oliveira R, et al. Augmented reality and physical hybrid model simulation for preoperative planning of metopic craniosynostosis surgery. Neurosurg Focus 2020 Mar 01;48(3):E19. [doi: 10.3171/2019.12.FOCUS19854] [Medline: 32114555]
61. Haerle SK, Daly MJ, Chan HH, Vescan A, Kucharczyk W, Irish JC. Virtual surgical planning in endoscopic skull base surgery. Laryngoscope 2013 Dec;123(12):2935-2939. [doi: 10.1002/lary.24004] [Medline: 24105632]
62. Morone PJ, Shah KJ, Hendricks BK, Cohen-Gadol AA. Virtual, 3-dimensional temporal bone model and its educational value for neurosurgical trainees. World Neurosurg 2019 Feb;122:e1412-e1415. [doi: 10.1016/j.wneu.2018.11.074] [Medline: 30471440]
63. Yudkowsky R, Luciano C, Banerjee P, Schwartz A, Alaraj A, Lemole Jr GM, et al. Practice on an augmented reality/haptic simulator and library of virtual brains improves residents' ability to perform a ventriculostomy. Simul Healthc 2013 Feb;8(1):25-31. [doi: 10.1097/SIH.0b013e3182662c69] [Medline: 23117306]
64. Khan R, Plahouras J, Johnston BC, Scaffidi MA, Grover SC, Walsh CM. Virtual reality simulation training for health professions trainees in gastrointestinal endoscopy. Cochrane Database Syst Rev 2018 Aug 17;8(8):CD008237 [FREE Full text] [doi: 10.1002/14651858.CD008237.pub3] [Medline: 30117156]
65. Walsh CM, Sherlock ME, Ling SC, Carnahan H. Virtual reality simulation training for health professions trainees in gastrointestinal endoscopy. Cochrane Database Syst Rev 2012 Jun 13(6):CD008237. [doi: 10.1002/14651858.CD008237.pub2] [Medline: 22696375]
66. Dharmawardana N, Ruthenbeck G, Woods C, Elmiyeh B, Diment L, Ooi EH, et al. Validation of virtual-reality-based simulations for endoscopic sinus surgery. Clin Otolaryngol 2015 Dec;40(6):569-579. [doi: 10.1111/coa.12414] [Medline: 25809675]
67. Alwani MM, Svenstrup TJ, Bandali EH, Sharma D, Higgins TS, Wu AW, et al. Validity testing of a three-dimensionally printed endoscopic sinonasal surgery simulator. Laryngoscope 2020 Dec;130(12):2748-2753. [doi: 10.1002/lary.28356] [Medline: 31714604]
68. Breimer GE, Haji FA, Bodani V, Cunningham MS, Lopez-Rios AL, Okrainec A, et al. Simulation-based education for endoscopic third ventriculostomy: a comparison between virtual and physical training models. Oper Neurosurg (Hagerstown) 2017 Feb 01;13(1):89-95. [doi: 10.1227/NEU.0000000000001317] [Medline: 28931258]
69. Aoki T, Koizumi T, Mansour DA, Fujimori A, Kusano T, Matsuda K, et al. Virtual reality with three-dimensional image guidance of individual patients' vessel anatomy in laparoscopic distal pancreatectomy. Langenbecks Arch Surg 2020 May;405(3):381-389. [doi: 10.1007/s00423-020-01871-6] [Medline: 32410077]
70. Locketz GD, Lui JT, Chan S, Salisbury K, Dort JC, Youngblood P, et al. Anatomy-specific virtual reality simulation in temporal bone dissection: perceived utility and impact on surgeon confidence. Otolaryngol Head Neck Surg 2017 Jun;156(6):1142-1149. [doi: 10.1177/0194599817691474] [Medline: 28322125]
71. Linke R, Leichtle A, Sheikh F, Schmidt C, Frenzel H, Graefe H, et al. Assessment of skills using a virtual reality temporal bone surgery simulator. Acta Otorhinolaryngol Ital 2013 Aug;33(4):273-281 [FREE Full text] [Medline: 24043916]
72. Plana NM, Diaz-Siso JR, Culnan DM, Cutting CB, Flores RL. The first year of global cleft surgery education through digital simulation: a proof of concept. Cleft Palate Craniofac J 2018 Apr;55(4):626-629. [doi: 10.1177/1055665617726544] [Medline: 29406778]
73. Kantar RS, Plana NM, Cutting CB, Diaz-Siso JR, Flores RL. Internet-based digital simulation for cleft surgery education: a 5-year assessment of demographics, usage, and global effect. J Surg Educ 2018;75(4):1120-1126. [doi: 10.1016/j.jsurg.2017.12.003] [Medline: 29396278]
74. Simpson RL. The virtual reality revolution: technology changes nursing education. Nurs Manage 2002 Sep;33(9):14-15. [doi: 10.1097/00006247-200209000-00007] [Medline: 12352788]
75. Girod S, Schvartzman SC, Gaudilliere D, Salisbury K, Silva R. Haptic feedback improves surgeons' user experience and fracture reduction in facial trauma simulation. J Rehabil Res Dev 2016;53(5):561-570 [FREE Full text] [doi: 10.1682/JRRD.2015.03.0043] [Medline: 27898160]
76. Roughley MA, Wilkinson CM. The affordances of 3D and 4D digital technologies for computerized facial depiction. Adv Exp Med Biol 2019;1138:87-101. [doi: 10.1007/978-3-030-14227-8_7] [Medline: 31313260]
77. Lin C, Gao J, Zheng H, Zhao J, Yang H, Zheng Y, et al. When to introduce three-dimensional visualization technology into surgical residency: a randomized controlled trial. J Med Syst 2019 Feb 09;43(3):71 [FREE Full text] [doi: 10.1007/s10916-019-1157-0] [Medline: 30739210]

78. Bric J, Connolly M, Kastenmeier A, Goldblatt M, Gould JC. Proficiency training on a virtual reality robotic surgical skills curriculum. Surg Endosc 2014 Dec;28(12):3343-3348. [doi: 10.1007/s00464-014-3624-5] [Medline: 24946742]
79. Davis MC, Can DD, Pindrik J, Rocque BG, Johnston JM. Virtual interactive presence in global surgical education: international collaboration through augmented reality. World Neurosurg 2016 Feb;86:103-111 [FREE Full text] [doi: 10.1016/j.wneu.2015.08.053] [Medline: 26342783]
80. Prat G, Costi D, Scarpelli F, Ruiz C, Lorea B. Anatomical variations of the coronary arteries and its correlation with digital anatomy. Ann Med 2019 May 28;51(sup1):57-58. [doi: 10.1080/07853890.2018.1561932]

Abbreviations

AR: augmented reality
MR: mixed reality
SWOT: Strength, Weakness, Opportunity, and Threat
VR: virtual reality

Edited by N Zary; submitted 03.11.21; peer-reviewed by K Reuter, A Joshi, C Moro; comments to author 15.12.21; revised version received 23.01.22; accepted 25.03.22; published 20.05.22.

Please cite as:
Wickramasinghe N, Thompson BR, Xiao J
The Opportunities and Challenges of Digital Anatomy for Medical Sciences: Narrative Review
JMIR Med Educ 2022;8(2):e34687
URL: https://mededu.jmir.org/2022/2/e34687
doi:10.2196/34687
PMID:35594064

Review

Usability Methods and Attributes Reported in Usability Studies of Mobile Apps for Health Care Education: Scoping Review

Susanne Grødem Johnson[1*], MSc; Thomas Potrebny[1*], PhD; Lillebeth Larun[2*], PhD; Donna Ciliska[3*], Prof Dr; Nina Rydland Olsen[1*], PhD

[1]Faculty of Health and Function, Western Norway University of Applied Sciences, Bergen, Norway

[2]Division of Health Services, Norwegian Institute of Public Health, Oslo, Norway

[3]Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Susanne Grødem Johnson, MSc
Faculty of Health and Function
Western Norway University of Applied Sciences
Inndalsveien 28
Bergen, 5063
Norway
Phone: 47 92213202
Email: susanne.grodem.johnson@hvl.no

Abstract

Background: Mobile devices can provide extendable learning environments in higher education and motivate students to engage in adaptive and collaborative learning. Developers must design mobile apps that are practical, effective, and easy to use, and usability testing is essential for understanding how mobile apps meet users' needs. No previous reviews have investigated the usability of mobile apps developed for health care education.

Objective: The aim of this scoping review is to identify usability methods and attributes in usability studies of mobile apps for health care education.

Methods: A comprehensive search was carried out in 10 databases, reference lists, and gray literature. Studies were included if they dealt with health care students and usability of mobile apps for learning. Frequencies and percentages were used to present the nominal data, together with tables and graphical illustrations. Examples include a figure of the study selection process, an illustration of the frequency of inquiry usability evaluation and data collection methods, and an overview of the distribution of the identified usability attributes. We followed the Arksey and O'Malley framework for scoping reviews.

Results: Our scoping review collated 88 articles involving 98 studies, mainly related to medical and nursing students. The studies were conducted from 22 countries and were published between 2008 and 2021. Field testing was the main usability experiment used, and the usability evaluation methods were either inquiry-based or based on user testing. Inquiry methods were predominantly used: 1-group design (46/98, 47%), control group design (12/98, 12%), randomized controlled trials (12/98, 12%), mixed methods (12/98, 12%), and qualitative methods (11/98, 11%). User testing methods applied were all think aloud (5/98, 5%). A total of 17 usability attributes were identified; of these, satisfaction, usefulness, ease of use, learning performance, and learnability were reported most frequently. The most frequently used data collection method was questionnaires (83/98, 85%), but only 19% (19/98) of studies used a psychometrically tested usability questionnaire. Other data collection methods included focus group interviews, knowledge and task performance testing, and user data collected from apps, interviews, written qualitative reflections, and observations. Most of the included studies used more than one data collection method.

Conclusions: Experimental designs were the most commonly used methods for evaluating usability, and most studies used field testing. Questionnaires were frequently used for data collection, although few studies used psychometrically tested questionnaires. The usability attributes identified most often were satisfaction, usefulness, and ease of use. The results indicate that combining different usability evaluation methods, incorporating both subjective and objective usability measures, and specifying which usability attributes to test seem advantageous. The results can support the planning and conduct of future usability studies for the advancement of mobile learning apps in health care education.

International Registered Report Identifier (IRRID): RR2-10.2196/19072

(JMIR Med Educ 2022;8(2):e38259) doi:10.2196/38259

KEYWORDS

user-computer interface; mobile apps; online learning; health education; students

Introduction

Background

Mobile devices can provide extendable learning environments and motivate students to engage in adaptive and collaborative learning [1,2]. Mobile devices offer various functions, enable convenient access, and support the ability to share information with other learners and teachers [3]. Most students own a mobile phone, which makes mobile learning easily accessible [4]. However, there are some challenges associated with mobile devices in learning situations, such as small screen sizes, connectivity problems, and multiple distractions in the environment [5].

Developers of mobile learning apps need to consider usability to ensure that apps are practical, effective, and easy to use [1] and to ascertain that mobile apps meet users' needs [6]. According to the International Organization for Standardization, usability is defined as "the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [7]. Better mobile learning usability will be achieved by focusing on user-centered design and attention to context, ensuring that the technology corresponds to the user's requirements and putting the user at the center of the process [8,9]. In addition, it is necessary to be conscious of the interrelatedness between usability and pedagogical design [9].

A variety of usability evaluation methods exists to test the usability of mobile apps, and Weichbroth [10] categorized them into the following 4 categories: inquiry, user testing, inspection, and analytical modeling. Inquiry methods are designed to gather data from users through questionnaires (quantitative data) and interviews and focus groups (qualitative data). User testing methods include think-aloud protocols, question-asking protocols, performance measurements, log analysis, eye tracking, and remote testing. Inspection methods, in contrast, involve experts testing apps, heuristic evaluation, cognitive walk-through, perspective-based inspections, and guideline reviews. Analytical modeling methods include cognitive task analysis and task environment analysis [10]. Across these 4 usability evaluation methods, the most commonly used data collection methods are controlled observations and surveys, whereas eye tracking, think-aloud methods, and interviews are applied less often [10].

Usability evaluations are normally performed in a laboratory or in field testing. Previous reviews have reported that usability evaluation methods are mainly conducted in a laboratory, which means in a controlled environment [1,11]. By contrast, field testing is conducted in real-life settings. There are pros and cons to the 2 different approaches. Field testing allows data collection within a dynamic environment, whereas in a laboratory data collection and conditions are easier to control [1]. A variety of data collection methods are appropriate for usability studies; for instance, in laboratories, participants performing predefined tasks, such as using questionnaires and observations, are often applied [1]. In field testing, logging mechanisms and diaries have been applied to capture user interaction with mobile apps [1].

In all, 2 systematic reviews examined various psychometrically tested usability questionnaires as a means of enhancing the usability of apps. Sousa and Lopez [12] identified 15 such questionnaires and Sure [13] identified 13. In all, 5 of the questionnaires have proven to be applicable in usability studies in general: the System Usability Scale (SUS), Questionnaire for User Interaction Satisfaction, After-Scenario Questionnaire, Post-Study System Usability Questionnaire, and Computer System Usability Questionnaire [12]. The SUS questionnaire and After-Scenario Questionnaire are most widely applied [13]. The most frequently reported usability attributes of these 5 questionnaires are learnability, efficiency, and satisfaction [12].

Usability attributes are features that measure the quality of mobile apps [1]. The most commonly reported usability attributes are effectiveness, efficiency, and satisfaction [5], which are part of the usability definition [7]. In the review by Weichbroth [10], 75 different usability attributes were identified. Given the wide selection of usability attributes, choosing appropriate attributes depends on the nature of the technology and the research question in the usability study [14]. Kumar and Mohite [1] recommended that researchers present and explain which usability attributes are being tested when mobile apps are being developed.

Previous reviews have examined the usability of mobile apps in general [5,10,11,14,15]; however, only one systematic review has specifically explored the usability of mobile learning apps [1]. However, studies from health care education were not included. Similarly, usability has not been widely explored in medical education apps [16]. Thus, there is a need to develop a better understanding of how the usability of mobile learning apps developed for health care education has been evaluated and conceptualized in previous studies.

Objectives

The aim of this scoping review has therefore been to identify usability methods and attributes in usability studies of mobile apps for health care education.

Methods

Framework

We have used the framework for scoping reviews developed by Arksey and O'Malley [17] and further developed by Levac et al [18] and Khalil et al [19]. We adopted the following five stages of this framework: (1) identifying the research question, (2) identifying relevant studies, (3) selecting studies, (4) charting the data, and (5) summarizing and reporting the results [17-19].

A detailed presentation of each step can be found in the published protocol for this scoping review [20]. We followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist for reporting scoping reviews (Multimedia Appendix 1 [21]).

Stage 1: Identifying the Research Question

The following two research questions have been formulated:

1. Which usability methods are used to evaluate the usability of mobile apps for health care education?
2. Which usability attributes are reported in the usability studies of mobile apps for health care education?

Stage 2: Identifying Relevant Studies

A total of 10 electronic databases on technology, education, and health care from January 2008 to October 2021 and February 2022 were searched. These databases were as follows: Engineering Village, Scopus, ACM Digital Library, IEEE Xplore, Education Resource Information Center, PsycINFO, CINAHL, MEDLINE, EMBASE, and Web of Science. The search string was developed by the first author and a research librarian and then peer reviewed by another research librarian. The search terms used in the Web of Science, in addition to all relevant subject headings, included: *((student* or graduate* or undergraduate* or postgraduate*) NEAR/3 nurs*)*. This search string was repeated for other types of students and combined with the Boolean operator OR. The search string for all types of health care students was then combined with various search terms for mobile apps and mobile learning using the Boolean operator AND. Similar search strategies were used and adapted for all 10 databases as shown in Multimedia Appendix 2. In addition, a citation search in Google Scholar, screening reference lists of included studies, and searching for gray literature in OpenGrey were conducted.

Stage 3: Selecting Studies

Two of the authors independently screened titles and abstracts using Rayyan web-based management software [22]. Studies deemed eligible by one of the authors were included for full-text screening and imported into the EndNote X9 (Clarivate) reference management system [23]. Eligibility for full-text screening was determined independently by two of the authors and disagreements were resolved by consensus-based discussions. Research articles with different designs were included, and there were no language restrictions. As mobile apps started appearing in 2008, this year was set as the starting point for the search. Eligibility criteria are presented in Table 1.

Table 1. Study eligibility.

	Inclusion criteria	Exclusion criteria
Population	Health care and allied health care students at the undergraduate and postgraduate levels	Health care professionals or students from education, engineering, or other nonhealth sciences
Concept	Studies of usability testing or methods of usability evaluation of mobile learning apps where the purpose relates to the development of the apps	Studies relating to learner management systems, e-learning platforms, open online courses, or distance education
Context	Typical educational setting (eg, classroom teaching, clinical placement, or simulation training), including both synchronous and asynchronous teaching	Noneducational settings not involving clinical placement or learning situations (eg, hospital or community settings)

Stage 4: Charting the Data (Data Abstraction)

The extracted data included information about the study (eg, authors, year of publication, title, and country), population (eg, number of participants), concepts (usability methods, usability attributes, and usability phase), and context (educational setting). The final data extraction sheet can be found in Multimedia Appendix 3 [24-111]. One review author extracted the data from the included studies using Microsoft Excel software [21], which was checked by another researcher.

Descriptions of usability attributes have not been standardized, making categorization challenging. Therefore, a review author used deductive analysis to interpret the usability attributes reported in the included studies. This interpretation was based on a review of usability attributes as defined in previous literature. These definitions were assessed on the basis of the results of the included studies. This analysis was reviewed and discussed by another author. Disagreements were resolved through a consensus-based discussion.

Stage 5: Summarizing and Reporting the Results

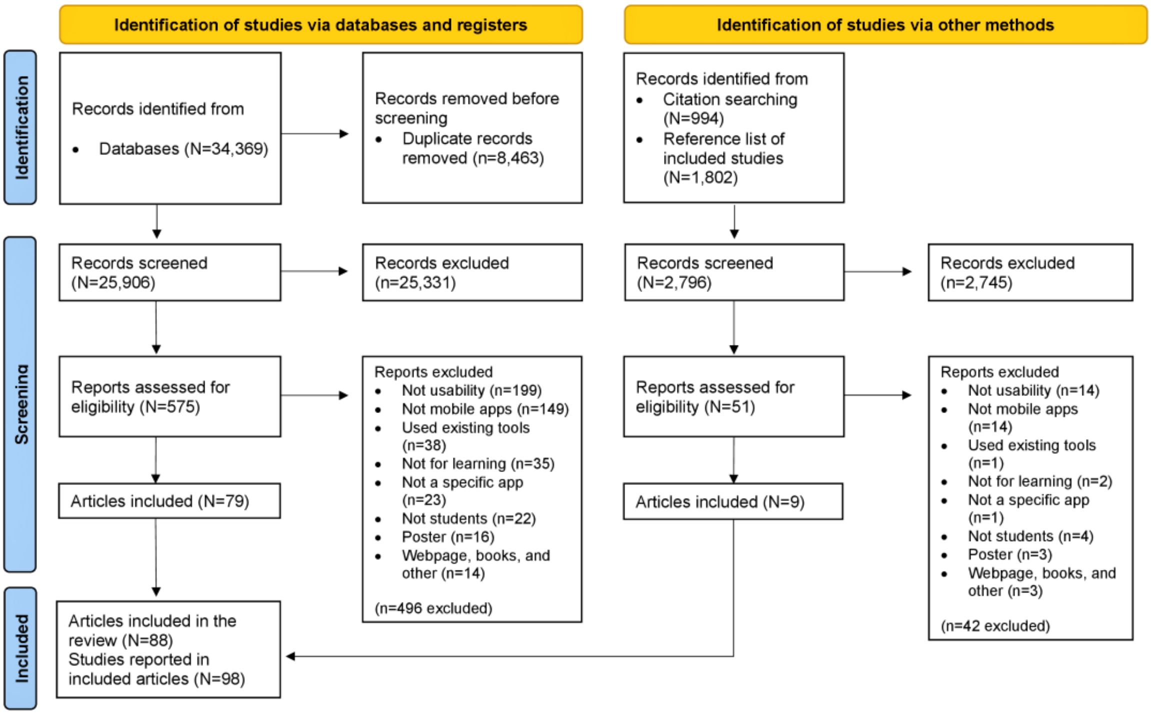
Frequencies and percentages were used to present nominal data, together with tables and graphical illustrations. For instance, a figure showing the study selection process, an illustration of the frequency of inquiry-based usability evaluation and data collection methods, and an overview of the distribution of identified usability attributes were provided.

Results

Eligible Studies

Database searches yielded 34,369 records, and 2796 records were identified using other methods. After removing duplicates, 28,702 records remained. A total of 626 reports were examined in full text. In all, 88 articles were included in the scoping review [24-111] (Figure 1). A total of 8 articles comprised results from several studies in the same article, presented as study A, study B, or study C in Multimedia Appendix 3. Therefore, a total of 98 studies were reported in the 88 articles included.

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



The included studies comprised a total sample population of 7790, with participant numbers ranging from 5 to 736 participants per study. Most of the studies included medical students (34/88, 39%) or nursing students (25/88, 28%). Other participants included students from the following disciplines: pharmacy (9/88, 10%), dentistry (5/88, 6%), physiotherapy (5/88, 6%), health sciences (3/88, 3%), and psychology (2/88, 2%). Further information is provided in Multimedia Appendix 3. There were 22 publishing countries, with most studies being from the United States (22/88, 25%), Spain (9/88, 10%), the United Kingdom (8/88, 9%), Canada (7/88, 8%), and Brazil (7/88, 8%), with an increasing number of publications from 2014. Table 2 provides an overview and characteristics of the included articles.

Table 2. Characteristics of included articles.

Study number	Study	Population (N)	Research design: data collection method	Usability attributes
1	Aebersold et al [24], 2018, United States	Nursing (N=69)	Mixed methods: questionnaire; task and knowledge performance[a]	Ease of use; learning performance; satisfaction; usefulness
2	Akl et al [25], 2008, United States	Resident (N=30)	Qualitative methods: focus groups; written qualitative reflections	Satisfaction
3	Al-Rawi et al [26], 2015, United States	Dentist (N=61)	Posttest 1-group design: questionnaire	Ease of use; frequency of use; satisfaction; usefulness
4	Albrecht et al [27], 2013, Germany	Medicine (N=6)	Posttest 1-group design: questionnaire[b]	Satisfaction
5	Alencar Neto et al [28], 2020, Brazil	Medicine (N=132)	Posttest 1-group design: questionnaire[b]	Ease of use; learnability; satisfaction; usefulness
6	Alepis and Virvou [29], 2010, Greece	Medicine (N=110)	Mixed methods: questionnaire; interviews	Ease of use; usefulness; user-friendliness
7	Ameri et al [30], 2020, Iran	Pharmacy (N=241)	Posttest 1-group design: questionnaire[b]	Context of use; efficiency; usefulness
8	Balajelini and Ghezeljeh [31], 2018, Iran	Nursing (N=41)	Posttest 1-group design: questionnaire	Ease of use; frequency of use; navigation; satisfaction; simplicity; usefulness
9	Barnes et al [32], 2015, United Kingdom	Medicine (N=42)	Randomized controlled trial: questionnaire; task and knowledge performance	Ease of use; effectiveness; learning performance; satisfaction
10	Busanello et al [33], 2015, Brazil	Dentist (N=62)	Pre-post test, nonrandomized control group design: questionnaire[b]	Learnability; learning performance; satisfaction
11	Cabero-Almenara and Roig-Vila [34], 2019, Spain	Medicine (N=50)	Pre-post test, 1-group design: questionnaire[b]	Learning performance; satisfaction
12	Choi et al [35], 2015, South Korea	Nursing (N=5)	Think-aloud methods: interviews; data from app	Context of use; ease of use; learnability; satisfaction; usefulness
13	Choi et al [36], 2018, South Korea	Nursing (N=75)	Pre-post test, nonrandomized control group design: questionnaire	Ease of use; learning performance; satisfaction; usefulness
14	Choo et al [37], 2019, Singapore	Psychology (N=8)	Mixed methods: questionnaire[b]; written qualitative reflections	Ease of use; learning performance; satisfaction; usefulness; user-friendliness
15	Chreiman et al [38], 2017, United States	Medicine (N=30)	Posttest 1-group design: questionnaire; data from app	Context of use; ease of use; frequency of use; usefulness
16	Colucci et al [39], 2015, United States	Medicine (N=115)	Posttest 1-group design: questionnaire	Effectiveness; efficiency; satisfaction; usefulness
17	Davids et al [40], 2014, South Africa	Residents (N=82)	Randomized controlled trial: questionnaire[b]; data from app	Effectiveness; efficiency; learnability; navigation; satisfaction; user-friendliness
18A	Demmans et al [41], 2018, Canada	Nursing (N=60)	Pre-post test, nonrandomized control group design: questionnaire; observations	Ease of use; effectiveness; learnability; learning performance; navigation; satisfaction
18B	Demmans et al [41], 2018, Canada	Nursing (N=85)	Pre-post test, nonrandomized control group design: questionnaire; observations	Ease of use; effectiveness; learnability; learning performance; navigation; satisfaction
19	Devraj et al [42], 2021, United States	Pharmacy (N=89)	Posttest 1-group design: questionnaire; data from app	Ease of use; errors; frequency of use; learning performance; navigation; operational usability; satisfaction; usefulness

Study number	Study	Population (N)	Research design: data collection method	Usability attributes
20	Díaz-Fernández et al [43], 2016, Spain	Physiotherapy (N=110)	Posttest 1-group design: questionnaire	Comprehensibility; ease of use; usefulness
21	Docking et al [44], 2018, United Kingdom	Paramedic (N=24)	Think-aloud methods: focus groups	Context of use; learnability; satisfaction; usefulness
22	Dodson and Baker [45], 2020, United States	Nursing (N=23)	Qualitative methods: focus groups	Ease of use; operational usability; satisfaction; usefulness; user-friendliness
23	Duarte Filho et al [46], 2014, Brazil	Medicine (N=10)	Posttest nonrandomized control group design: questionnaire	Ease of use; efficiency; satisfaction; usefulness
24	Duggan et al [47], 2020, Canada	Medicine (N=80)	Posttest 1-group design: questionnaire; data from app	Ease of use; frequency of use; satisfaction; usefulness
25	Fernandez-Lao et al [48], 2016, Spain	Physiotherapy (N=49)	Randomized controlled trial: questionnaire[b]; task and knowledge performance	Learning performance; satisfaction
26	Fralick et al [49], 2017, Canada	Medicine (N=62)	Pre-post test, nonrandomized control group design: questionnaire	Ease of use; frequency of use; learning performance; usefulness
27	Ghafari et al [50], 2020, Iran	Nursing (N=8)	Posttest 1-group design: questionnaire	Ease of use; operational usability; satisfaction; usefulness
28	Goldberg et al [51], 2014, United States	Medicine (N=18)	Posttest 1-group design: questionnaire; task and knowledge performance	Ease of use; effectiveness
29	Gutiérrez-Puertas et al [52], 2021, Spain	Nursing (N=184)	Randomized controlled trial: questionnaire; task and knowledge performance	Learning performance; satisfaction
30	Herbert et al [53], 2021, United States	Nursing (N=33)	Randomized controlled trial: questionnaire; task and knowledge performance	Ease of use; learning performance; navigation; operational usability; usefulness
31	Hsu et al [54], 2019, Taiwan	Nursing (N=16)	Qualitative methods: interviews	Context of use; operational usability; satisfaction; usefulness
32	Huang et al [55], 2010, Taiwan	Not clear (N=28)	Posttest 1-group design: questionnaire	Ease of use; satisfaction, usefulness
33	Hughes and Kearney [56], 2017, United States	Occupational therapy (N=19)	Qualitative methods: focus groups	Efficiency; satisfaction
34	Ismail et al [57], 2018, Malaysia	Health science (N=124)	Pre-post test, 1-group design: questionnaire	Ease of use; learning performance; satisfaction; user-friendliness
35	Johnson et al [58], 2021, Norway	Occupational therapy, physiotherapy, and social education (N=15)	Qualitative methods: focus groups	Context of use; ease of use; operational usability
36A	Kang Suh [59], 2018, South Korea	Nursing (N=92)	Pre-post test, nonrandomized control group design: questionnaire; data from app	Effectiveness; frequency of use; learning performance; satisfaction
36B	Kang Suh [59], 2018, South Korea	Nursing (N=49)	Qualitative methods: focus groups	Effectiveness; frequency of use; learning performance; satisfaction
37	Keegan et al [60], 2016, United States	Nursing (N=116)	Posttest nonrandomized control group design: questionnaire; task and knowledge performance	Learning performance; satisfaction; usefulness
38	Kim-Berman et al [61], 2019, United States	Dentist (N=93)	Posttest 1-group design: questionnaire; task and knowledge performance	Context of use; ease of use; effectiveness; usefulness

Study number	Study	Population (N)	Research design: data collection method	Usability attributes
39	Kojima et al [62], 2011, Japan	Physiotherapy and occupational therapy (N=41)	Pre-post test, 1-group design: questionnaire	Ease of use; learning performance; satisfaction; usefulness
40	Koulias et al [63], 2012, Australia	Medicine (N=171)	Posttest 1-group design: questionnaire	Ease of use; operational usability; satisfaction
41	Kow et al [64], 2016, Singapore	Medicine (N=221)	Pre-post test, 1-group design: questionnaire	Learning performance; satisfaction
42	Kurniawan and Witjaksono [65], 2018, Indonesia	Medicine (N=30)	Posttest 1-group design: questionnaire	Satisfaction; usefulness
43A	Lefroy et al [66], 2017, United Kingdom	Medicine (N=21)	Qualitative methods: focus groups; data from app	Context of use; frequency of use; satisfaction
43B	Lefroy et al [66], 2017, United Kingdom	Medicine (N=405)	Quantitative methods: data from app	Context of use; frequency of use; satisfaction
44	Li et al [67], 2019, Taiwan	Health care (N=70)	Pre-post test, nonrandomized control group design: questionnaire[b]	Ease of use; usefulness
45	Lin and Lin [68], 2016, Taiwan	Nursing (N=36)	Pre-post test, nonrandomized control group design: questionnaire	Cognitive load; ease of use; learnability; learning performance; usefulness
46	Lone et al [69], 2019, Ireland	Dentist (N=59)	Randomized controlled trial: questionnaire; task and knowledge performance	Ease of use; learnability; learning performance; operational usability; satisfaction
47A	Long et al [70], 2016, United States	Nursing (N=158)	Pre-post test, 1-group design: questionnaire; data from app	Ease of use; efficiency; learnability; learning performance; satisfaction
47B	Long et al [70], 2016, United States	Health science (N=159)	Randomized controlled trial: questionnaire; data from app	Ease of use; efficiency; learnability; learning performance; satisfaction
48	Longmuir [71], 2014, United States	Medicine (N=56)	Posttest 1-group design: questionnaire; data from app	Efficiency; learnability; operational usability; satisfaction
49	López et al [72], 2016, Spain	Medicine (N=67)	Posttest 1-group design: questionnaire[b]	Context of use; ease of use; errors; satisfaction; usefulness
50	Lozano-Lozano et al [73], 2020, Spain	Physiotherapy (N=110)	Randomized controlled trial: questionnaire; task and knowledge performance	Learning performance; satisfaction; usefulness
51	Lucas et al [74], 2019, Australia	Pharmacy (N=39)	Pre-post test, 1-group design: questionnaire; task and knowledge performance	Satisfaction; usefulness
52	Mathew et al [75], 2014, Canada	Medicine (N=5)	Think-aloud methods: questionnaire[b]; interviews; task and knowledge performance	Learnability; satisfaction
53	McClure [76], 2019, United States	Nursing (N=16)	Posttest 1-group design: questionnaire[b]	Learnability; satisfaction; usefulness
54	McDonald et al [77], 2018, Canada	Medicine (N=20)	Pre-post test, 1-group design: questionnaire; data from app	Effectiveness; satisfaction
55	McLean et al [78], 2014, Australia	Medicine (N=58)	Mixed methods: questionnaire; focus groups; interviews	Satisfaction
56	McMullan [79], 2018, United Kingdom	Health science (N=60)	Pre-post test, 1-group design: questionnaire	Learning performance; navigation; satisfaction; usefulness; user-friendliness
57	Mendez-Lopez et al [80], 2021, Spain	Psychology (N=67)	Pre-post test, 1-group design: questionnaire; task and knowledge performance	Cognitive load; ease of use; learning performance; satisfaction; usefulness

Study number	Study	Population (N)	Research design: data collection method	Usability attributes
58	Meruvia-Pastor et al [81], 2016, Canada	Nursing (N=10)	Pre-post test, 1-group design: questionnaire; task and knowledge performance	Ease of use; learning performance; satisfaction; usefulness
59	Mettiäinen [82], 2015, Finland	Nursing (N=121)	Mixed methods: questionnaire; focus groups	Ease of use; usefulness
60	Milner et al [83], 2020, United States	Medicine and nursing (N=66)	Posttest 1-group design: questionnaire	Satisfaction; usefulness
61	Mladenovic et al [84], 2021, Serbia	Dentist (N=56)	Posttest 1-group design: questionnaire	Context of use; ease of use; satisfaction; usefulness
62	Morris and Maynard [85], 2010, United Kingdom	Physiotherapy and nursing (N=19)	Pre-post test, 1-group design: questionnaire	Context of use; ease of use; navigation; operational usability; usefulness
63A	Nabhani et al [86], 2020, United Kingdom	Pharmacy (N=56)	Posttest 1-group design: questionnaire	Ease of use; learnability; learning performance; satisfaction; usefulness
63B	Nabhani et al [86], 2020, United Kingdom	Pharmacy (N=152)	Posttest 1-group design: questionnaire	Ease of use; learnability; learning performance; satisfaction; usefulness
63C	Nabhani et al [86], 2020, United Kingdom	Pharmacy (N=33)	Posttest 1-group design: task and knowledge performance	Ease of use; learnability; learning performance; satisfaction; usefulness
64A	Noguera et al [87], 2013, Spain	Physiotherapy (N=84)	Posttest 1-group design: questionnaire	Learning performance; satisfaction; usefulness
64B	Noguera et al [87], 2013, Spain	Physiotherapy (N=76)	Randomized controlled trial: questionnaire	Learning performance; satisfaction; usefulness
65	O'Connell et al [88], 2016, Ireland	Medicine, nursing, and pharmacy (N=89)	Randomized controlled trial: questionnaire[b]	Ease of use; learning performance; operational usability; satisfaction; simplicity
66	Oliveira et al [89], 2019, Brazil	Medicine (N=110)	Randomized controlled trial: questionnaire; task and knowledge performance	Frequency of use; learning performance; satisfaction
67	Orjuela et al [90], 2015, Colombia	Medicine (N=22)	Posttest 1-group design: questionnaire	Ease of use; satisfaction
68	Page et al [91], 2016, United States	Medicine (N=356)	Mixed methods: questionnaire; interviews	Context of use; efficiency; satisfaction
69	Paradis et al [92], 2018, Canada	Medicine and nursing (N=108)	Posttest 1-group design: questionnaire[b]	Ease of use; satisfaction; usefulness
70	Pereira et al [93], 2017, Brazil	Medicine (N=20)	Posttest 1-group design: questionnaire[b]	Ease of use; learnability; satisfaction; usefulness
71	Pereira et al [94], 2019, Brazil	Nursing (N=60)	Posttest 1-group design: questionnaire	Ease of use; operational usability; satisfaction
72A	Pinto et al [95], 2008, Brazil	Biomedical informatics (N=5)	Qualitative methods: observations; task and knowledge performance	Efficiency; errors; learnability; learning performance; operational usability; satisfaction
72B	Pinto et al [95], 2008, Brazil	Medicine (N=not clear)	Posttest nonrandomized control group design: questionnaire	Efficiency; errors; learnability; learning performance; operational usability; satisfaction
73	Quattromani et al [96], 2018, United States	Nursing (N=181)	Randomized controlled trial: questionnaire[b]	Learnability; learning performance; satisfaction; usefulness
74	Robertson and Fowler [97], 2017, United States	Medicine (N=18)	Qualitative methods: focus groups	Satisfaction

Study number	Study	Population (N)	Research design: data collection method	Usability attributes
75A	Romero et al [98], 2021, Germany	Medicine (N=22)	Think-aloud methods: questionnaire; interviews; task and knowledge performance	Effectiveness; efficiency; errors; navigation; satisfaction
75B	Romero et al [98], 2021, Germany	Medicine (N=22)	Posttest 1-group design: questionnaire[b]	Learnability; satisfaction
75C	Romero et al [98], 2021, Germany	Medicine (N=736)	Posttest 1-group design: questionnaire	Frequency of use; satisfaction
76	Salem et al [99], 2020, Australia	Pharmacy (N=33)	Posttest 1-group design: questionnaire	Operational usability; satisfaction; usefulness
77	San Martín-Rodríguezet al [100], 2020, Spain	Nursing (N=77)	Posttest 1-group design: questionnaire; task and knowledge performance	Learning performance; operational usability; satisfaction
78	Schnepp and Rogers [101], 2017, United States	Not clear (N=72)	Think-aloud methods: questionnaire[b]; interviews; task and knowledge performance	Learnability; satisfaction
79	Smith et al [102], 2016, United Kingdom	Medicine and nursing (N=74)	Mixed methods: questionnaire; focus groups	Navigation; operational usability; satisfaction; user-friendliness
80	Strandell-Laine et al [103], 2019, Finland	Nursing (N=52)	Mixed methods: questionnaire[b]; written qualitative responses	Learnability; operational usability; satisfaction
81	Strayer et al [104], 2010, United States	Medicine (N=122)	Mixed methods: questionnaire; focus groups	Context of use; learnability; learning performance; satisfaction; usefulness
82	Taylor et al [105], 2010, United Kingdom	A total of 8 different health care educations (N=79)	Qualitative methods: focus groups; written qualitative reflections	Context of use; learnability
83	Toh et al [106], 2014, Singapore	Pharmacy (N=31)	Posttest 1-group design: questionnaire	Ease of use; learnability; navigation; usefulness
84	Tsopra et al [107], 2020, France	Medicine (N=57)	Mixed methods: questionnaire; focus groups	Ease of use; operational usability; satisfaction; usefulness
85	Wu [108], 2014, Taiwan	Nursing (N=36)	Mixed methods: questionnaire; interviews	Cognitive load; effectiveness; satisfaction; usefulness
86	Wyatt et al [109], 2012, United States	Nursing (N=12)	Qualitative methods: focus groups	Ease of use; efficiency; errors; learnability; memorability; navigation; satisfaction
87	Yap [110], 2017, Singapore	Pharmacy (N=123)	Posttest 1-group design: questionnaire	Comprehensibility; learning performance; memorability; navigation; satisfaction; usefulness
88	Zhang et al [111], 2015, Singapore	Medicine (N=185)	Mixed methods: questionnaire; focus groups	Usefulness

[a]Performances measured, comparing paper and app results, quiz results, and exam results.

[b]Reported use of validated questionnaires.

Usability Evaluation Methods

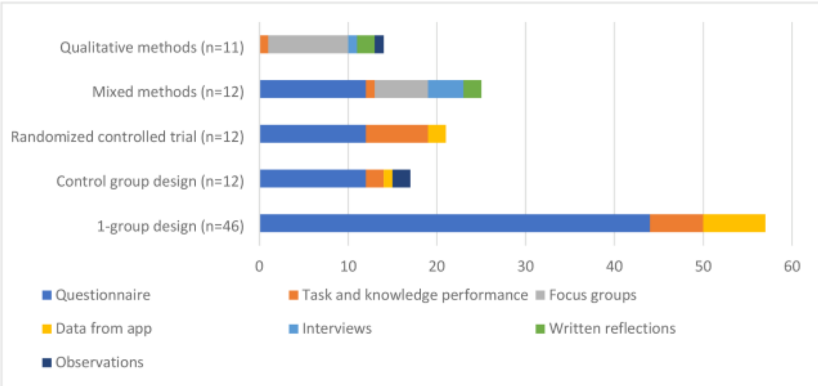
The usability evaluation methods found were either inquiry-based or based on user testing. The following inquiry methods were used: 1-group design (46/98, 47%), control group design (12/98, 12%), randomized controlled trials (12/98, 12%), mixed methods (12/98, 12%), and qualitative methods (11/98, 11%). Several studies that applied inquiry-based methods used more than one data collection method, with questionnaires being used most often (80/98, 82%), followed by task and knowledge performance testing (17/98, 17%), focus groups (15/98, 15%), collection of user data from the app (10/98, 10%), interviews (5/98, 5%), written qualitative reflections (4/98, 4%), and observations (3/98, 3%). Additional information can be found in the data extraction sheet (Multimedia Appendix 3). Figure 2 illustrates the frequency of the inquiry-based usability evaluation methods and data collection methods.

The only user testing methods found were think-aloud methods (5/98, 5%), and 4 (80%) of these studies applied more than one data collection method. The data collection methods used included interviews (4/98, 4%), questionnaires (3/98, 3%), task and knowledge performance (3/98, 3%), focus groups (1/98, 1%), and collection of user data from the app (1/98, 1%).

A total of 19 studies used a psychometrically tested usability questionnaire, including the SUS, Technology Acceptance Model, Technology Satisfaction Questionnaire, and Technology Readiness Index. SUS [112] was used in most (9/98, 9%) of the studies.

Field testing was the most frequent type of usability experiment, accounting for 72% (71/98) of usability experiments. A total of 22 (22%) studies performed laboratory testing, and 5 (5%) studies did not indicate the type of experiment performed. Multimedia Appendix 3 provides an overview of the type of experiment conducted in each study. The usability testing of the mobile apps took place in a classroom setting (41/98, 42%), in clinical placement (29/98, 30%), during simulation training (14/98, 14%), other (7/98, 7%), or the setting was not specified (5/98, 5%).

Figure 2. Inquiry usability evaluation methods and data collection methods.



Usability Attributes

A total of 17 usability attributes have been identified among the included studies. The most frequently identified attributes were satisfaction, usefulness, ease of use, learning performance, and learnability. The least frequent were errors, cognitive load, comprehensibility, memorability, and simplicity. Table 3 provides an overview of the usability attributes identified in the included studies.

Table 3. Distribution of usability attributes (n=17) and affiliated reports (N=88).

Usability attribute	Distribution, n (%)	Reports (references)
Satisfaction	74 (84)	[24-28,31-37,39-42,44-48,50,52,54-57,59,60,62-66,69-81,83,84,86-104,107-110]
Usefulness	51 (58)	[24,26,28-31,35-39,42-47,49,50,53-55,60-62,65,67,68,72-74,76,79-87,92,93,96,99,104,106-108,110,111]
Ease of use	45 (51)	[24,26,28,29,31,32,35-38,41-43,45-47,49-51,53,55,57,58,61-63,67-70,72,80-82,84-86,88,90,92-94,106,107,109]
Learning performance	33 (38)	[24,32-34,36,37,41,42,48,49,52,53,57,59,60,62,64,68-70,73,79-81,86-89,95,96,100,104,110]
Learnability	23 (26)	[28,33,35,40,41,44,68-71,75,76,86,93,95,96,98,101,103-106,109]
Operational usability	19 (22)	[42,45,50,53,54,58,63,69,71,85,88,90,94,95,99-101,103,107]
Context of use	14 (16)	[30,35,38,44,54,58,61,66,72,84,85,91,104,105]
Navigation	12 (14)	[31,40-42,53,79,85,98,102,106,109,110]
Efficiency	11 (13)	[30,39,40,46,56,70,71,91,95,98,109]
Effectiveness	10 (11)	[32,39-41,51,59,61,77,98,108]
Frequency of use	10 (11)	[26,31,38,42,47,49,59,66,89,98]
User-friendliness	7 (8)	[29,37,40,45,57,79,102]
Errors	5 (6)	[42,72,95,98,109]
Cognitive load	3 (3)	[68,80,108]
Comprehensibility	2 (2)	[43,110]
Memorability	2 (2)	[109,110]
Simplicity	2 (2)	[31,88]

Discussion

Principal Findings

This scoping review sought to identify the usability methods and attributes reported in usability studies of mobile apps for health care education. A total of 88 articles, with a total of 98 studies reported in these 88 articles, were included in this review. Our findings indicate a steady increase in publications from 2014, with studies being published in 22 different countries. Field testing was used more frequently than laboratory testing. Furthermore, the usability evaluation methods applied were either inquiry-based or based on user testing. Most of the inquiry-based methods were experiments that used questionnaires as a data collection method, and all of the studies with user testing methods applied think-aloud methods. Satisfaction, usefulness, ease of use, learning performance, and learnability were the most frequently identified usability attributes.

Comparison With Prior Work

Usability Evaluation Methods

The studies included in this scoping review mainly applied inquiry-based methods, primarily the collection of self-reported data through questionnaires. This is congruent with the results of Weichbroth [10], in which controlled observations and surveys were the most frequently applied methods. Asking users to respond to a usability questionnaire may provide relevant and valuable information. Among the 83 studies that used questionnaires in our review, only 19 (23%) used a psychometrically tested usability questionnaire; of these, the SUS questionnaire [112] was used most frequently. In line with the review on usability questionnaires [12], we recommend using a psychometrically tested usability questionnaire to support the advancement of usability science. As questionnaires address only certain usability attributes, mainly learnability, efficiency, and satisfaction [12], it would be helpful to also include additional methods, such as interviews or mixed methods, and to incorporate additional open-ended questions when using questionnaires.

Furthermore, the application of usability evaluation methods other than inquiry methods, such as user testing methods and inspection methods [10], could be beneficial and lead to more objective measures of app usability. Among other things, subjective data are collected via self-reported questionnaires, and objective data are collected based on task completion rates [40]. For example, in one of the included studies, the participants reported that the usability of the app was satisfactory by subjective measures, but the participants did not use the app [75]. Another study reported a lack of coherence between subjective and objective data; thus, these results indicate the importance of not relying solely on subjective measures of usability [40]. Therefore, it is suggested that various usability evaluation methods, including subjective and objective usability measures, are used in future usability studies.

Our review found that most of the included studies in health care education (71/98, 72%) performed field testing, whereas previous literature suggests that usability experiments in other fields are more often conducted in a laboratory [1,113]. For instance, Kumar and Mohite [1] found that 73% of the studies included in their review of mobile learning apps used laboratory testing. Mobile apps in health care education have been developed to support students' learning, on-campus and during clinical placement, in various settings and on the move. Accordingly, it is especially important to test how the apps are perceived in specific environments [5]; hence, field testing is required. However, many usability issues can be discovered in a laboratory. Particularly in the early phases of app development, testing an app with several participants in a laboratory may make it more feasible to test and improve the app [8]. Usability testing in a laboratory can provide rapid feedback on usability issues, which can then be addressed before testing the app in a real-world environment. Therefore, it may be beneficial to conduct small-scale laboratory testing before field testing.

Usability Attributes

Previous systematic reviews of mobile apps in general identified satisfaction, efficiency, and effectiveness as the most common usability attributes [5,10]. In this review, efficiency and effectiveness were explored to a limited extent, whereas satisfaction, usefulness, and ease of use were the most frequently identified usability attributes. Our results coincide with those from a previous review on the usability of mobile learning apps [1], possibly because satisfaction, usefulness, and ease of use are usability attributes of particular importance when examining mobile learning apps.

Learning performance was assessed frequently in the included studies. For ensuring that apps are valuable in a given learning context, it is relevant to test additional usability attributes such as cognitive load [9]. However, few studies included in our review examined cognitive load [68,80,108]. Mobile apps are often used in an environment with multiple distractions, which may contribute to an increased cognitive load [5], affecting the learning performance. Testing both learning performance and app users' cognitive load may improve the understanding of the app's usability.

We found that several of the included studies did not use terminology from usability literature to describe which usability attributes they were testing. For instance, studies that tested satisfaction often used words such as "likes and dislikes" and "recommend use to others" and did not specify that they tested the usability attribute satisfaction. Specifying which usability attributes are investigated will be important when performing a usability study of mobile apps, as this will influence transparency and enable comparison between different studies. In addition, evaluating a wider range of usability attributes may enable researchers to expand their perspective regarding the app's usability problems and ensure quicker improvement of the app. Defining and presenting different usability attributes in a reporting guideline can assist in deciding on and reporting relevant usability attributes. As such, a reporting guideline would be beneficial for researchers planning and conducting usability studies, a point that is also supported by the systematic review conducted by Kumar and Mohite [1].

Future Directions

Combining different usability evaluation methods that incorporate both subjective and objective usability measures can add various and important perspectives when developing apps. In future studies, it would be advantageous to use psychometrically tested usability questionnaires to support the advancement of the usability science. In addition, developers of mobile apps should determine which usability attributes are relevant before conducting usability studies (eg, by registering a protocol). Incorporating these perspectives into the development of a reporting guideline would be beneficial to future usability studies.

Strengths and Limitations

First, the search strategy was designed in collaboration with a research librarian and peer reviewed by another research librarian and included 10 databases and other sources. This broad search strategy resulted in a high number of references, which may be associated with a lower level of precision. To ensure the retrieval of all potentially pertinent articles, two of the authors independently screened titles and abstracts; studies deemed eligible by one of the authors were included for full-text screening.

Second, the full-text evaluation was challenging because the term *usability* has multiple meanings that do not always relate to usability testing. For instance, the term was used when testing students' experience of a commercially developed app but not in connection with the app's further development. In addition, many studies did not explicitly state that a mobile app was being investigated, which also created a challenge when deciding whether they satisfied the eligibility criteria. Nevertheless, reading the full-text articles independently by 2 reviewers and solving disagreements through consensus-based discussions ensured the inclusion of relevant articles.

Conclusions

This scoping review was performed to provide an overview of the usability methods used and the attributes identified in usability studies of mobile apps in health care education. Experimental designs were commonly used to evaluate usability and most studies used field testing. Questionnaires were frequently used for data collection, although few studies used psychometrically tested questionnaires. Usability attributes identified most often were satisfaction, usefulness, and ease of use. The results indicate that combining different usability evaluation methods, incorporating both subjective and objective usability measures, and specifying which usability attributes to test seem advantageous. The results can support the planning and conduct of future usability studies of the advancement of learning apps in health care education.

Acknowledgments

The research library at Western Norway University of Applied Sciences provided valuable assistance in developing and performing the search strategy for this scoping review. Gunhild Austrheim, a research librarian, provided substantial guidance in the planning and performance of the database searches. Marianne Nesbjørg Tvedt peer reviewed the search string. Malik Beglerovic also assisted with database searches. The authors would also like to thank Ane Kjellaug Brekke Gjerland for assessing the data extraction sheet.

Authors' Contributions

SGJ, LL, DC, and NRO proposed the idea for this review. SGJ, DC, and NRO contributed to the screening of titles and abstracts, and SGJ and TP decided on eligibility based on full-text examinations. SGJ extracted data from the included studies. SGJ, TP, LL, DC, and NRO contributed to the drafts of the manuscript, and all authors approved the final version for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist for reporting scoping reviews.
[DOCX File , 107 KB - mededu_v8i2e38259_app1.docx]

Multimedia Appendix 2
The search strategies for the 10 databases.
[DOCX File , 84 KB - mededu_v8i2e38259_app2.docx]

Multimedia Appendix 3
Data extraction sheet.
[XLSX File (Microsoft Excel File), 156 KB - mededu_v8i2e38259_app3.xlsx]

References

1. Kumar BA, Mohite P. Usability of mobile learning applications: a systematic literature review. J Comput Educ 2018;5(1):1-17. [doi: 10.1007/s40692-017-0093-6]
2. Asarbakhsh M, Sandars J. E-learning: the essential usability perspective. Clin Teach 2013 Feb;10(1):47-50. [doi: 10.1111/j.1743-498X.2012.00627.x] [Medline: 23294744]
3. Lall P, Rees R, Law GC, Dunleavy G, Cotič Ž, Car J. Influences on the implementation of mobile learning for medical and nursing education: qualitative systematic review by the digital health education collaboration. J Med Internet Res 2019 Feb 28;21(2):e12895 [FREE Full text] [doi: 10.2196/12895] [Medline: 30816847]
4. Sophonhiranrak S, Promsaka Na Sakonnak N. Limitations of mobile learning: a systematic review. In: E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education. 2017 Presented at: E-Learn '17; October 17, 2017; Vancouver, British Columbia, Canada p. 965-971.
5. Harrison R, Flood D, Duce D. Usability of mobile applications: literature review and rationale for a new usability model. J Interact Sci 2013 May 7;1(1):1-16. [doi: 10.1186/2194-0827-1-1]
6. Paz F, Pow-Sang JA. A systematic mapping review of usability evaluation methods for software development process. Int J Softw Eng Appl 2016 Jan 31;10(1):165-178. [doi: 10.14257/ijseia.2016.10.1.16]
7. ISO 9241-11:2018. Ergonomics of human-system interaction — part 11: usability: definitions and concepts. Geneva, Switzerland: International Organization for Standardization; 2018.
8. Rubin J, Chisnell D, Spool J. Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. 2nd edition. Hoboken, NJ, USA: Wiley; May 2008.
9. Kukulska-Hulme A. Mobile usability in educational contexts: what have we learnt? Int Rev Res Open Dis Learn 2007 Jun 15;8(2):1-16. [doi: 10.19173/irrodl.v8i2.356]
10. Weichbroth P. Usability of mobile applications: a systematic literature study. IEEE Access 2020 Mar 19;8:55563-55577. [doi: 10.1109/access.2020.2981892]
11. Nayebi F, Desharnais JM, Abran A. The state of the art of mobile application usability evaluation. In: Proceedings of the 25th IEEE Canadian Conference on Electrical and Computer Engineering. 2012 Presented at: CCECE '12; April 29-May 02, 2012; Montreal, Canada p. 1-4. [doi: 10.1109/ccece.2012.6334930]
12. Sousa VE, Dunn Lopez K. Towards usable e-health. A systematic review of usability questionnaires. Appl Clin Inform 2017 May 10;8(2):470-490 [FREE Full text] [doi: 10.4338/ACI-2016-10-R-0170] [Medline: 28487932]
13. Sure M. Questionnaires for Usability: A Systematic Literature Review. Linköping, Sweden: Linköping University; 2014.
14. Zhang D, Adipat B. Challenges, methodologies, and issues in the usability testing of mobile applications. Int J Hum Comput Interact 2005 Jul;18(3):293-308. [doi: 10.1207/s15327590ijhc1803_3]
15. Ismail NA, Ahmad F, Kamaruddin NA, Ibrahim R. A review on usability issues in mobile applications. J Mob Comput Appl 2016;3(3):47-52. [doi: 10.9790/0050-03034752]
16. Sandars J. The importance of usability testing to allow e-learning to reach its potential for medical education. Educ Prim Care 2010 Jan;21(1):6-8. [doi: 10.1080/14739879.2010.11493869] [Medline: 20202315]
17. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
18. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
19. Khalil H, Peters M, Godfrey CM, McInerney P, Soares CB, Parker D. An evidence-based approach to scoping reviews. Worldviews Evid Based Nurs 2016 Apr;13(2):118-123. [doi: 10.1111/wvn.12144] [Medline: 26821833]
20. Johnson SG, Potrebny T, Larun L, Ciliska D, Olsen NR. Usability methods and attributes reported in usability studies of mobile apps for health care education: protocol for a scoping review. JMIR Res Protoc 2020 Aug 04;9(8):e19072 [FREE Full text] [doi: 10.2196/19072] [Medline: 32750011]
21. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
22. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
23. The EndNote Team. EndNote. X9.2. Clarivate. Philadelphia, PA; 2013. URL: https://endnote.com/ [accessed 2022-02-15]
24. Aebersold M, Voepel-Lewis T, Cherara L, Weber M, Khouri C, Levine R, et al. Interactive anatomy-augmented virtual simulation training. Clin Simul Nurs 2018 Feb;15:34-41 [FREE Full text] [doi: 10.1016/j.ecns.2017.09.008] [Medline: 29861797]
25. Akl EA, Mustafa R, Slomka T, Alawneh A, Vedavalli A, Schünemann HJ. An educational game for teaching clinical practice guidelines to Internal Medicine residents: development, feasibility and acceptability. BMC Med Educ 2008 Nov 18;8:50 [FREE Full text] [doi: 10.1186/1472-6920-8-50] [Medline: 19017400]
26. Al-Rawi W, Easterling L, Edwards PC. Development of a mobile device optimized cross platform-compatible oral pathology and radiology spaced repetition system for dental education. J Dent Educ 2015 Apr;79(4):439-447. [Medline: 25838016]
27. Albrecht UV, Noll C, von Jan U. Explore and experience: mobile augmented reality for medical training. Stud Health Technol Inform 2013;192:382-386. [Medline: 23920581]

28. Alencar Neto JB, Araújo RL, Barroso Filho EM, Silva PG, Garrido RJ, Rocha PH, et al. Development and validation of a smartphone application for orthopedic residency education. Rev Bras Educ Med 2020 May;44(4):561-567. [doi: 10.1590/1981-5271v44.4-20200212.ING]
29. Alepis E, Virvou M. Evaluation of mobile authoring and tutoring in medical issues. US China Educ Rev 2010 Jul;7(7):84-92 [FREE Full text]
30. Ameri A, Khajouei R, Ameri A, Jahani Y. Acceptance of a mobile-based educational application (LabSafety) by pharmacy students: an application of the UTAUT2 model. Educ Inf Technol 2019 Aug 1;25(1):419-435. [doi: 10.1007/s10639-019-09965-5]
31. Taziki Balajelini F, Najafi Ghezeljeh T. Prehospital trauma management: evaluation of a new designed smartphone application. J Client Centered Nurs Care 2018 Nov 30;4(4):193-202. [doi: 10.32598/jccnc.4.4.193]
32. Barnes J, Duffy A, Hamnett N, McPhail J, Seaton C, Shokrollahi K, et al. The Mersey Burns App: evolving a model of validation. Emerg Med J 2015 Aug;32(8):637-641. [doi: 10.1136/emermed-2013-203416] [Medline: 25371408]
33. Busanello FH, da Silveira PF, Liedke GS, Arús NA, Vizzotto MB, Silveira HE, et al. Evaluation of a digital learning object (DLO) to support the learning process in radiographic dental diagnosis. Eur J Dent Educ 2015 Nov;19(4):222-228. [doi: 10.1111/eje.12125] [Medline: 25393811]
34. Cabero-Almenara J, Roig-Vila R. The motivation of technological scenarios in augmented reality (AR): results of different experiments. Appl Sci 2019 Jul 19;9(14):2907-2916. [doi: 10.3390/app9142907]
35. Choi M, Lee HS, Park JH. Usability of academic electronic medical record application for nursing students' clinical practicum. Healthc Inform Res 2015 Jul;21(3):191-195 [FREE Full text] [doi: 10.4258/hir.2015.21.3.191] [Medline: 26279956]
36. Choi M, Lee H, Park JH. Effects of using mobile device-based academic electronic medical records for clinical practicum by undergraduate nursing students: a quasi-experimental study. Nurse Educ Today 2018 Feb;61:112-119. [doi: 10.1016/j.nedt.2017.11.018] [Medline: 29197263]
37. Choo CC, Devakaran B, Chew PK, Zhang MW. Smartphone application in postgraduate clinical psychology training: trainees' perspectives. Int J Environ Res Public Health 2019 Oct 30;16(21):4206 [FREE Full text] [doi: 10.3390/ijerph16214206] [Medline: 31671592]
38. Chreiman KM, Prakash PS, Martin ND, Kim PK, Mehta S, McGinnis K, et al. Staying connected: service-specific orientation can be successfully achieved using a mobile application for onboarding care providers. Trauma Surg Acute Care Open 2017;2(1):e000085 [FREE Full text] [doi: 10.1136/tsaco-2017-000085] [Medline: 29766089]
39. Colucci PG, Kostandy P, Shrauner WR, Arleo E, Fuortes M, Griffin AS, et al. Development and utilization of a Web-based application as a robust radiology teaching tool (radstax) for medical student anatomy teaching. Acad Radiol 2015 Feb;22(2):247-255 [FREE Full text] [doi: 10.1016/j.acra.2014.09.014] [Medline: 25964956]
40. Davids MR, Chikte UM, Halperin ML. Effect of improving the usability of an e-learning resource: a randomized trial. Adv Physiol Educ 2014 Jun;38(2):155-160 [FREE Full text] [doi: 10.1152/advan.00119.2013] [Medline: 24913451]
41. Demmans Epp C, Horne J, Scolieri BB, Kane I, Bowser AS. PsychOut! a mobile app to support mental status assessment training. In: Proceedings of the 13th European Conference on Technology Enhanced Learning. 2018 Presented at: EC-TEL '18; September 3-5, 2018; Leeds, UK p. 216-230. [doi: 10.1007/978-3-319-98572-5_17]
42. Devraj R, Colyott L, Cain J. Design and evaluation of a mobile serious game application to supplement instruction. Curr Pharm Teach Learn 2021 Sep;13(9):1228-1235. [doi: 10.1016/j.cptl.2021.06.032] [Medline: 34330403]
43. Díaz-Fernández Á, Jiménez-Delgado JJ, Osuna-Pérez MC, Rueda-Ruiz A, Paulano-Godino F. Development and implementation of a mobile application to improve university teaching of electrotherapy. In: Proceedings of the 2016 International Conference on Interactive Mobile Communication, Technologies and Learning. 2016 Presented at: IMCL '16; October 17-19, 2016; San Diego, CA, USA p. 33-37. [doi: 10.1109/imctl.2016.7753766]
44. Docking RE, Lane M, Schofield PA. Usability testing of the iPhone app to improve pain assessment for older adults with cognitive impairment (prehospital setting): a qualitative study. Pain Med 2018 Jun 01;19(6):1121-1131. [doi: 10.1093/pm/pnx028] [Medline: 28340247]
45. Dodson CH, Baker E. Focus group testing of a mobile app for pharmacogenetic-guided dosing. J Am Assoc Nurse Pract 2020 Feb 04;33(3):205-210. [doi: 10.1097/JXX.0000000000000392] [Medline: 32039960]
46. Duarte Filho NF, Gonçalves CF, Pizetta DC. Experimental analysis of the efficiency of application E-Mages in medical imaging visualization. In: Proceedings of the 9th Iberian Conference on Information Systems and Technologies. 2014 Presented at: CISTI '14; June 18-21, 2014; Barcelona, Spain p. 1-6. [doi: 10.1109/cisti.2014.6876925]
47. Duggan N, Curran VR, Fairbridge NA, Deacon D, Coombs H, Stringer K, et al. Using mobile technology in assessment of entrustable professional activities in undergraduate medical education. Perspect Med Educ 2021 Dec;10(6):373-377 [FREE Full text] [doi: 10.1007/s40037-020-00618-9] [Medline: 33095399]
48. Fernández-Lao C, Cantarero-Villanueva I, Galiano-Castillo N, Caro-Morán E, Díaz-Rodríguez L, Arroyo-Morales M. The effectiveness of a mobile application for the development of palpation and ultrasound imaging skills to supplement the traditional learning of physiotherapy students. BMC Med Educ 2016 Oct 19;16(1):274 [FREE Full text] [doi: 10.1186/s12909-016-0775-1] [Medline: 27756288]

49. Fralick M, Haj R, Hirpara D, Wong K, Muller M, Matukas L, et al. Can a smartphone app improve medical trainees' knowledge of antibiotics? Int J Med Educ 2017 Nov 30;8:416-420 [FREE Full text] [doi: 10.5116/ijme.5a11.8422] [Medline: 29200402]
50. Ghafari S, Yazdannik A, Mohamadirizi S. Education promotion based on "mobile technology" in the Critical Care Nursing Department: four-phase intervention. J Educ Health Promot 2020;9:325 [FREE Full text] [doi: 10.4103/jehp.jehp_34_20] [Medline: 33426129]
51. Goldberg H, Klaff J, Spjut A, Milner S. A mobile app for measuring the surface area of a burn in three dimensions: comparison to the Lund and Browder assessment. J Burn Care Res 2014;35(6):480-483. [doi: 10.1097/BCR.0000000000000037] [Medline: 24918947]
52. Gutiérrez-Puertas L, García-Viola A, Márquez-Hernández VV, Garrido-Molina JM, Granados-Gámez G, Aguilera-Manrique G. Guess it (SVUAL): an app designed to help nursing students acquire and retain knowledge about basic and advanced life support techniques. Nurse Educ Pract 2021 Jan;50:102961. [doi: 10.1016/j.nepr.2020.102961] [Medline: 33421681]
53. Herbert VM, Perry RJ, LeBlanc CA, Haase KN, Corey RR, Giudice NA, et al. Developing a smartphone app with augmented reality to support virtual learning of nursing students on heart failure. Clin Simul Nurs 2021 May;54:77-85. [doi: 10.1016/j.ecns.2021.02.003]
54. Hsu LL, Hsiang HC, Tseng YH, Huang SY, Hsieh SI. Nursing students' experiences of using a smart phone application for a physical assessment course: a qualitative study. Jpn J Nurs Sci 2019 Apr;16(2):115-124. [doi: 10.1111/jjns.12215] [Medline: 29947101]
55. Huang HM, Chen YL, Chen KY. Investigation of three-dimensional human anatomy applied in mobile learning. In: Proceedings of the 2010 International Computer Symposium. 2010 Presented at: ICS '10; December 16-18, 2010; Tainan, Taiwan p. 358-363. [doi: 10.1109/compsym.2010.5685486]
56. Hughes JK, Kearney P. Impact of an iDevice application on student learning in an occupational therapy kinesiology course. Mhealth 2017;3:43 [FREE Full text] [doi: 10.21037/mhealth.2017.08.02] [Medline: 29184895]
57. Ismail SN, Rangga JU, Rasdi I, Rahman UR, Samah MA. Mobile apps application to improve safety and health knowledge, attitude and practice among university students. Malays J Med Health Sci 2018;14:47-55.
58. Johnson SG, Titlestad KB, Larun L, Ciliska D, Olsen NR. Experiences with using a mobile application for learning evidence-based practice in health and social care education: an interpretive descriptive study. PLoS One 2021;16(7):e0254272 [FREE Full text] [doi: 10.1371/journal.pone.0254272] [Medline: 34252136]
59. Kang J, Suh EE. Development and evaluation of "chronic illness care smartphone apps" on nursing students' knowledge, self-efficacy, and learning experience. Comput Inform Nurs 2018 Nov;36(11):550-559. [doi: 10.1097/CIN.0000000000000447] [Medline: 29901475]
60. Keegan RD, Oliver MC, Stanfill TJ, Stevens KV, Brown GR, Ebinger M, et al. Use of a mobile device simulation as a preclass active learning exercise. J Nurs Educ 2016 Jan;55(1):56-59. [doi: 10.3928/01484834-20151214-14] [Medline: 26812386]
61. Kim-Berman H, Karl E, Sherbel J, Sytek L, Ramaswamy V. Validity and user experience in an augmented reality virtual tooth identification test. J Dent Educ 2019 Nov;83(11):1345-1352. [doi: 10.21815/JDE.019.139] [Medline: 31406007]
62. Kojima S, Mitani M, Ishikawa A. Development of an E-learning resource on mobile devices for kinesiology: a pilot study. J Phys Ther Sci 2011;23(4):667-672. [doi: 10.1589/jpts.23.667]
63. Scott KM, Kitching S, Burn D, Koulias M, Campbell D, Phelps M. "Wherever, whenever" learning in medicine: interactive mobile case-based project. In: Proceedings of the 27th Annual conference of the Australian Society for Computers in Tertiary Education. 2010 Presented at: Ascilite '10; December 5-8, 2010; New South Wales, Australia p. 888-890.
64. Kow AW, Ang BL, Chong CS, Tan WB, Menon KR. Innovative patient safety curriculum using iPAD game (PASSED) improved patient safety concepts in undergraduate medical students. World J Surg 2016 Nov;40(11):2571-2580. [doi: 10.1007/s00268-016-3623-x] [Medline: 27417109]
65. Kurniawan MH, Suharjito, Diana, Witjaksono G. Human anatomy learning systems using augmented reality on mobile application. Procedia Comput Sci 2018;135:80-88. [doi: 10.1016/j.procs.2018.08.152]
66. Lefroy J, Roberts N, Molyneux A, Bartlett M, Gay S, McKinley R. Utility of an app-based system to improve feedback following workplace-based assessment. Int J Med Educ 2017 May 31;8:207-216 [FREE Full text] [doi: 10.5116/ijme.5910.dc69] [Medline: 28578320]
67. Li YJ, Lee LH, Cheng YT, Ou YY. Design and evaluation of a healthcare management terminology mobile learning application. In: Proceedings of the 2019 IEEE International Conference on Healthcare Informatics. 2019 Presented at: ICHI '19; June 10-13, 2019; Xi'an, China p. 1-9. [doi: 10.1109/ichi.2019.8904572]
68. Lin YT, Lin YC. Effects of mental process integrated nursing training using mobile device on students' cognitive load, learning attitudes, acceptance, and achievements. Comput Human Behav 2016 Feb;55(B):1213-1221 [FREE Full text] [doi: 10.1016/j.chb.2015.03.076]
69. Lone M, Vagg T, Theocharopoulos A, Cryan JF, Mckenna JP, Downer EJ, et al. Development and assessment of a three-dimensional tooth morphology quiz for dental students. Anat Sci Educ 2019 May;12(3):284-299. [doi: 10.1002/ase.1815] [Medline: 30378278]

70. Long JD, Gannaway P, Ford C, Doumit R, Zeeni N, Sukkarieh-Haraty O, et al. Effectiveness of a technology-based intervention to teach evidence-based practice: the EBR tool. Worldviews Evid Based Nurs 2016 Feb;13(1):59-65. [doi: 10.1111/wvn.12132] [Medline: 26773417]
71. Longmuir KJ. Interactive computer-assisted instruction in acid-base physiology for mobile computer platforms. Adv Physiol Educ 2014 Mar;38(1):34-41 [FREE Full text] [doi: 10.1152/advan.00083.2013] [Medline: 24585467]
72. López MM, López MM, de la Torre Díez I, Jimeno JC, López-Coronado M. A mobile decision support system for red eye diseases diagnosis: experience with medical students. J Med Syst 2016 Jun;40(6):151. [doi: 10.1007/s10916-016-0508-3] [Medline: 27142275]
73. Lozano-Lozano M, Galiano-Castillo N, Fernández-Lao C, Postigo-Martin P, Álvarez-Salvago F, Arroyo-Morales M, et al. The Ecofisio mobile app for assessment and diagnosis using ultrasound imaging for undergraduate health science students: multicenter randomized controlled trial. J Med Internet Res 2020 Mar 10;22(3):e16258 [FREE Full text] [doi: 10.2196/16258] [Medline: 32154784]
74. Lucas C, Gibson A, Shum SB. Pharmacy students' utilization of an online tool for immediate formative feedback on reflective writing tasks. Am J Pharm Educ 2019 Aug;83(6):6800 [FREE Full text] [doi: 10.5688/ajpe6800] [Medline: 31507274]
75. Mathew D, Archer N, McKibbon KA, Dillenburg R. Integrating clinical practice guidelines within a mobile tablet app. In: Proceedings of the 2014 International Conference on E-Commerce. 2014 Presented at: EC '14; July 17-19, 2014; Lisbon, Portugal p. 357-361.
76. McClure KC. Usability of a mobile website focused on preoperative and intraoperative anesthetic considerations for the cardiac patient with valvular dysfunction. Franciscan Missionaries of Our Lady University, Baton Rouge, LA, USA: ProQuest Dissertations Publishing; Dec 2019.
77. McDonald H, Gawad N, Raîche I, Rubens FD. Transition to residency: the successful development and implementation of a nonclinical elective in perioperative management. J Surg Educ 2018;75(3):628-638. [doi: 10.1016/j.jsurg.2017.08.022] [Medline: 28888420]
78. McLean M, Brazil V, Johnson P. How we "breathed life" into problem-based learning cases using a mobile application. Med Teach 2014 Oct;36(10):849-852. [doi: 10.3109/0142159X.2014.886771] [Medline: 24571614]
79. McMullan M. Evaluation of a medication calculation mobile app using a cognitive load instructional design. Int J Med Inform 2018 Oct;118:72-77. [doi: 10.1016/j.ijmedinf.2018.07.005] [Medline: 30153925]
80. Mendez-Lopez M, Juan MC, Molla R, Fidalgo C. Evaluation of an augmented reality application for learning neuroanatomy in psychology. Anat Sci Educ 2022 May;15(3):535-551. [doi: 10.1002/ase.2089] [Medline: 33866682]
81. Meruvia-Pastor O, Patra P, Andres K, Twomey C, Peña-Castillo L. OMARC: an online multimedia application for training health care providers in the assessment of respiratory conditions. Int J Med Inform 2016 May;89:15-24. [doi: 10.1016/j.ijmedinf.2016.02.007] [Medline: 26980355]
82. Mettiäinen S. Electronic assessment and feedback tool in supervision of nursing students during clinical training. Electron J E Learn 2015;13(1):42-55.
83. Milner KA, McCloud R, Cullen J. The evidence appraisal game: an innovative strategy for teaching step 3 in evidence-based practice. Worldviews Evid Based Nurs 2020 Apr;17(2):173-175. [doi: 10.1111/wvn.12433] [Medline: 32233017]
84. Mladenovic R, Davidovic B, Tusek I, Trickovic-Janjic O, Mladenovic K. The effect of a mobile application for learning about traumatic dental injuries during the COVID-19 pandemic. Srp Arh Celok Lek 2021;149(3-4):202-207. [doi: 10.2298/sarh201110007m]
85. Morris J, Maynard V. Pilot study to test the use of a mobile device in the clinical setting to access evidence-based practice resources. Worldviews Evid Based Nurs 2010 Dec;7(4):205-213. [doi: 10.1111/j.1741-6787.2009.00171.x] [Medline: 19804588]
86. Nabhani S, Harrap N, Ishtiaq S, Ling V, Dudzinski M, Greenhill D, et al. Development and evaluation of an educational game to support pharmacy students. Curr Pharm Teach Learn 2020 Jul;12(7):786-803. [doi: 10.1016/j.cptl.2020.02.006] [Medline: 32540040]
87. Noguera JM, Jiménez JJ, Osuna-Pérez MC. Development and evaluation of a 3D mobile application for learning manual therapy in the physiotherapy laboratory. Comput Educ 2013 Nov;69(1):96-108. [doi: 10.1016/j.compedu.2013.07.007]
88. O'Connell E, Pegler J, Lehane E, Livingstone V, McCarthy N, Sahm LJ, et al. Near field communications technology and the potential to reduce medication errors through multidisciplinary application. Mhealth 2016;2:29 [FREE Full text] [doi: 10.21037/mhealth.2016.07.01] [Medline: 28293602]
89. Oliveira EY, Crosewski NI, Silva AL, Ribeiro CT, de Oliveira CM, Fogaça RT, et al. Profile of educational technology use by medical students and evaluation of a new mobile application designed for the study of human physiology. J Med Syst 2019 Aug 27;43(10):313. [doi: 10.1007/s10916-019-1438-7] [Medline: 31451942]
90. Orjuela MA, Uribe-Quevedo A, Jaimes N, Perez-Gutierrez B. External automatic defibrillator game-based learning app. In: Proceedings of the 2015 IEEE Games Entertainment Media Conference. 2015 Presented at: GEM '15; October 14-16, 2015; Toronto, Canada p. 1-4. [doi: 10.1109/gem.2015.7377206]
91. Page CP, Reid A, Coe CL, Carlough M, Rosenbaum D, Beste J, et al. Learnings from the pilot implementation of mobile medical milestones application. J Grad Med Educ 2016 Oct;8(4):569-575 [FREE Full text] [doi: 10.4300/JGME-D-15-00550.1] [Medline: 27777669]

92. Paradis M, Stiell I, Atkinson KM, Guerinet J, Sequeira Y, Salter L, et al. Acceptability of a mobile clinical decision tool among emergency department clinicians: development and evaluation of the Ottawa rules app. JMIR Mhealth Uhealth 2018 Jun 11;6(6):e10263 [FREE Full text] [doi: 10.2196/10263] [Medline: 29891469]
93. Victor Soares Pereira R, Kubrusly M, Marçal E. Desenvolvimento, utilização e avaliação de uma aplicação móvel para educação médica: um estudo de caso em Anestesiologia. RENOTE 2017 Jul 28;15(1):75104. [doi: 10.22456/1679-1916.75104]
94. Pereira FG, Rocha D, Melo GA, Jaques RM, Formiga LM. Building and validating a digital application for the teaching of surgical instrumentation. Cogitare Enferm 2019 Mar 11;24:e58334. [doi: 10.5380/ce.v24i0.58334]
95. Pinto VC, da Costa TM, Naveira MC, Sigulem D, Schor P, Pisa IT. MDFluxo: ophtalmology education with a PDA efficacy and usability evaluation. In: Proceedings of the 1st International Conference on Health Informatics. 2008 Presented at: HEALTHINF '08; January 28-31, 2008; Madeira, Portugal p. 227-230. [doi: 10.5220/0001042002270230]
96. Quattromani E, Hassler M, Rogers N, Fitzgerald J, Buchanan P. Smart pump app for infusion pump training. Clin Simul Nurs 2018 Apr 1;17:28-37. [doi: 10.1016/j.ecns.2017.11.004]
97. Robertson AC, Fowler LC. Medical student perceptions of learner-initiated feedback using a mobile Web application. J Med Educ Curric Dev 2017;4:2382120517746384 [FREE Full text] [doi: 10.1177/2382120517746384] [Medline: 29349345]
98. Roa Romero Y, Tame H, Holzhausen Y, Petzold M, Wyszynski JV, Peters H, et al. Design and usability testing of an in-house developed performance feedback tool for medical students. BMC Med Educ 2021 Jun 23;21(1):354 [FREE Full text] [doi: 10.1186/s12909-021-02788-4] [Medline: 34162382]
99. Salem S, Cooper J, Schneider J, Croft H, Munro I. Student acceptance of using augmented reality applications for learning in pharmacy: a pilot study. Pharmacy (Basel) 2020 Jul 21;8(3):122 [FREE Full text] [doi: 10.3390/pharmacy8030122] [Medline: 32708150]
100. San Martín-Rodríguez L, Escalada-Hernández P, Soto-Ruiz N. A themed game to learn about nursing theories and models: a descriptive study. Nurse Educ Pract 2020 Nov;49:102905. [doi: 10.1016/j.nepr.2020.102905] [Medline: 33137561]
101. Schnepp JC, Rogers CB. Evaluating the acceptability and usability of EASEL: a mobile application that supports guided reflection for experiential learning activities. J Inf Technol Educ Innov Pract 2017;16:195-214. [doi: 10.28945/3791]
102. Smith N, Rapley T, Jandial S, English C, Davies B, Wyllie R, et al. Paediatric musculoskeletal matters (pmm)--collaborative development of an online evidence based interactive learning tool and information resource for education in paediatric musculoskeletal medicine. Pediatr Rheumatol Online J 2016 Jan 05;14(1):1 [FREE Full text] [doi: 10.1186/s12969-015-0062-4] [Medline: 26728031]
103. Strandell-Laine C, Leino-Kilpi H, Löyttyniemi E, Salminen L, Stolt M, Suomi R, et al. A process evaluation of a mobile cooperation intervention: a mixed methods study. Nurse Educ Today 2019 Sep;80:1-8. [doi: 10.1016/j.nedt.2019.05.037] [Medline: 31175963]
104. Strayer SM, Pelletier SL, Martindale JR, Rais S, Powell J, Schorling JB. A PDA-based counseling tool for improving medical student smoking cessation counseling. Fam Med 2010 May;42(5):350-357 [FREE Full text] [Medline: 20461567]
105. Taylor JD, Dearnley CA, Laxton JC, Coates CA, Treasure‐Jones T, Campbell R, et al. Developing a mobile learning solution for health and social care practice. Dist Educ 2010 Jul 30;31(2):175-192. [doi: 10.1080/01587919.2010.503343]
106. Yap KY, Toh TW, Chui WK. Development of a virtual patient record mobile app for pharmacy practice education. Arch Pharma Pract 2014;5(2):66-71. [doi: 10.4103/2045-080X.132650]
107. Tsopra R, Courtine M, Sedki K, Eap D, Cabal M, Cohen S, et al. AntibioGame®: a serious game for teaching medical students about antibiotic use. Int J Med Inform 2020 Apr;136:104074 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104074] [Medline: 31926355]
108. Wu TT. The use of a mobile assistant learning system for health education based on project-based learning. Comput Inform Nurs 2014 Oct;32(10):497-503. [doi: 10.1097/CIN.0000000000000089] [Medline: 25140444]
109. Wyatt TH, Li X, Indranoi C, Bell M. Developing iCare v.1.0: an academic electronic health record. Comput Inform Nurs 2012 Jun;30(6):321-329. [doi: 10.1097/NXN.0b013e31824af81f] [Medline: 22411413]
110. Yap KY. Usefulness of the mobile interactive pharmacy education enhancement resource (miPEER) mobile Web-app as a learning tool for electronic health records. Int J Clin Skills 2017;11(6):1-7. [doi: 10.4172/clinical-skills.1000131]
111. Zhang MW, Cheok CC, Ho RC. Global outreach of a locally-developed mobile phone app for undergraduate psychiatry education. JMIR Med Educ 2015 Jun 08;1(1):e3 [FREE Full text] [doi: 10.2196/mededu.4179] [Medline: 27731838]
112. Brooke J. SUS: a retrospective. J Usabil Stud Arch 2013;8(2):29-40 [FREE Full text] [doi: 10.5555/2817912.2817913]
113. Bastien JM. Usability testing: a review of some methodological and technical aspects of the method. Int J Med Inform 2010 Apr;79(4):e18-e23. [doi: 10.1016/j.ijmedinf.2008.12.004] [Medline: 19345139]

Abbreviations

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
SUS: System Usability Scale

Edited by T Leung; submitted 25.03.22; peer-reviewed by L Rutter, H Mehdizadeh, L Gutierrez-Puertas; comments to author 12.05.22; revised version received 02.06.22; accepted 05.06.22; published 29.06.22.

Please cite as:
Johnson SG, Potrebny T, Larun L, Ciliska D, Olsen NR
Usability Methods and Attributes Reported in Usability Studies of Mobile Apps for Health Care Education: Scoping Review
JMIR Med Educ 2022;8(2):e38259
URL: https://mededu.jmir.org/2022/2/e38259
doi:10.2196/38259
PMID:35767323

Viewpoint

Fundraising in Education: Road Map to Involving Medical Educators in Fundraising

Alireza Jalali[1*], MD; Jacline A Nyman[1*], PhD; Elaine Hamelin-Mitchell[1*], MSc

Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Alireza Jalali, MD
Faculty of Medicine
University of Ottawa
451 Smyth Road
Ottawa, ON, K1H 8M5
Canada
Phone: 1 6135625800
Email: ajalali@uottawa.ca

Abstract

Traditional funding models must change as governments decrease funding and often freeze tuition at a domestic level. As a result, universities face an increasing need to diversify their business models, including revenue streams. Therefore, interest in raising significant funds from other sources is stronger than ever, leading to the need for a fundraising approach that is more sophisticated. Medical educators and health professionals are some of the most trusted members of society, and with this paper, the authors aim to raise awareness of the critical role they play in helping universities with their global impact and fundraising efforts.

(JMIR Med Educ 2022;8(2):e32597) doi:10.2196/32597

KEYWORDS

fundraising; philanthropy; crowdfunding; funding; charity; higher education; university; business model; revenue streams; medical education; educators; academia; academic environments

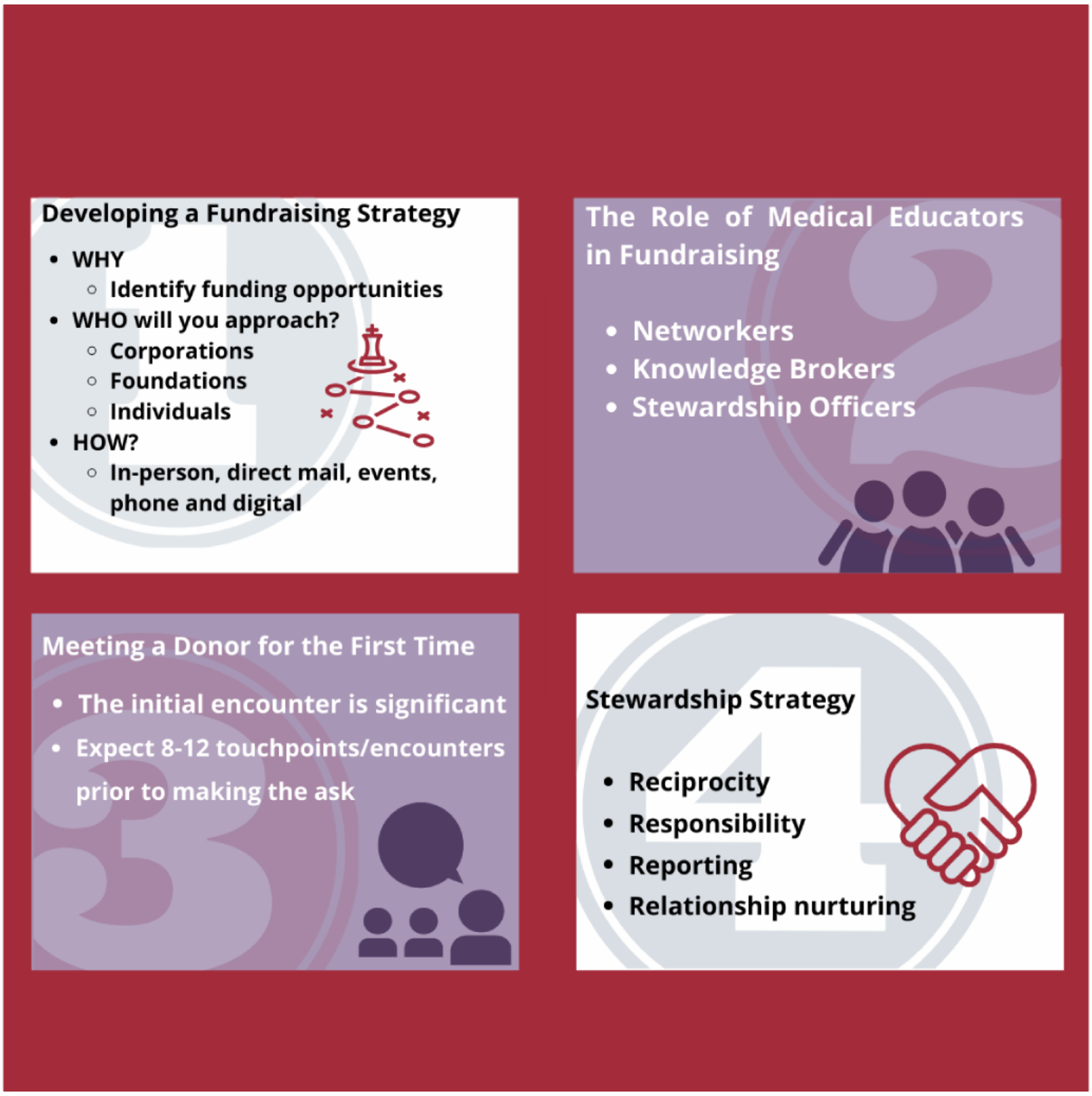
Introduction

This is a time of economic, societal, and political challenges in higher education. Therefore, universities today focus more than ever on fundraising. To diversify their traditional revenue streams, universities are increasing their international student recruitment efforts and are expanding their research boundaries as global actors and agents of change. Therefore, interest in fundraising from private sources is stronger than ever, leading to the need for a fundraising approach that is more sophisticated. As such, medical educators must become agents of change and reflect on strategies to ensure the successful implementation of fundraising programs in academic environments toward achieving maximum impact. *Fundraising* is defined as seeking financial support for a nonprofit organization or cause. However, in the Greek language, the word *philanthropy* means "love [Philos] of humanity [Anthropos]." Because philanthropy fuels the act of fundraising, fundraising is not only about raising money through tactics and transactions but also about changing lives, making an impact, and cultivating long-lasting and meaningful relationships with donors and the community.

Defining a Fundraising Road Map

The first step on this journey is to develop a *fundraising road map*: a plan of action that guides an institute's fundraising activities (Figure 1). A road map sets out the fundraising objectives and the ways the institute will meet them. Fundraising objectives should align with the institute's strategic plan and areas with the most significant impact. A fundraising plan should also consider donors' needs by connecting them to benefits that affect them, such as life-changing clinical research.

Figure 1. The fundraising road map.



To develop a successful road map, institutions need more than a vision and a strategic plan. Educators need to thoroughly understand their budgets, knowing why they need the funds and how much they need in advance. As such, some basic knowledge of fundraising is essential to any medical educator venturing into this field. First, the road map needs to be adapted to the amount of financial support required to fund a given project successfully. A sophisticated fundraising program will include an annual fund or a fundraising event aiming to acquire donors through a tactical approach typically aimed at raising funds up to US $25,000. Relationship-oriented major gift fundraising programs require significant development of authentic longer-term relationships with donors. These programs aim to raise gifts of US $25,000 or more, including top-level "principal" gifts of US $1 million or more, and even transformational giving that exceeds US $5 million. Donors may also consider leaving significant planned gifts, whether during their lifetimes or upon death, as part of an overall financial and estate plan. Gifts in kind, where goods and services are given as donations, are another option. Each of these types of fundraising programs requires a well-established fundraising strategy.

For example, a researcher approaches you with a crucial need for incubators for the new Advanced Medical Research Centre. After careful budget analysis, it appears there are insufficient funds to acquire this equipment. Estimating that a state-of-the-art incubator costs roughly US $30,000, you can begin to build a "major gift" fundraising campaign plan. However, if researchers require multiple incubators, you may need to target "principal" level donations depending on the amount. The key here is to clarify and confirm the exact funding requirement and align it to the giving capacity of your potential donors.

Developing a Fundraising Strategy

Your strategy defines who you will approach and in what way. This can be achieved by conducting a funding gap analysis to determine the amount needed to fund your project's ongoing operations or future development and to match interested donors to suitable initiatives. Most donations come from three types of donors: corporations (eg, corporate matching gift programs and volunteer grants), foundations (eg, grants from nongovernmental organizations), and individuals. Each of these donor types can be approached through any of the five main types of fundraising: in-person, direct mail, events, phone (SMS text messaging), and digital [1]. Recent studies have shown social media also plays an important role in fundraising, mostly in crowdfunding campaigns [2,3]. However, for major gift fundraising, personal, one-on-one solicitation remains the most successful means of gift solicitation.

The development of a well-researched fundraising strategy is essential to the success of one's efforts, and fund raisers must decide on donors and approaches carefully based on their needs and environment.

The Role of Medical Educators in Fundraising

Medical educators can become an essential part of the fundraising team because they provide a unique perspective from within the faculty. As expert witnesses to a fundraising strategy, they can elaborate on the faculty's vision and goals, provide details of programs, explain the impact of medical research, and often have the most significant working relationships with potential donors (this is key to prospect identification). Participating in fundraising expands upon their existing role of expert witnesses for their work or research funding [4].

Medical educators can play three roles in fundraising [5], which are as follows: (1) networker—by networking with patients, clinicians, researchers, and other educators, they can serve as a connection between the fundraising team and the groups mentioned above, as well as open doors within the medical community; (2) knowledge broker—as knowledge brokers, medical educators can serve as the fundraising team's academic lead and can provide access to knowledge about the faculty's strategic priorities, research innovations, and medical advances. In this way, they can assist the fundraising team in establishing links between donor interests and the strategic preferences of the institute; and (3) stewardship officer—in the capacity of stewardship officers, medical educators can meet with donors and alumni to update them on the latest institutional achievements and answer their questions.

Meeting a Donor for the First Time

The first encounter with a donor is significant. A few essential tips on a successful first meeting include engaging the prospect and listening more than talking about them. The process should focus on donors instead of the individuals doing the fundraising. Make a note of how many times you use the word "I" versus "you." The conversation should always be "you" dominant. During your encounters, always be enthusiastic, authentic, and passionate. Discuss your strengths, and bring along colleagues who might be able to help in areas of uncertainty. Always take notes (with permission) and end meetings by establishing the next steps. Typically, it takes 12-18 months to develop meaningful relationships that result in asking prospective donors to give, and within that period, 8-10 significant "moves" or actions, including in-person meetings, can occur. When ready to ask for a donation, ask for a specific amount and keep your sights high. The amount asked should be based on both the donor's wealth and the value of the project. Praise other peoples' commitments to your program without breaching any confidentiality.

Stewardship Strategy

Lastly, develop a customized stewardship strategy for the donor, which means having a systematic approach to cultivate and improve your relationships with donors. The four stewardship strategies are reciprocity (must demonstrate gratitude), responsibility (must act in a socially responsible manner), reporting (need to keep its public informed), and relationship nurturing (should make sure donors receive thank you letters and annual reports and are invited to special events). As the relationship strengthens, fundraisers may also send handwritten cards for special occasions [6].

Ethical Practice in Fundraising

A critical point for any medical educator venturing into fundraising is always to follow strict ethical rules regarding interaction with patients and learners. Unwarranted pressure on patients or learners to contribute should be avoided, and patient or learner confidentiality and trust should always be a priority. The main recommendation is for institutions to have a fundraising committee made of educators, physicians, and fundraisers to mitigate these risks and for fundraisers to follow ethical guidelines established by their organization [7].

In conclusion, in this modern age, having medical educators collaborate closely with fundraising teams is essential. Educators serve as the link between the donor, the fundraiser, and the life-changing medical advances toward which we work. As academic leaders and subject matter experts, medical educators bring imperative knowledge and connections vital to any fundraising campaign's success.

Conflicts of Interest

None declared.

References

1. Camilleri MA. Higher education marketing: Opportunities and challenges in the digital era. Academia 2019;16-17:4-28. [doi: 10.26220/aca.3169]
2. Ortiz RA, Witte S, Gouw A, Sanfilippo A, Tsai R, Fumagalli D, et al. Engaging a Community for Rare Genetic Disease: Best Practices and Education From Individual Crowdfunding Campaigns. Interact J Med Res 2018 Feb 05;7(1):e3. [doi: 10.2196/ijmr.7176]

3. Ren J, Raghupathi V, Raghupathi W. Understanding the Dimensions of Medical Crowdfunding: A Visual Analytics Approach. J Med Internet Res 2020 Jul 03;22(7):e18813 [FREE Full text] [doi: 10.2196/18813] [Medline: 32618573]
4. Byrnes JEK, Ranganathan J, Walker BLE, Faulkes Z. To Crowdfund Research, Scientists Must Build an Audience for Their Work. PLoS One 2014;9(12):e110329 [FREE Full text] [doi: 10.1371/journal.pone.0110329] [Medline: 25494306]
5. Nyman J, Pilbeam C, Baines P, Maklan S. Identifying the roles of university fundraisers in securing transformational gifts: lessons from Canada. Studies in Higher Education 2016 Oct 24;43(7):1227-1240. [doi: 10.1080/03075079.2016.1242565]
6. Waters RD. Measuring stewardship in public relations: A test exploring impact on the fundraising relationship. Public Relations Review 2009 Jun;35(2):113-119. [doi: 10.1016/j.pubrev.2009.01.012]
7. Prokopetz JJZ, Lehmann LS. Physicians as fundraisers: medical philanthropy and the doctor-patient relationship. PLoS Med 2014 Mar;11(2):e1001600 [FREE Full text] [doi: 10.1371/journal.pmed.1001600] [Medline: 24523665]

Edited by T Leung; submitted 03.08.21; peer-reviewed by FNU Amrita, PH Liao; comments to author 24.01.22; revised version received 28.02.22; accepted 18.03.22; published 05.04.22.

Please cite as:
Jalali A, Nyman JA, Hamelin-Mitchell E
Fundraising in Education: Road Map to Involving Medical Educators in Fundraising
JMIR Med Educ 2022;8(2):e32597
URL: https://mededu.jmir.org/2022/2/e32597
doi:10.2196/32597
PMID:35380542

Original Paper

Influence of Factors Relating to Sex and Gender on Rank List Decisions and Perceptions of Residency Training: Survey Study

Ryan Gibney[1], MD; Christina Cantwell[1], MSc; Alisa Wray[1], MAEd, MD; Megan Boysen-Osborn[1], MHPE, MD; Warren Wiechmann[1], MBA, MD; Soheil Saadat[1], MD, PhD; Jonathan Smart[1], MD; Shannon Toohey[1], MAEd, MD

Department of Emergency Medicine, University of California, Irvine Medical Center, Orange, CA, United States

Corresponding Author:
Shannon Toohey, MAEd, MD
Department of Emergency Medicine
University of California, Irvine Medical Center
101 The City Drive, Route 128-01
Orange, CA, 92868
United States
Phone: 1 714 456 5239
Fax: 1 714 456 3714
Email: stoohey@hs.uci.edu

Abstract

Background: Females make up more than half of medical school matriculants but only one-third of emergency medicine (EM) residents. Various factors may contribute to why fewer females choose the field of EM, such as the existing presence of females in the specialty.

Objective: This study is a follow-up to previous work, and a survey is used to assess current residents' attitudes and perceptions on various factors, including those relating to sex and gender on creating rank lists as medical students and in perceived effects on residency education.

Methods: A web-based survey consisting of Likert scale questions regarding a variety of factors influencing a student's decision to create a rank list and in perceived effects on residency education was sent to current EM residents in 2020.

Results: Residents from 17 programs participated in the survey with an 18.2% (138/758) response rate. The most important factors in creating a rank list were the personality of residents in the program, location, and facility type. For factors specifically related to gender, respondents who answered affirmatively to whether the gender composition of residents affected the selection of a program in making a rank list were more likely to also answer affirmatively to subsequent questions related to the gender of program leadership (P<.001) and gender composition of attending physicians (P<.001). The personality of residents was also the most important factor perceived to affect residency education. For factors influencing rank list and residency education, female respondents placed higher importance on subcategories related to gender (ie, gender composition of the residents, of the program leadership, and of the attending physicians) to a significant degree compared with their male counterparts.

Conclusions: Although factors such as location and resident personality show the most importance in influencing residency selection, when stratifying based on respondent sex, females tend to indicate that factors relating to gender have more influence on rank list and residency education compared with males.

(JMIR Med Educ 2022;8(2):e33592) doi:10.2196/33592

KEYWORDS

residency; sex; gender; graduate medical education; emergency medicine; residents; program leadership; rank list

Introduction

Background

Although females now make up more than half of medical school graduates, they compose only approximately one-third of emergency medicine (EM) residents [1,2]. It is unclear why fewer females choose to pursue EM than males. A possibility is the lack of availability of female mentorship among EM faculty. Although one could assume that female students may be more likely to attend an EM program with a higher proportion of female faculty, a study found that there was no correlation between the presence of women in leadership roles and the percentage of female residents in a program [1]. Still, women are the minority in academic medicine, with only 9.3% of the chair and 25.9% of the program director (PD) positions being held by women [1]. The presence of other female core faculty may have more influence on an applicant's decision to choose a program. Furthermore, using resident gender distribution may not be the best surrogate for determining whether faculty gender plays a role in an applicant's ranking decision because applicants consider many factors in their decisions and may not match with their top-choice program [2].

Objectives

The aim of this study is to determine whether residents feel that gender distribution is an important factor when choosing a residency program. This study is a follow-up to previous work examining sex ratios across EM programs of entering years from 2014 to 2017 (Gibney et al, unpublished data, 2021). The authors approached a cross-section of the programs identified in the previous study with varied sex diversity and asked their residents to complete a survey on factors that were important in residency selection and residency education to determine what sex or gender factors were perceived as important and if any factors showed differences in importance between males and females.

Throughout this paper we use the terms *male* and *female* to discuss topics relating to sex because of the limitations of the previous study upon which this work is based, which relied on publicly available data to calculate male-female ratios. Gender, on the other hand, refers to the social construct of masculinity or femininity, or *man*, *woman*, and *nonbinary*. In the survey we created and discuss in this paper, respondents were given the opportunity to self-identify their gender as nonbinary.

Methods

Recruitment and Survey

The University of California, Irvine Institutional Review Board registered this study and survey as exempt, given that it was an anonymous survey with minimal risk. We designed a survey using SurveyMonkey software (Momentive). The questionnaire was not externally validated; however, it was created by several educators in EM and trialed by 5 colleagues to ensure clarity and understanding. The survey questions which were sent to participants can be found in Multimedia Appendix 1. The survey was sent to programs willing to participate further in the study, with a target population of current EM residents at Accreditation Council for Graduate Medical Education–accredited US residency programs in October-November 2019 to serve as a representative pool of residents across the country. Voluntary response sampling was used in that the survey and study information sheet were sent to the program coordinator or PD who then forwarded the survey to their residents. A chance to win a US $100 Amazon gift card was offered as an incentive, and the survey was anonymous with an option to supply an email address at the end to enter the draw.

The survey was distributed electronically through an emailed link. The link was not publicly available or advertised. Contact with participants as a group, not individuals, was through the program coordinator or PD through email. Survey data were captured automatically when participants submitted their answers. The survey was voluntary, and participants could choose to stop answering at any time. The time frame for the survey was 30 days. Items were not randomized. Adaptive questioning was used to display the Likert scale only if a participant responded *Yes* to a question about whether a factor was important to them. The survey consisted of 2 pages. The first page contained 47 demographic data and yes or no questions that displayed Likert scales for questions that were answered *yes*. The second page had 13 questions with Likert responses if answered *yes*. A completeness check was performed by making the questions mandatory. Respondents were not able to change their responses with a Back or Review step. View, participation, and completion rates were not tracked. Cookies were used by the SurveyMonkey site to assign unique respondent IDs, and there were no duplicate entries. The IP address was not used to identify duplicates. Participants did not need to register or create a survey log-in. Incomplete questionnaires could be submitted. Atypical timestamp was not used to exclude data. Items were not weighted, nor were propensity scores used. The survey methods comply with CHERRIES (Checklist for Reporting Results of Internet E-Surveys) [3]. The programs that participated in the survey included a broad spectrum of sex distributions ranging from highly male-dominated to highly female-dominated ratios (one >3:1, three 2-3:1, ten 1-2:1, one 1:1, two <1:1, male:female).

Statistical Analysis

Categorical variables are presented as relative frequencies and continuous variables as mean (SD). We compared the association of categorical variables using a chi-square test. We compared the distribution of continuous variables among study groups by using the Mann–Whitney U test because they were not normalized. $P<.05$ was considered statistically significant. We used SPSS software (version 26.0; IBM Corp) for data analysis.

Results

Survey Demographics

Of the 171 EM residency programs included in our study, 17 (9.9%) agreed to participate in the follow-up survey on residency selection and education. Surveys were sent to the 758 residents, and 138 (18.2% response rate) responded. These respondents represented 17 programs across 10 states (California, Delaware, Florida, Iowa, Louisiana, Michigan, New Jersey, New York, North Carolina, and Texas) and the District of Columbia. The median age of the respondents was 28 (IQR 4) years. Of the 138 respondents, 56 (40.6%) were male, 81 (58.7%) were female, and 1 (0.7%) was nonbinary. The authors recognize the use of *male or female* terminology in the survey question self-identifying gender as a limitation of the study, and this is further addressed in the *Discussion* section.

Decision Factors in Determining Rank Lists

All residency selection factors and the rate at which respondents marked them as important are shown in Figure 1. The respondents noted the following as the most important factors when making their rank lists as fourth-year medical students: location (134/138, 97.1%), experience at the program (eg, interview day and externship; 133/138, 96.4%), personality of the residents in the program (133/138, 96.4%), reputation of the program (128/138, 92.7%), facility type (county vs academic vs private; 117/138, 84.8%), reputation or personality of faculty and attendings (116/138, 84.1%), variety of educational experiences (115/138, 83.3%), patient demographics (105/138, 76.1%), length of the program (98/138, 71%), annual patient visits (95/138, 68.8%), and schedule (shift length and numbers; 95/138, 68.8%).

After selecting the factors they used in determining their rank list, the respondents were asked to score on a Likert scale how important each of these factors was, with 1 being not very important and 4 being very important. Averages of the scored scales are reported in Figure 1. The single most important factor was the personality of residents in the program with a Likert scale average of 3.61 (SD 0.7). Other important factors included location (average 3.45, SD 0.8), facility type (average 3.41, SD 0.7), experience at program (average 3.36, SD 0.8), and patient demographics (average 3.34, SD 0.7).

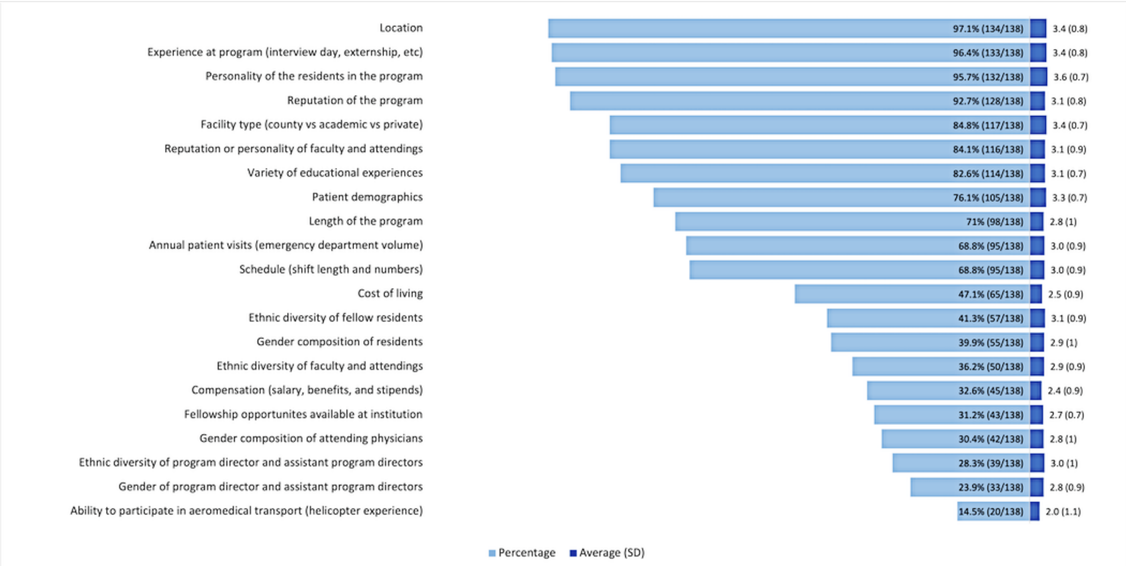


Figure 1. Decision factors in determining rank lists and average importance of rank list factors.

Decision Factors Relating to Gender and Sex Composition in Determining Rank Lists

Regarding the factors relating to gender makeup in the residency program and how they affected the rank list, the distribution of positive answers was as follows: gender composition of residents: 39.9% (55/138), 95% CI 31.6%-48.5%; gender composition of attending physicians: 30.4% (42/138), 95% CI 22.9%-38.8%; and gender of PD and assistant PDs (APDs): 23.9% (33/138), 95% CI 17.1%-31.1% (Figure 1). Gender composition of residents was more important than gender composition of PDs (*P*=.004), but the difference between the other categories listed was not statistically significant.

The survey evaluated whether the sex composition of residents, faculty, and leadership would affect residency selection if it was male or female predominant. Regarding the factors that would affect their selection of residency, of the 138 respondents, 64 (46.4%) indicated a program with only, or predominantly, male residents; 57 (41.3%) indicated a program with only, or predominantly, male faculty; and 56 (40.6%) indicated a program with only male leadership (PD or APDs). Of these respondents, 83% (53/64), 84% (48/57), and 89% (50/56), respectively, indicated at least a moderate effect on their selection of that program (Table 1).

In a subgroup analysis, respondents who answered affirmatively to whether the sex composition of residents affected selection of a program in making a rank list were more likely to also answer affirmatively to subsequent questions related to the sex of the PD and APDs (*P*<.001) and the sex composition of attending physicians (*P*<.001).

Table 1. Effect of male predominance among residents, faculty, and program leadership on rank list decisions.

	Values, n (%)
Would a program that has only or predominantly male residents affect your selection of residency? How much would it affect your selection of residency? (n=60)	
Major effect	29 (48)
Moderate effect	24 (40)
Mild effect	7 (12)
Would a program that has only or predominantly male faculty affect your selection of residency? How much would it affect your selection of residency? (n=56)	
Major effect	19 (34)
Moderate effect	29 (52)
Mild effect	8 (14)
Would a program that has only or predominantly male residency leadership (program director and assistant or associate program director) affect your selection of residency? How much would it affect your selection of residency? (n=54)	
Major effect	24 (44)
Moderate effect	26 (48)
Mild effect	4 (7)

Perceived Factors That Affect Residency Education

Next, we examined the factors that were perceived to affect residency education. Less than half of the respondents said that compensation was a factor that affected their residency education. The three factors that were indicated to have the least influence on residency education were gender of the PD and APDs, compensation, and ability to participate in aeromedical transport. The total number of respondents in this question set varies because of an incomplete data set used in analysis. The three most important factors were variety of education experiences (129/137, 94.2%), personality of residents in the program (129/138, 93.5%), and patient demographics (128/137, 93.4%; Figure 2). Respondents were again asked to score on a Likert scale how important their selected factors were, with 1 being not very important and 4 being very important. Scored averages are shown in Figure 2. Personality of residents in the program showed the highest average score when ranked on a 4-point Likert scale (average 3.58, SD 0.8; Figure 2). Other important factors included variety of educational experiences (average 3.41, SD 0.7), facility type (average 3.4, SD 0.6), and patient demographics (average 3.4, SD 0.8; Figure 2).

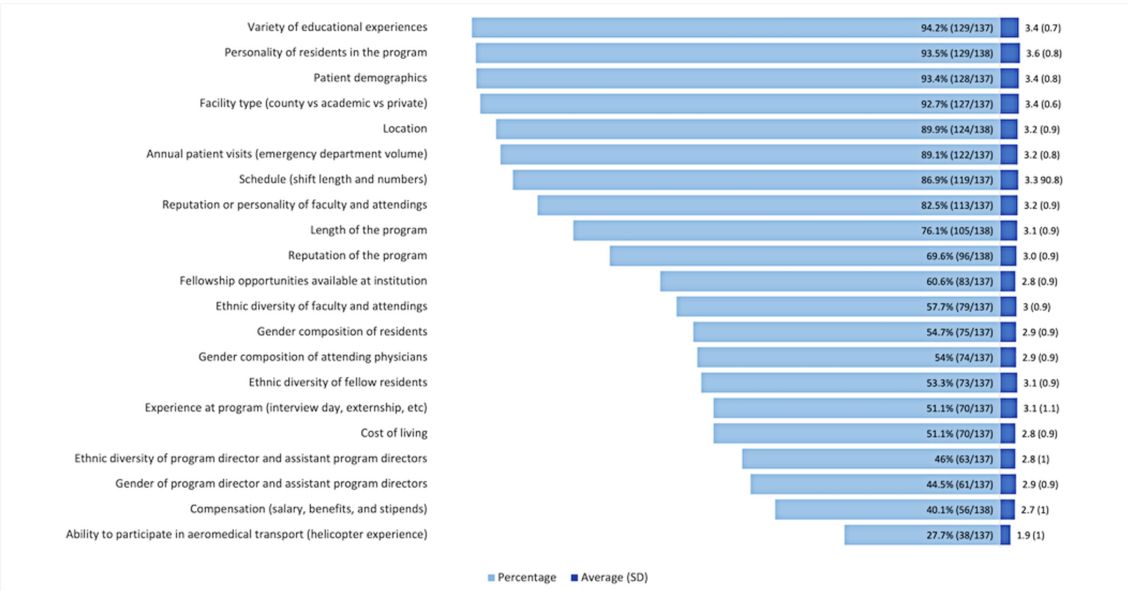


Figure 2. Factors perceived to affect residency education and their average importance.

Perceived Factors Relating to Gender and Sex That Affect Residency Education

The results of factors perceived as affecting residency education followed an identical pattern to those affecting residency selection. Gender composition of residents was more important than that of attendings, which in turn was more important than gender of the PD and APDs in having a perceived effect on one's residency education.

In the final set of questions, respondents were asked to select from >50%, >60%, >70%, >80%, and >90% to describe at what percentage they considered the faculty to be male or female predominant. The median at which a faculty was male predominant was >70%, whereas female predominant was indicated at only >60%. With respect to male- or female-predominant resident groups, the median was >60% for both. In establishing the point at which respondents felt that a faculty or residency was either male or female predominant, respondents showed a broader IQR in response to what percentage they considered to be female dominant. Respondents seemed to have a lower threshold for considering faculty or residency to be female predominant, with the IQR spanning from the minimum of >50% to >70%, whereas IQRs regarding male predominance were narrower, extending from >60% to >70%.

Factors Related to Residency Selection and Education Stratified by Respondent Sex

There were differences in responses to certain categories based on respondent sex. A chi-square analysis was performed with α of .05. Factors affecting residency selection and rank lists stratified by sex are presented in Table 2. Statistically significant differences were found in females placing more importance than males on the following factors: experience at the program on interview day (χ^2_1=7.5; *P*=.006; n=137), patient demographics (χ^2_1=5.0; *P*=.03; n=137), gender composition of residents (χ^2_1=13.8; *P*<.001; n=137), gender of PD and APDs (χ^2_1=9.3; *P*=.002; n=137), gender of attending physicians (χ^2_1=13.7; *P*<.001; n=137), and ethnic diversity of fellow residents (χ^2_1=4.3; *P*=.04; n=137; Table 2).

With respect to the perceived factors that affect residency education, females placed more importance than males on the following factors: gender of PD and APDs (χ^2_1=11.6; *P*=.001; n=136), gender composition of attending physicians (χ^2_1=6.1; *P*=.01; n=136), and ethnic diversity of PD and APDs (χ^2_1=6.9; *P*=.008; n=136; Table 3). In both categories of factors affecting rank lists and factors affecting residency education, females placed higher importance on subcategories specifically related to gender (ie, gender composition of residents, gender of PD and APDs, and gender composition of attending physicians).

Table 2. Pearson chi-square tests comparing distribution of responses between male and female respondents for factors relating to selection of residency programs. The chi-square statistic is significant at *P*=.05 level (N=137).

Factors relating to selection of residency programs and response	Respondent sex, n (%)			Chi-square (*df*)	Significance (*P* value)
	Female (n=81)	Male (n=56)	Total (N=137)		
Location					
No	3 (3.7)	1 (1.8)	4 (2.9)	0.4 (1)	—[a]
Yes	78 (96.3)	55 (98.2)	133 (97.1)	—	.51[b]
Reputation of the program					
No	5 (6.2)	5 (8.9)	10 (7.3)	0.4 (1)	—
Yes	76 (93.8)	51 (91.1)	127 (92.7)	—	.54[b]
Length of the program					
No.	27 (33.3)	12 (21.4)	39 (28.5)	2.3 (1)	—
Yes	54 (66.7)	44 (78.6)	98 (71.5)	—	.13
Compensation (salary, benefits, and stipends)					
No	56 (69.1)	36 (64.3)	92 (67.2)	0.4 (1)	—
Yes	25 (30.9)	20 (35.7)	45 (32.8)	—	.55
Personality of the residents in the program					
No	2 (2.5)	4 (7.1)	6 (4.4)	1.7 (1)	—
Yes	79 (97.5)	52 (92.9)	131 (95.6)	—	.19[b]
Experience at program (interview day, externship, etc)					
No	0 (0)	5 (8.9)	5 (3.6)	7.5 (1)	—
Yes	81 (100)	51 (91.1)	132 (96.4)	—	.006[b]
Fellowship opportunities available at institution					
No	55 (67.9)	39 (69.6)	94 (68.6)	0.5 (1)	—
Yes	26 (32.1)	17 (30.4)	43 (31.4)	—	.83
Patient demographics					
No	14 (17.3)	19 (33.9)	33 (24.1)	5.0 (1)	—
Yes	67 (82.7)	37 (66.1)	104 (75.9)	—	.03
Variety of educational experiences					
No	11 (13.6)	13 (23.2)	24 (17.5)	2.1 (1)	—
Yes	70 (86.4)	43 (76.8)	113 (82.5)	—	.15
Gender composition of residents					
No	38 (46.9)	44 (78.6)	82 (59.9)	13.8 (1)	—
Yes	43 (53.1)	12 (21.4)	55 (40.1)	—	<.001
Gender of program director and assistant program directors					
No	54 (66.7)	50 (89.3)	104 (75.9)	9.3 (1)	—
Yes	27 (33.3)	6 (10.7)	33 (24.1)	—	.002
Schedule (shift length and numbers)					
No	23 (28.4)	20 (35.7)	43 (31.4)	0.8 (1)	—
Yes	58 (71.6)	36 (64.3)	94 (68.6)	—	.36
Gender composition of attending physicians					
No	47 (58)	49 (87.5)	96 (70.1)	13.7 (1)	—
Yes	34 (42)	7 (12.5)	41 (29.9)	—	<.001

Factors relating to selection of residency programs and response	Respondent sex, n (%)			Chi-square (*df*)	Significance (*P* value)
	Female (n=81)	Male (n=56)	Total (N=137)		
Annual patient visits (emergency department volume)					
No	26 (32.1)	16 (28.6)	42 (30.7)	0.2 (1)	—
Yes	55 (67.9)	40 (71.4)	95 (69.3)	—	.66
Cost of living					
No	46 (56.8)	26 (46.4)	72 (52.6)	1.4 (1)	—
Yes	35 (43.2)	30 (53.6)	65 (47.4)	—	.23
Facility type (county vs academic vs private)					
No	14 (17.3)	7 (12.5)	21 (15.3)	0.6 (1)	—
Yes	67 (82.7)	49 (87.5)	116 (84.7)	—	.45
Ethnic diversity of fellow residents					
No	42 (51.9)	39 (69.6)	81 (59.1)	4.3 (1)	—
Yes	39 (48.1)	17 (30.4)	56 (40.9)	—	.04
Ethnic diversity of faculty and attendings					
No	48 (59.3)	40 (71.4)	88 (64.2)	2.1 (1)	—
Yes	33 (40.7)	16 (28.6)	49 (35.8)	—	.14
Ethnic diversity of program director and assistant program directors					
No	55 (67.9)	44 (78.6)	99 (72.3)	1.9 (1)	—
Yes	26 (32.1)	12 (21.4)	38 (27.7)	—	.17
Ability to participate in aeromedical transport (helicopter experience)					
No	71 (87.7)	46 (82.1)	117 (85.4)	0.8 (1)	—
Yes	10 (12.3)	10 (17.9)	20 (14.6)	—	.37
Reputation or personality of faculty and attendings					
No	10 (12.3)	11 (19.6)	21 (15.3)	1.4 (1)	—
Yes	71 (87.7)	45 (80.4)	116 (84.7)	—	.24

[a]Not available.

[b]More than 20% of the cells in this subtable have expected cell counts <5. Chi-square results may be invalid.

Table 3. Pearson chi-square tests comparing distribution of responses between male and female respondents for factors relating to education in residency programs. The chi-square statistic is significant at *P*=.05 level (N=136-137).

Factors relating to education in residency programs and response	Respondent sex, n (%)			Chi-square (*df*)	Significance (*P* value)
	Female (n=81)	Male (n=56)	Total (N=136-137)		
Location					
No	9 (11.1)	5 (8.9)	14 (10.2)	0.2 (1)	—[a]
Yes	72 (88.9)	51 (91.1)	123 (89.8)	—	.68
Reputation of the program					
No	20 (24.7)	22 (39.3)	42 (30.7)	3.3 (1)	—
Yes	61 (75.3)	34 (60.7)	95 (69.3)	—	.07
Length of the program					
No	23 (28.4)	10 (17.9)	33 (24.1)	2.0 (1)	—
Yes	58 (71.6)	46 (82.1)	104 (75.9)	—	.16
Compensation (salary, benefits, and stipends)					
No	48 (59.3)	34 (60.7)	82 (59.9)	0.3 (1)	—
Yes	33 (40.7)	22 (39.3)	55 (40.1)	—	.86
Personality of the residents in the program					
No	4 (4.9)	5 (8.9)	9 (6.6)	0.9 (1)	—
Yes	77 (95.1)	51 (91.1)	128 (93.4)	—	.35[b]
Experience at program (interview day, externship, etc)[c]					
No	36 (45)	30 (53.6)	66 (48.5)	1.0 (1)	—
Yes	44 (55)	26 (46.4)	70 (51.5)	—	.33
Fellowship opportunities available at institution[c]					
No	31 (38.8)	22 (39.3)	53 (39)	0.0 (1)	—
Yes	49 (61.3)	34 (60.7)	83 (61)	—	.95
Patient demographics[c]					
No	5 (6.3)	4 (7.1)	9 (6.6)	0.0 (1)	—
Yes	75 (93.8)	52 (92.9)	127 (93.4)	—	.84[b]
Variety of educational experiences[c]					
No	3 (3.8)	4 (7.1)	7 (5.1)	0.8 (1)	—
Yes	77 (96.3)	52 (92.9)	129 (94.9)	—	.38[b]
Gender composition of residents[c]					
No	31 (38.8)	31 (55.4)	62 (45.6)	3.7 (1)	—
Yes	49 (61.3)	25 (44.6)	74 (54.4)	—	.06
Gender of program director and assistant program directors[c]					
No	35 (43.8)	41 (73.2)	76 (55.9)	11.6 (1)	—
Yes	45 (56.3)	15 (26.8)	60 (44.1)	—	.001
Schedule (shift length and numbers)[c]					
No	8 (10)	10 (17.9)	18 (13.2)	1.8 (1)	—
Yes	72 (90)	46 (82.1)	118 (86.8)	—	.18
Gender composition of attending physicians[c]					

Factors relating to education in residency programs and response	Respondent sex, n (%)			Chi-square (*df*)	Significance (*P* value)
	Female (n=81)	Male (n=56)	Total (N=136-137)		
No	30 (37.5)	33 (58.9)	63 (46.3)	6.1 (1)	—
Yes	50 (62.5)	23 (41.1)	73 (53.7)	—	.01
Annual patient visits (emergency department volume)[c]					
No	8 (10)	7 (12.5)	15 (11)	0.2 (1)	—
Yes	72 (90)	49 (87.5)	121 (89)	—	.65
Cost of living[c]					
No	42 (52.5)	24 (42.9)	66 (48.5)	1.2 (1)	—
Yes	38 (47.5)	32 (57.1)	70 (51.5)	—	.27
Facility type (county vs academic vs private)[c]					
No	8 (10)	2 (3.6)	10 (7.4)	2.0 (1)	—
Yes	72 (90)	54 (96.4)	126 (92.6)	—	.16[b]
Ethnic diversity of fellow residents[c]					
No	32 (40)	31 (55.4)	63 (46.3)	3.1 (1)	—
Yes	48 (60)	25 (44.6)	73 (53.7)	—	.08
Ethnic diversity of faculty and attendings[c]					
No	30 (37.5)	28 (50)	58 (42.6)	2.1 (1)	—
Yes	50 (62.5)	28 (50)	78 (57.4)	—	.15
Ethnic diversity of program director and assistant program directors[c]					
No	36 (45)	38 (67.9)	74 (54.4)	6.9 (1)	—
Yes	44 (55)	18 (32.1)	62 (45.6)	—	.008
Ability to participate in aeromedical transport (helicopter experience)[c]					
No	58 (72.5)	40 (71.4)	98 (72.1)	0.2 (1)	—
Yes	22 (27.5)	16 (28.6)	38 (27.9)	—	.90
Reputation or personality of faculty and attendings[c]					
No	13 (16.3)	11 (19.6)	24 (17.6)	0.3 (1)	—
Yes	67 (83.8)	45 (80.4)	112 (82.4)	—	.61

[a]Not available.

[b]More than 20% of the cells in this subtable have expected cell counts <5. Chi-square results may be invalid.

[c]Sample size changes from 137 to 136 because of incomplete data set used in analysis.

Discussion

Principal Findings

We hypothesized that program leadership may influence a student's rank list in a way that tends to favor the propagation of a similar sex distribution among residents. Although medical school matriculation rates are nearly equal between male and female students, the percentage of females represented in academic medicine remains disparate [4]. Within EM, the Association of American Medical Colleges reports that as of 2015, women made up only 33% of the EM faculty, with only 17% representation seen among full professors [4]. Potential reasons that account for the disparity include a lack of mentors, greater work–life balance prioritization, and gender discrimination and bias [5]. Our findings on gender diversity within EM residency leadership was consistent with previous data showing 76% of the programs with male directors [6]; however, direct influence of the PD and faculty gender had not previously been evaluated.

To avoid bias regarding the topic of the survey, we included numerous aspects unrelated to gender in the questionnaire to allow survey respondents to appropriately consider all relevant aspects of selection and education. Many of the topics were similar to those found in the National Resident Matching Program (NRMP) Applicant Survey, and the factors that were considered most important were similar to those found in the 2019 NRMP Applicant Survey results for EM [7].

As a cohort, our survey results suggest that faculty and residency leadership gender makeup have minimal effect on residency selection. However, when stratified by respondent sex, our study found that female respondents viewed gender distribution of PD and APDs and attendings as influencing their residency selection and education more than their male counterparts. In programs with lower male–female ratios in leadership, incoming female residents may rank these institutions higher for the potential for female mentorship. A previous study on sex distribution in radiology residencies showed that programs with a female PD had a higher concentration of female residents [8], and a similar study in EM showed no difference [1]. However, in our study, these factors do not seem as important as others.

The relatively short time a medical student spends at a given program during a rotation or interview necessitates that the student must make inferences about how their own experience will be if matched into that program. The concept of homophily—the tendency to favor those like oneself—can potentially explain this trend. In the 8 programs with a higher percentage of female residents than male residents, program leadership also reflected a low male-female ratio, and all these programs had at least one female represented in leadership. In addition, residents who seek to perform research and achieve publication throughout their training may also place high value on such opportunities when determining rank lists. Although homophily has been described as being more commonly observed among females, the phenomenon has been observed in the publication realm and shown to be stronger in male journal editors [9]. Within EM publications, only 26% of the first authorships are female [10]. A study in 2011 reported that only 15.9% of the editors-in-chief and 17.5% of the editorial board members of 60 top-ranking journals were female [11]. Thus, if seeking research opportunities during residency, incoming female trainees may gravitate toward programs that visibly promote advancement and career development that is more favorable toward females than toward programs with predominantly or exclusively male leadership.

Sex distributions among current residents seem to play a role in creating a rank list. In residency programs that are predominantly male, incoming female residents may perceive a lack of fit in these programs because they do not see as many female colleagues. Our study shows that the primary gender makeup affecting applicants' ranking of programs was that of the residents. Female respondents indicated that the gender composition of residents influenced their rank list and education more than their male counterparts. Incoming female residents may view programs with a higher female presence more favorably because they can see themselves being successful. The same can be said for male residents as well. However, because EM as a specialty is predominantly male, females may be influenced to rank programs with a greater percentage of female residents higher because they see other females being successful as a resident at that institution. As Bandura [12] describes in his concept of self-efficacy, observing people similar to oneself be successful increases one's belief in achieving the same success. Therefore, in the recruitment of residents, programs should examine their own sex and gender makeup to determine ways to address cognitive biases that may result from a skewed distribution.

Limitations

Limitations to our study include having a low survey response rate (138/758, 18.2%) and low overall sample size; however, our sample was well distributed in terms of geographic location and age of respondents. Furthermore, for the non–gender-related items, the results are similar to the 2019 NRMP Applicant Survey results, suggesting that there may not be significant bias from the response rate. In addition, there is a factor of retrospective recall in that the residents were surveyed after matriculation, rather than at the time of ranking decisions. There is also response bias with having more females than males take the survey; therefore, the factors identified in our study may represent elements more important to females than to males. There is also the uncontrollable wildcard inherent in the NRMP. Programs may rank incoming residents with a nearly even male-female split, but, depending on algorithms and student choice, the sex ratios expected may not match the outcome. However, it is still important to recognize the possibility for existing sex distributions among residents and program leadership to influence a fourth-year medical student's decision to rank that program.

Another limitation of our study is the use of *male* and *female* as response options in self-identifying gender in our survey. Future work would benefit from clear distinctions in demographic data with separate questions for respondent sex (eg, male or female) and gender (eg, man, woman, and nonbinary) to capture characteristics of respondents more accurately.

Comparison With Prior Work

More recently, Mannix et al [13] examined sex distribution among chief residents in EM. The group found that females have increased representation among chief residents compared with their overall proportion among EM residents, with females and males having a similar presence in the chief positions. Our study did not examine the perceptions of having female representation among chief residents, although Mannix et al [13] also suggest that increased numbers of female chief residents will help bridge the sex gap in academic medicine and program leadership observed currently. Our data similarly show that current female residents place higher value on female leadership among PDs and APDs than male residents.

DeSantis and Marco [14] previously reported that friendliness, environment, and interview day experience were the top 3 factors that were important to residents in selecting their program. Our study echoed similar results, with aspects such as location, experience at the program, and personality of the residents having a significant influence on a resident's choice and education. As Laskey and Cydulka [15] reported in a 2009 study, female residents valued opportunities to serve specific populations as more important than their male counterparts, and in our study, we also found a similar trend with regard to patient demographics as being an influencing factor in rank list and residency education.

Conclusions

Follow-up surveys to quantify the importance of sex and gender in residency selection showed that other factors such as location, interview day experience, personality of residents, and educational experiences were rated as much more important than gender differences within a program. In stratifying results based on male and female respondents, female respondents tended to indicate that factors relating to gender had more influence on their decisions in creating a rank list and perceptions of residency education more often than male respondents.

Acknowledgments

The authors would like to thank the residents and their respective institutions who participated in this survey for their time and interest in the work.

Authors' Contributions

RG provided the idea for the study and design, acquisition and interpretation of data, and manuscript drafting. CC was involved with data acquisition, analysis, and interpretation, as well as manuscript drafting and submission preparation. AW, MBO, WW, and JS were involved with study design. SS performed data analysis and interpretation and provided statistical expertise. ST was involved in study concept, design, data interpretation, and advising. All authors provided critical revision of the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questions sent to participants.
[DOCX File , 17 KB - mededu_v8i2e33592_app1.docx]

References

1. DeFazio CR, Cloud SD, Verni CM, Strauss JM, Yun KM, May PR, et al. Women in emergency medicine residency programs: an analysis of data from accreditation council for graduate medical education-approved residency programs. AEM Educ Train 2017 Jul;1(3):175-178 [FREE Full text] [doi: 10.1002/aet2.10030] [Medline: 30051031]
2. Beran RL. National resident matching program. Gastroenterology 1999 Aug;117(2):P509. [doi: 10.1053/gast.1999.0029900509a]
3. 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 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
4. 2015-2016 - The state of women in academic medicine statistics. Association of American Medical Colleges, Washington, DC. 2016. URL: https://www.aamc.org/data-reports/faculty-institutions/data/2015-2016-state-women-academic-medicine-statistics [accessed 2020-07-17]
5. Edmunds LD, Ovseiko PV, Shepperd S, Greenhalgh T, Frith P, Roberts NW, et al. Why do women choose or reject careers in academic medicine? A narrative review of empirical evidence. Lancet 2016 Dec 10;388(10062):2948-2958. [doi: 10.1016/S0140-6736(15)01091-0] [Medline: 27105721]
6. Greenstein J, Hardy R, Chacko J, Husain A. Demographics and fellowship training of residency leadership in EM: a descriptive analysis. West J Emerg Med 2017 Jan;18(1):129-132 [FREE Full text] [doi: 10.5811/westjem.2016.10.31452] [Medline: 28116024]
7. Results of the 2019 NRMP applicant survey by preferred specialty and applicant type. National Resident Matching Program, Data Release and Research Committee, Washington, DC. 2019. URL: https://vdocument.in/results-of-the-2019-nrmp-applicant-survey-this-report-presents-survey-results.html [accessed 2022-02-02]
8. Campbell JC, Yoon SC, Cater SW, Grimm LJ. Factors influencing the gender breakdown of academic radiology residency programs. J Am Coll Radiol 2017 Jul;14(7):958-962. [doi: 10.1016/j.jacr.2017.02.045] [Medline: 28427906]
9. Callaham M. Gender bias and peer review: annals seeks greater diversity. Ann Emerg Med 2019 Dec;74(6):736-739. [doi: 10.1016/j.annemergmed.2019.10.018] [Medline: 31779950]
10. Kaji AH, Meurer WJ, Napper T, Nigrovic LE, Mower WR, Schriger DL, Annals of Emergency Medicine Diversity Task Force. State of the journal: women first authors, peer reviewers, and editorial board members at annals of emergency medicine. Ann Emerg Med 2019 Dec;74(6):731-735. [doi: 10.1016/j.annemergmed.2019.05.011] [Medline: 31280922]
11. Amrein K, Langmann A, Fahrleitner-Pammer A, Pieber TR, Zollner-Schwetz I. Women underrepresented on editorial boards of 60 major medical journals. Gend Med 2011 Dec;8(6):378-387. [doi: 10.1016/j.genm.2011.10.007] [Medline: 22153882]

12. Bandura A. Self-efficacy. In: Ramachandran VS, editor. Encyclopedia of Human Behavior. New York: Academic Press; 1994:71-81.
13. Mannix A, Parsons M, Krzyzaniak SM, Black LP, Alvarez A, Mody S, et al. Emergency Medicine Gender in Resident Leadership Study (EM GIRLS): the gender distribution among chief residents. AEM Educ Train 2020 Jul;4(3):262-265 [FREE Full text] [doi: 10.1002/aet2.10436] [Medline: 32704596]
14. DeSantis M, Marco CA. Emergency medicine residency selection: factors influencing candidate decisions. Acad Emerg Med 2005 Jun 01;12(6):559-561 [FREE Full text] [doi: 10.1197/j.aem.2005.01.006] [Medline: 15930408]
15. Laskey S, Cydulka RK. Applicant considerations associated with selection of an emergency medicine residency program. Acad Emerg Med 2009 Apr;16(4):355-359 [FREE Full text] [doi: 10.1111/j.1553-2712.2009.00361.x] [Medline: 19469049]

Abbreviations

APD: assistant program director
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
EM: emergency medicine
NRMP: National Resident Matching Program
PD: program director

Edited by T Leung; submitted 15.09.21; peer-reviewed by E Mohammadi, L Suppan; comments to author 19.10.21; revised version received 29.10.21; accepted 07.01.22; published 05.04.22.

Please cite as:
Gibney R, Cantwell C, Wray A, Boysen-Osborn M, Wiechmann W, Saadat S, Smart J, Toohey S
Influence of Factors Relating to Sex and Gender on Rank List Decisions and Perceptions of Residency Training: Survey Study
JMIR Med Educ 2022;8(2):e33592
URL: https://mededu.jmir.org/2022/2/e33592
doi:10.2196/33592
PMID:35380547

Original Paper

Factors Associated With Specialists' Intention to Adopt New Behaviors After Taking Web-Based Continuing Professional Development Courses: Cross-sectional Study

Lysa Bergeron[1,2], BSc, MSc; Simon Décary[3], BSc, MSc, PhD; Codjo Djignefa Djade[1], BSc, MSc; Sam J Daniel[4,5], MSc, MD; Martin Tremblay[4], BSc, MSc, PhD; Louis-Paul Rivest[6], BSc, MSc, PhD; France Légaré[1,7], BSc, MD, MSc, PhD, CCMF

[1]VITAM - Centre de recherche en santé durable, Centre intégré universitaire de santé et de services sociaux de la Capitale-Nationale, Quebec, QC, Canada

[2]Department of Social and Preventive Medicine, Faculty of Medicine, Université Laval, Quebec, QC, Canada

[3]School of Rehabilitation, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada

[4]Direction du Développement Professionnel Continu, Fédération des Médecins Spécialistes du Québec, Montreal, QC, Canada

[5]Department of Pediatric Surgery, McGill University, Montreal, QC, Canada

[6]Department of Mathematics and Statistics, Faculty of Science and Engineering, Université Laval, Quebec, QC, Canada

[7]Department of Family Medicine and Emergency Medicine, Faculty of Medicine, Université Laval, Quebec, QC, Canada

Corresponding Author:
France Légaré, BSc, MD, MSc, PhD, CCMF
VITAM - Centre de recherche en santé durable
Centre intégré universitaire de santé et de services sociaux de la Capitale-Nationale
2480, chemin de la Canardière
Quebec, QC, G1G2G1
Canada
Phone: 1 4186635313
Email: france.legare@fmed.ulaval.ca

Abstract

Background: Web-based continuing professional development (CPD) is a convenient and low-cost way for physicians to update their knowledge. However, little is known about the factors that influence their intention to put this new knowledge into practice.

Objective: We aimed to identify sociocognitive factors associated with physicians' intention to adopt new behaviors as well as indications of Bloom's learning levels following their participation in 5 web-based CPD courses.

Methods: We performed a cross-sectional study of specialist physicians who had completed 1 of 5 web-based CPD courses offered by the Federation of Medical Specialists of Quebec. The participants then completed CPD-Reaction, a questionnaire based on Godin's integrated model for health professional behavior change and with evidence of validity that measures behavioral intention (dependent variable) and psychosocial factors influencing intention (n=4). We also assessed variables related to sociodemographics (n=5), course content (n=9), and course format (eg, graphic features and duration) (n=8). Content variables were derived from CanMEDS competencies, Bloom's learning levels, and Godin's integrated model. We conducted ANOVA single-factor analysis, calculated the intraclass correlation coefficient (ICC), and performed bivariate and multivariate analyses.

Results: A total of 400 physicians participated in the courses (range: 38-135 physicians per course). Average age was 50 (SD 12) years; 56% (n=223) were female, and 44% (n=177) were male. Among the 259 who completed CPD-Reaction, behavioral intention scores ranged from 5.37 (SD 1.17) to 6.60 (SD 0.88) out of 7 and differed significantly from one course to another (P<.001). The ICC indicated that 17% of the total variation in the outcome of interest, the behavioral intention of physicians, could be explained at the level of the CPD course (ICC=0.17). In bivariate analyses, social influences (P<.001), beliefs about capabilities (P<.001), moral norm (P<.001), beliefs about consequences (P<.001), and psychomotor learning (P=.04) were significantly correlated with physicians' intention to adopt new behaviors. Multivariate analysis showed the same factors, except for social influences and psychomotor learning, as significantly correlated with intention.

Conclusions: We observed average to high behavioral intention scores after all 5 web-based courses, with some variations by course taken. Factors affecting physicians' intention were beliefs about their capabilities and about the consequences of adopting new clinical behaviors, as well as doubts about whether the new behavior aligned with their moral values. Our results will inform design of future web-based CPD courses to ensure they contribute to clinical behavior change.

(JMIR Med Educ 2022;8(2):e34299) doi:10.2196/34299

KEYWORDS

continuing professional development; CPD-Reaction; behavioral intention; medical specialists; continuing professional development; web-based training; medical education; education; physician; psychosocial; online course

Introduction

Improving patient outcomes requires that health professionals constantly adjust their practices in light of new evidence. Continuing professional development (CPD) is one of the most common strategies for achieving this, and indeed is a requirement for continued practice in many countries, including Canada [1]. The use of web-based CPD increased 10-fold from 2002 to 2008 in the United States and continues to grow rapidly [2], paralleling the increasing use of other web-based tools by health professionals [3]. A combination of the high costs of in-person CPD and the sanitary measures imposed during the COVID-19 pandemic have accelerated this increase in web-based CPD [4,5]. It is not clear whether web-based CPD has a real impact on clinical practice [6-9] or if physicians want CPD delivered this way [10]. However, in times of pandemic there is little choice, and some of the advantages of distance learning have been highlighted in this context [11]. High-quality CPD courses should translate the evidence presented not only into new awareness, but also into new practices; yet most studies only evaluate their impact on clinical practice using measures of satisfaction and changes in knowledge [6,12,13]. Several meta-analyses on the impact of CPD on physician performance have recommended that new research should focus less on whether CPD is effective and more on why it is effective [14]. This requires a better understanding of the theory-based mechanisms underlying the impact of web-based CPD courses on clinical practice [15]. Future courses could then be based on these evidence-based and theory-informed mechanisms.

Sociocognitive theories describe these mechanisms by identifying key variables and the interrelationship of determinants in predicting health behaviors [16]. Studies based on such theories provide the empirical evidence to guide many behavior-change interventions. Learning is based not only on absorbing information, but on other factors that produce social behaviors, such as social modeling and a personal sense of control [16-19]. To ensure that CPD courses lead to physicians adopting the desired behavior in clinical practice, it is essential that they be informed by sound, theory-based factors known to influence the adoption of a given behavior [20-22]. According to Godin's integrated model for health professional behavior change [17], behavioral intention is the central factor influencing the adoption of a given behavior. In turn, this intention is influenced by a number of other sociocognitive factors. Incorporating these modifiable sociocognitive factors in the design of CPD has proven acceptable and feasible [23] and holds great promise for improved clinical practices [15,24].

Godin's comprehensive list of these factors, gathered from evidence produced by multiple studies in numerous domains, informed our CPD-Reaction tool, designed to assess behavioral intention after CPD activities [19,25]. The questionnaire consists of 12 items related to intention and 4 of the following influences on intention: (1) social influence (perception of approval or disapproval by persons significant to the individual regarding the adoption of the behavior); (2) beliefs about capabilities (belief that one is capable of performing the behavior); (3) moral norm (feeling of personal obligation regarding the adoption of the behavior); and (4) beliefs about consequences (perception that the behavior will have harmful or beneficial effects). For CPD to result in adopting a new clinical or organizational practice, "deep" learning also needs to occur [26]. Many CPD developers use Bloom's taxonomy to design the learning objectives of CPD activities, which also provides measures for their effects [27]. Bloom's taxonomy is related to Kirkpatrick's model, one of the best-known models for analyzing and evaluating the results of training programs [28]. Bloom's taxonomy provides additional detail by defining 3 domains of learning, which are affective, psychomotor, and cognitive. Affective learning relates to attitudes, psychomotor to physical skills, and cognitive to six learning levels, each of increasing "depth" or complexity [29].

Therefore, to address the lack of theory-informed assessment of CPD activities, we aimed to identify sociocognitive factors associated with physicians' intention as well as indications of Bloom's learning levels following their participation in 5 web-based CPD courses.

Methods

Study Design

We performed a cross-sectional study of a convenience sample of specialist physicians who had completed 1 of 5 different web-based CPD courses [30]. We report data according to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) reporting guidelines for cross-sectional studies [31]. Data were collected between November 2015 and April 2019 following completion of the CPD courses by participants using a web-based interactive platform (MÉDUSE) designed by the Federation of Medical Specialists of Québec (FMSQ) [32]. The FMSQ consists of 35 medical associations and represents 59 medical specialties in the province of Quebec. Its members include more than 10,000 medical specialists [33].

Ethics Approval

Approval for this study was obtained from the research ethics boards of the Centre intégré universitaire de santé et de services sociaux (CIUSSS) de la Capitale-Nationale (Project 2020-1889_SPPL).

Study Participants

To be eligible, the physicians had to have completed (1) one of the 5 available FMSQ CPD web-based courses and (2) the CPD-Reaction questionnaire.

CPD Courses

The 5 web-based courses were all accredited by the Royal College of Physicians and Surgeons of Canada [34] and targeted the following five behaviors: (1) to adapt the frequency of cytological exams for gynecological patients 25-45 years old to new human papillomavirus recommendations; (2) to use recommended lung cancer treatment and monitoring algorithms; (3) to use a systematic leadership approach in community health endeavors (eg, preventing instances of suicide from a bridge in Montreal); (4) to respect best practices in record keeping; and (5) to identify patients who meet the criteria for identifying a potential organ donor. Courses were free of commercial support (paid by FMSQ members' annual contributions). They were secured and accessible 24/7 by Quebec specialist physicians. The courses lasted from 90 to 120 minutes, and participants could stop or reinitiate courses at any time. Course objectives were based on Godin's integrated model for health professionals' behavior and aimed to encourage physicians to adopt new behaviors (or cease old ones). Each targeted behavior was designed according to 3 of the TACT principles: "target," "action," and "context" ("time" was excluded, as the targeted behaviors were not dependent on a specific time frame) [35]. The courses were also designed to develop core competencies as described in the CanMEDS Competency Framework [36]. CanMEDS is a framework created by the Royal College to ensure that CPD courses, regardless of their specialist content, allow physicians to develop one or more of the following core roles: medical expert, communicator, collaborator, leader, health advocate, scholar, or professional. We inserted the relevant learning objectives into each of the 5 CPD-Reaction questionnaires and attached them to the end of each respective course (Figure S1 in Multimedia Appendix 1).

Data Collection Procedure

Data were collected in 2 separate databases by the FMSQ, one for the sociodemographic variables of those attending each course (henceforth referred to as "participants") and another for those who had completed CPD-Reaction (henceforth referred to as "respondents"). Individual participant sociodemographic data could not be linked to individual respondent CPD-Reaction questionnaire scores and were analyzed at the level of the CPD course. For sociodemographic variables and variables collected at the course level (content and format of the CPD courses), the same values were then attributed to all respondents in the same course.

At the Level of Participants

Psychosocial Determinant Variables (1 Dependent Variable and 4 Independent Variables)

The CPD-Reaction is a self-administered questionnaire based on sociocognitive theories of behavioral change. The questionnaire had been developed and validated earlier with participants in 18 different CPD activities and had a Cronbach α ranging from .77 to .85 [22,25]. The CPD-Reaction questionnaire consists of 12 items grouped into the five following constructs: (1) behavioral intention (dependent variable; 2 items); (2) beliefs about capabilities (3 items); (3) social influences (3 items); (4) beliefs about consequences (2 items); and (5) moral norm (2 items). The specific clinical behavior targeted by the CPD course is inserted into each item of the questionnaire. There is no overall score for CPD-Reaction. The score for each construct is computed as the average of each item (Likert scale of 1, which is low, to 7, which is high), except for social influence, which is rated on a Likert scale of 1 to 5 [37]. Thus, a moral norm score of 7, for example, indicates that the respondent feels a strong obligation to adopt this behavior, while a score of 1 for beliefs about capabilities indicates that the respondent does not feel confident in their ability to adopt the behavior [37]. All physicians who completed CPD courses were invited to fill out the CPD-Reaction questionnaire afterward.

At the Level of the CPD Courses

Participant Profile (Sociodemographic) Variables (5 Independent Variables)

All participants (n=400) provided information about their age, number of years in practice, sex (female or male), their medical association (clinical area), and administrative region.

Characteristics of Course Content (9 Independent Variables)

Two coders independently noted the presence of slides in each CPD course in which they could identify the following elements: (a) targeting of a CanMEDS role—medical expert, communicator, collaborator, leader, health advocate, scholar, or professional (when more than one role was targeted, the reference category was "not applicable"); (b) Bloom's learning levels; and (c) constructs of Godin's theoretical framework for the study of health care professionals' behavior and intention [17,27,36].

Characteristics of Course Format (8 Independent Variables)

Informed by literature on presentation of material for optimal learning [38,39], 2 coders independently assessed the presence of the following factors: use of virtual characters, use of a reflective approach, duration of the course, presence of nonfunctional links to references, presence of slides with a video of a health professional (opinion leader), presence of slides with a figure or a diagram, presence of slides with a quiz, and presence of women on the scientific committee for the course development.

Data Analysis

The data set had a hierarchical structure consisting of 2 levels, which were respondents and CPD courses. Individual participant sociodemographic data were in a distinct database and could not be linked to individual respondent CPD-Reaction questionnaire scores. Therefore, analysis focused primarily on variables at the level of CPD courses. These variables were only retained if they could be collected for all 5 courses. All variables retained at the course level had fewer than 1% missing values. At the respondent level (n=259), only the dependent variable, intention, and the 4 independent psychosocial variables (social influences, beliefs about capabilities, moral norm, and beliefs about consequences) were accessible and analyzed. At the CPD course level (n=5 CPD courses), data analyzed included the 5 sociodemographic variables of participants, 8 course format variables, and 9 course content variables.

We used descriptive statistics and frequency counts to describe all variables. We performed an ANOVA single-factor analysis to assess whether the topic of the courses had an impact on intention. We also computed the intraclass coefficient (ICC) to assess the percentage of variance in behavioral intention and its psychosocial determinants attributable to the CPD course [40]. We performed exploratory bivariate analysis using Spearman correlations for each one of our independent variables at the level of the CPD courses to assess their association with physicians' intention scores. Lastly, we performed bivariate and multivariate analysis on the 4 psychosocial determinants of intention at the respondent level (n=259) to explore their impact on intention scores. We used a linear model and introduced a random effect for the CPD courses. A threshold of .05 was set for statistical significance. We verified all assumptions for the linear regression model [41]. All analyses were performed with SAS, version 9.4 (SAS Institute Inc).

Results

Characteristics of the Participants

Table 1 shows the characteristics of participants from across Quebec attending each course.

Table 1. Profile of the participants.

Variables	Courses[a]					
	1	2	3	4	5	All courses
Participants, n (%)	96 (24)	60 (15)	71 (18)	135 (34)	38 (9)	400 (100)
Age (years), mean (SD)	48 (11)	49 (11)	51 (13)	52 (12)	46 (12)	50 (12)
Gender, n (%)						
Female	70 (73)	20 (33)	43 (61)	74 (55)	16 (42)	223 (56)
Male	26 (27)	40 (67)	28 (39)	61 (45)	22 (58)	177 (44)
Main clinical area for each course, n (%)	67 (70)[b]	24 (40)[c]	11 (15)[d]	23 (17)[e]	7 (18)[f]	87 (22)[b]
Most frequent administrative area of main practice site, n (%)	18 (19)[g]	18 (30)[h]	14 (20)[h]	35 (26)[g]	6 (16)[h]	77 (19)[g]
Years of practice, mean (SD)	19 (13)	19 (13)	20 (14)	22 (14)	14 (13)	19 (14)

[a]Course details: behaviors used in the questionnaire for each course; course 1—to adapt the frequency of cytological examinations according to new recommendations; course 2—to use lung cancer treatment and monitoring algorithms; course 3—to use a systematic leadership approach; course 4—to respect good practices in record keeping; and course 5—to identify patients who meet the criteria for identifying a potential organ donor.

[b]Obstetrics and gynecology.

[c]Pneumology.

[d]Preventive medicine.

[e]Psychiatry.

[f]Anesthesiology.

[g]Montérégie.

[h]Montreal.

CPD Course Characteristics

Table 2 shows details of course characteristics, including course formatting, content, presence of Godin's constructs, and Bloom's learning levels. Three courses lasted 90 minutes and 2 lasted 120 minutes. Two out of 5 courses focused on the CanMEDS role of medical expert, 1 on the role of leader, 1 on the role of professional, and 1 on several of the roles at once (classified as "not applicable"). Moreover, 3 course characteristics showed no variability—all contained slides with a quiz (format variable), slides on beliefs about capabilities, and slides on role and identity (content variables; Table 2).

Table 2. CPD[a] course characteristics.

Variables	Course[b]				
	1	2	3	4	5
Characteristics of course format					
Virtual character use in course	Yes	No	Yes	Yes	No
Use of a reflective approach	Yes	No	No	Yes	Yes
Duration, min	90	120	120	90	90
Presence of nonfunctional references	Yes	No	No	No	Yes
Presence of slides with a video of health professional—leader opinion	No	No	No	No	Yes
Presence of slides with a figure or a diagram	No	Yes	Yes	Yes	Yes
Presence of slides with a quiz	Yes	Yes	Yes	Yes	Yes
Presence of women on the scientific committee	No	—[c]	—	Yes	Yes
Characteristics of course content					
Main CanMED role	Medical expert	Medical expert	Leader	Professional	N/A[d]
Presence of slides per constructs of Godin's integrated model for health professional behavior change					
Intention	Yes	Yes	Yes	No	Yes
Social influences	No	No	Yes	No	Yes
Beliefs about capabilities	Yes	Yes	Yes	Yes	Yes
Role and identity	Yes	Yes	Yes	Yes	Yes
Beliefs about consequences	Yes	No	Yes	Yes	Yes
Presence of which level of Bloom's taxonomy					
Cognitive	Yes	Yes	Yes	Yes	No
Affective	No	No	No	No	Yes
Psychomotor	No	No	No	Yes	Yes

[a]CPD: continuing professional development.

[b]Course details: course 1—to adapt the frequency of cytological exams for gynecological patients 25-45 years old to new Human Papillomavirus recommendations; course 2—to use recommended lung cancer treatment and monitoring algorithms; course 3—to use a systematic leadership approach in community health endeavors (eg, preventing suicides from a bridge in Montreal); course 4—to respect best practices in record keeping; and course 5—to identify patients who meet the criteria for identifying a potential organ donor.

[c]Not available.

[d]N/A: not applicable.

CPD-Reaction Questionnaires Scores and ICC

Of the 400 participants, 259 (65%) respondents fully completed the CPD-Reaction questionnaire. Table 3 shows details of respondents' mean intention and psychosocial determinants and SD scores for each course. The behavioral intention score was medium to high and varied depending on the course undertaken (intention score between 5.37, SD 1.17 and 6.60, SD 0.88). ANOVA analysis of the variable intention showed significant differences between courses (*F* value=12.50, *P*<.001). The ICC (0.17) indicated that 17% of the total variation in the behavioral intention of physicians to adopt new behaviors could be explained at the level of the course in which they had registered. Some courses showed significantly different means of intention, that is, respondents shared more intracourse similarities than extracourse similarities (ie, within CPD courses vs between courses).

Table 3. CPD-Reaction questionnaire mean scores and ICC[a].

Variables	Courses[b]					Mean (SD)	ICC
	1	2	3	4	5		
Number of respondents, n (%)	53 (20)	44 (17)	63 (24)	61 (24)	38 (15)	N/A[c]	N/A
Psychosocial determinants[d], mean (SD)							
Intention	6.22 (1.15)	6.22 (0.87)	5.37 (1.17)	6.57 (1.11)	6.60 (0.88)	6.15 (1.16)	16.8
Social influences	5.44 (0.97)	5.41 (1.12)	4.39 (0.98)	5.17 (0.95)	5.16 (1.33)	5.07 (1.12)	13.0
Beliefs about capabilities	6.28 (0.80)	6.22 (0.75)	5.07 (0.81)	6.16 (0.89)	6.25 (1.11)	5.94 (0.99)	25.9
Moral norm	6.58 (0.66)	6.35 (0.82)	5.94 (0.93)	6.61 (1.11)	6.68 (0.80)	6.41 (0.93)	8.98
Beliefs about consequences	6.28 (0.94)	6.30 (0.94)	5.99 (0.91)	6.72 (0.47)	6.68 (0.77)	6.38 (0.86)	10.9

[a]ICC: intraclass correlation coefficient.

[b]Course details: course 1—to adapt the frequency of cytological exams for gynecological patients 25-45 years old to new Human Papillomavirus recommendations; course 2—to use recommended lung cancer treatment and monitoring algorithms; course 3—to use a systematic leadership approach in community health endeavors (eg, preventing suicides from a bridge in Montreal); course 4—to respect best practices in record keeping; and course 5—to identify patients who meet the criteria for identifying a potential organ donor.

[c]N/A: not applicable.

[d]Score range 1-7.

Factors Associated With Physicians' Intention to Adopt a New Behavior

Only one of the course variables, psychomotor learning level (Bloom's taxonomy), was significantly associated with the physicians' intention to change their behavior, and this was the case in all 5 courses (R=0.89, P=.04) (data not shown). Bivariate regression analysis of psychosocial determinants showed that all 4 variables were significantly associated with intention (P<.001) (Table 4). Multivariate regression analysis of the same variables showed 3 out of the 4 were significantly correlated with intention, namely beliefs about capabilities (0.49, P<.001), moral norm (0.37, P<.001), and beliefs about consequences (0.40, P<.001) (Table 4). When we analyzed the courses separately, we found similar results (Multimedia Appendix 2).

Table 4. Bivariate regression analysis and multivariate regression analysis of psychosocial determinants associated with intention to adopt a clinical behavior (n=259 respondents).

Variables	β	95% CI	*P* value
Bivariate regression analysis			
Social influences	.42	0.31 to 0.53	<.001
Beliefs about capabilities	.95	0.86 to 1.04	<.001
Moral norm	.82	0.71 to 0.93	<.001
Beliefs about consequences	.80	0.67 to 0.92	<.001
Multivariate regression analysis			
Social influences	–.04	–0.12 to 0.04	.30
Beliefs about capabilities	.49	0.37 to 0.62	<.001
Moral norm	.37	0.27 to 0.48	<.001
Beliefs about consequences	.40	0.30 to 0.50	<.001

Discussion

Principal Findings

We identified factors associated with physicians' intention to adopt new behaviors following the completion of 5 different web-based CPD courses. Behavioral intention scores were average to high but differed significantly from one course to another. The differences between CPD courses (higher level in our hierarchical database) explained a significant proportion of this variance in intention (ICC=0.17). We saw no influence of course characteristics (content-wise or format-wise) on intention except the targeting of Bloom's psychomotor learning level. Finally, we observed that together, beliefs about capabilities, moral norm, and belief about consequences (3 of the psychosocial variables included in Godin's integrated model for health professional behavior change) partially explained physicians' behavioral intentions.

Significance and Comparison With Prior Work

First, we found that behavioral intention scores were average to high but varied by course. Some courses seemed to be associated with higher physician intention to adapt their practice compared with others. Course 3, on using a systematic leadership approach in community health endeavors, had the lowest intention score of all—although this course also targets more complex and ambiguous outcomes than the others, and leadership skills are difficult to develop in 120 minutes. We also found that the variance in intention explained by the difference in CPD courses had significant magnitude. We obtained an ICC of 0.17; thus, the intergroup variance represented 17% of the total variance. Interestingly, higher ICCs are more often seen in studies in specialty settings than in primary care studies [42]. It is possible that specialist physicians have more in common with each other, even diverse specialist physicians attending the same course, than do general practitioners. Some studies have observed that medical professional culture ensures there is more similarity than diversity within specific medical specialties [43-45].

Second, in bivariate analyses at the CPD course level, the only variable significantly correlated with the intention to adopt new behaviors was targeting Bloom's psychomotor learning level. This level of learning, unlike the cognitive or affective levels, is more closely related to physical changes in behavior. However, this variable was not retained in our final model, suggesting that as an influence on adoption of new behaviors, it does not supersede the psychosocial variables included in the integrated model [17]. Regarding the other nonsignificant variables, previous studies have also found little significant association between sociodemographic characteristics and intention to adopt new behaviors [46,47]. Our results validate the assumption of the integrated model for health professional behavior change: modifiable psychosocial factors are the variables most likely to explain behavior change, and CPD courses should therefore focus on these factors to be more effective.

Third, we found that the 3 variables most significantly associated with intention to adopt a behavior among respondents were all psychosocial factors included in our integrated theoretical framework—beliefs about capabilities, moral norm, and beliefs about consequences (ie, their confidence about adopting the behavior, its ethical acceptability, and their perception that the behavior would be useful and beneficial). Based on our results, CPD courses should use behavior change techniques that focus on these 3 variables [48,49]. To improve beliefs about capabilities, courses could provide more experience to give participants confidence in their abilities, such as identifying barriers and management strategies, providing feedback, and encouraging monitoring of future actions (eg, noting and recording when the new behaviors have been adopted) [17,50]. To improve beliefs about consequences, courses could provide information about the proven benefits of the behavior and personalized information about its consequences. Regarding moral norm, courses could emphasize the felt obligation to adopt behaviors or help participants focus on moral considerations such as being aware of others' needs [51]. In addition, according to Godin's theory, when people hold two ideas that are not psychologically consistent, to reduce cognitive dissonance, they do all in their power to change them until they become consistent [17]. One way to reduce cognitive dissonance is to solicit arguments from the subject in favor of the behavior to be adopted even if they are against it [52]. While producing such arguments may cause discomfort, the subject will ultimately adjust their initial attitude to be more consistent with the arguments they fabricated in favor of the behavior. Surprising as it may seem, when we are led to act contrary to our convictions, we tend to justify our actions, and we adapt our opinions to our behavior. Other work on "provisional selves" suggests that playing a role with which one is unfamiliar, or even against which one resists, opens new moral possibilities and can help one envisage adapting one's current role or adopting new ones [53,54]. Including this technique in a CPD course would be an interesting challenge.

Limitations and Strengths

This was a cross-sectional study, which limited our interpretation to assuming that attending the CPD courses improved intention scores. Indeed, we are unaware if respondents already had moderate-to-high intention to adopt these behaviors before completing the CPD courses—using the CPD-Reaction questionnaire both before and after the course would have better indicated a change in intention due to the course topic. A future study with a more robust study design (eg, pre-post controlled trial) could further verify the impact of courses [55]. Moreover, to increase power, we brought data from all 5 CPD courses (each targeting a different clinical behavior) into one hierarchical data set. Although aggregating data on distinct behaviors is not always advisable [56,57], this limitation was mitigated by respect for the theory archetypes that structure the study. In addition, our sensitivity analysis (Multimedia Appendix 2) showed similar results to those obtained with the aggregated database. The literature suggests that at least 30 units at each level of analysis are needed to reach sufficient power [58]. New ways to assess CPD courses are needed as few individual CPD courses recruit hundreds of participants. Moreover, sociodemographic data collected at the group level could not be applied to the respondent level (ie, to the individual level). Inferring results of analysis at the upper level (where determinants and outcomes are related at the group [course] level) to the individual level (ecological fallacy) or the reverse (atomistic fallacy) can result in bias [59]. Finally, intention is recognized as a limited proxy for behavior. Meta-analytic syntheses have found that intention accounts, on average, for only about 25% of the variance in behavior [35,60], although finding other reliable measures of behavior is challenging [19,56]. While a 2006 review by Eccles et al [61] "provide[d] encouragement for the contention that there is a predictable relationship between the intentions of a health professional and their subsequent behaviour," CPD activities making use of the determinants of intention as dependent variables should also integrate methods to close the intention-behavior gap such as audit and feedback, eliciting of implementation intentions ("if-then" plans), commitment to change statements, and supervision to support clinicians in following through on their intentions [62-64].

Conclusions

Beliefs about capabilities, moral norm, and belief about consequences partially explained physicians' behavioral intention. To address these beliefs, CPD activities could focus on building physicians' confidence about overcoming obstacles and on strategies for helping them align moral values with new behaviors, as well as providing information about their proven benefits.

As mentioned in our previous work [12], the use of CPD-Reaction helps CPD developers reflect on the nature of their training objectives in relation to the impact they seek. This study provides insights as to how to optimize physicians' intention to adopt a new behavior as a result of web-based CPD activities. CPD-Reaction contains the relevant theory-informed and validated items needed to assess intention and its determinants for CPD developers targeting clinical behavior change.

Acknowledgments

The authors thank Louisa Blair for editing this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CPD-Reaction questionnaire.

[PDF File (Adobe PDF File), 169 KB - mededu_v8i2e34299_app1.pdf]

Multimedia Appendix 2

Bivariate regression analysis and multivariate regression analysis of psychosocial determinants associated with intention to change a clinical behavior for each continuing professional development course (n=5). Course topic 1: to adapt the frequency of cytological exams for gynecological patients 25-45 years old to new human papillomavirus recommendations; course topic 2: to use recommended lung cancer treatment and monitoring algorithms; course topic 3: to use a systematic leadership approach in community health endeavors (eg, preventing suicides from a bridge in Montreal); course topic 4: to respect best practices in record keeping; and course topic 5: to identify patients who meet the criteria for identifying a potential organ donor.

[PDF File (Adobe PDF File), 39 KB - mededu_v8i2e34299_app2.pdf]

References

1. A Concise Guide to Maintenance of Certification. Royal College of Physicians and Surgeons of Canada. URL: https://www.royalcollege.ca/rcsite/cpd/moc-program/concise-guide-maintenance-certification-e [accessed 2021-10-01]
2. Sullivan LG. A systematic review to compare the effectiveness of face-to-face versus online (including blended learning) delivery of CME/CPD for healthcare practitioners (HCPs). University of Queensland. 2017. URL: https://bit.ly/3lpsRxK [accessed 2022-05-03]
3. Archambault PM, van de Belt TH, Grajales FJ, Faber MJ, Kuziemsky CE, Gagnon S, et al. Wikis and collaborative writing applications in health care: a scoping review. J Med Internet Res 2013 Oct 08;15(10):e210 [FREE Full text] [doi: 10.2196/jmir.2787] [Medline: 24103318]
4. O'Brien Pott M, Blanshan AS, Huneke KM, Baasch Thomas BL, Cook DA. Barriers to identifying and obtaining CME: a national survey of physicians, nurse practitioners and physician assistants. BMC Med Educ 2021 Mar 19;21(1):168 [FREE Full text] [doi: 10.1186/s12909-021-02595-x] [Medline: 33740962]
5. Gravas S, Ahmad M, Hernández-Porras A, Furriel F, Alvarez-Maestro M, Kumar A, Office of Education and SIU Board of Directors. Impact of COVID-19 on medical education: introducing homo digitalis. World J Urol 2021 Jun 29;39(6):1997-2003 [FREE Full text] [doi: 10.1007/s00345-020-03417-3] [Medline: 32860535]
6. Samuel A, Cervero RM, Durning SJ, Maggio LA. Effect of continuing professional development on health professionals' performance and patient outcomes: a scoping review of knowledge syntheses. Acad Med 2021 Jun 01;96(6):913-923. [doi: 10.1097/ACM.0000000000003899] [Medline: 33332905]
7. Lahti M, Hätönen H, Välimäki M. Impact of e-learning on nurses' and student nurses knowledge, skills, and satisfaction: a systematic review and meta-analysis. Int J Nurs Stud 2014 Jan;51(1):136-149. [doi: 10.1016/j.ijnurstu.2012.12.017] [Medline: 23384695]
8. Fredericks S, Martorella G, Catallo C. A systematic review of web-based educational interventions. Clin Nurs Res 2015 Feb 26;24(1):91-113. [doi: 10.1177/1054773814522829] [Medline: 24571963]
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] [Medline: 18780847]
10. van der Keylen P, Tomandl J, Wollmann K, Möhler R, Sofroniou M, Maun A, et al. The online health information needs of family physicians: systematic review of qualitative and quantitative studies. J Med Internet Res 2020 Dec 30;22(12):e18816 [FREE Full text] [doi: 10.2196/18816] [Medline: 33377874]

11. Slater BJ, Kashyap MV, Calkins CM, Powell D, Rothstein DH, Clifton M, et al. Global dissemination of knowledge through virtual platforms: Reflections and recommendations from APSA/IPEG. J Pediatr Surg 2022 Jan 15:ahead of print [FREE Full text] [doi: 10.1016/j.jpedsurg.2022.01.006] [Medline: 35086680]
12. Légaré F, Freitas A, Thompson-Leduc P, Borduas F, Luconi F, Boucher A, et al. The majority of accredited continuing professional development activities do not target clinical behavior change. Academic Medicine 2015;90(2):197-202. [doi: 10.1097/acm.0000000000000543]
13. Rouleau G, Gagnon M, Côté J, Payne-Gagnon J, Hudson E, Dubois C, et al. Effects of e-Learning in a continuing education context on nursing care: systematic review of systematic qualitative, quantitative, and mixed-studies reviews. J Med Internet Res 2019 Oct 02;21(10):e15118 [FREE Full text] [doi: 10.2196/15118] [Medline: 31579016]
14. Cervero RM, Gaines JK. The impact of CME on physician performance and patient health outcomes: an updated synthesis of systematic reviews. J Contin Educ Health Prof 2015;35(2):131-138. [doi: 10.1002/chp.21290] [Medline: 26115113]
15. Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, 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 [FREE Full text] [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]
16. Conner M, Norman P. Predicting Health Behaviour. Berkshire, UK: McGraw-Hill Education; 2005.
17. Godin G. Le comportement clinique des professionnels de la santé: déterminants d'adoption et méthodes de changement. Université Laval. 2012 Jan 19. URL: http://research.chusj.org/getmedia/c47e1a6f-1206-4750-ad85-efd1abb9a460/recherche-clinique_URCA_presentation-conferencier_gaston-godin.pdf.aspx [accessed 2022-05-03]
18. Godin G, Bélanger-Gravel A, Amireault S, Vohl M, Pérusse L. The effect of mere-measurement of cognitions on physical activity behavior: a randomized controlled trial among overweight and obese individuals. Int J Behav Nutr Phys Act 2011 Jan 11;8(1):2 [FREE Full text] [doi: 10.1186/1479-5868-8-2] [Medline: 21223565]
19. 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 Jul 16;3:36 [FREE Full text] [doi: 10.1186/1748-5908-3-36] [Medline: 18631386]
20. Accme data report shows increasing growth and diversity in accredited continuing medical education. Accreditation Council for Continuing Medical Education. 2018 Jul 12. URL: http://www.accme.org/news-releases/2017-data-report [accessed 2022-05-03]
21. Casper ES. The theory of planned behavior applied to continuing education for mental health professionals. Psychiatr Serv 2007 Oct;58(10):1324-1329. [doi: 10.1176/ps.2007.58.10.1324] [Medline: 17914010]
22. Légaré F, Borduas F, Jacques A, Laprise R, Voyer G, Boucher A, et al. Developing a theory-based instrument to assess the impact of continuing professional development activities on clinical practice: a study protocol. Implement Sci 2011 Mar 07;6(1):17 [FREE Full text] [doi: 10.1186/1748-5908-6-17] [Medline: 21385369]
23. Légaré F, Stacey D, Brière N, Fraser K, Desroches S, Dumont S, et al. Healthcare providers' intentions to engage in an interprofessional approach to shared decision-making in home care programs: a mixed methods study. J Interprof Care 2013 May;27(3):214-222 [FREE Full text] [doi: 10.3109/13561820.2013.763777] [Medline: 23394265]
24. Goulet F, Hudon E, Gagnon R, Gauvin E, Lemire F, Arsenault I. Effects of continuing professional development on clinical performance: Results of a study involving family practitioners in Quebec. Canadian Family Physician 2013;59(5):518-525. [doi: 10.1111/ceo.12106]
25. Légaré F, Freitas A, Turcotte S, Borduas F, Jacques A, Luconi F, et al. Responsiveness of a simple tool for assessing change in behavioral intention after continuing professional development activities. PLoS One 2017 May 1;12(5):e0176678 [FREE Full text] [doi: 10.1371/journal.pone.0176678] [Medline: 28459836]
26. Moon J. Using reflective learning to improve the impact of short courses and workshops. J Contin Educ Health Prof 2004;24(1):4-11. [doi: 10.1002/chp.1340240103] [Medline: 15069907]
27. Armson H, Elmslie T, Roder S, Wakefield J. Is the cognitive complexity of commitment-to-change statements associated with change in clinical practice? An application of Bloom's taxonomy. J Contin Educ Health Prof 2015;35(3):166-175. [doi: 10.1002/chp.21303] [Medline: 26378422]
28. Kirkpatrick D. The Four Levels of Evaluation: Measurement and Evaluation. Alexandria, VA: American Society for Training & Development; 2007.
29. Armstrong P. Bloom’s Taxonomy. Vanderbilt University Center for Teaching. 2010. URL: https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/ [accessed 2022-05-03]
30. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. The Lancet 2002 Jan;359(9300):57-61. [doi: 10.1016/s0140-6736(02)07283-5]
31. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med 2007 Oct 16;4(10):e296 [FREE Full text] [doi: 10.1371/journal.pmed.0040296] [Medline: 17941714]
32. Training. Fédération des médecins spécialistes du Québec. URL: https://fmsq.org/en/our-profession/training [accessed 2021-10-01]
33. Qui sommes nous? Fédération des médecins spécialistes du Québec. URL: https://fmsq.org/fr/propos-de-nous/qui-sommes-nous [accessed 2021-10-01]

34. National Standard for Support of Accredited CPD Activities. Royal College of Physicians and Surgeons of Canada. URL: https://www.royalcollege.ca/rcsite/cpd/providers/tools-resources-accredited-cpd-providers/national-standard-accredited-cpd-activities-e [accessed 2021-10-01]
35. Ajzen I. The theory of planned behavior: Frequently asked questions. Human Behav and Emerg Tech 2020 Jun 10;2(4):314-324. [doi: 10.1002/hbe2.195]
36. About CanMEDS. The Royal College of Physicians and Surgeons of Canada. URL: http://canmeds.royalcollege.ca/en/about [accessed 2021-10-01]
37. Légaré F, Borduas F, Freitas A, Turcotte S. The Continuing Professional Development (CPD) Reaction Questionnaire User Manual. Université Laval. 2015. URL: https://www.dropbox.com/s/f7nrygvd835fz65/CPD%20Reaction_User%20Manual.pdf?dl=0# [accessed 2022-05-03]
38. Martha ASD, Santoso HB. The design and impact of the pedagogical agent: a systematic literature review. Journal of Educators Online 2019 Jan;16(1):n1. [doi: 10.9743/jeo.2019.16.1.8]
39. Mayer RE. Applying the science of learning: Evidence-based principles for the design of multimedia instruction. American Psychologist 2008;63(8):760-769. [doi: 10.1037/0003-066x.63.8.760]
40. Ben Charif A, Croteau J, Adekpedjou R, Zomahoun HTV, Adisso EL, Légaré F. Implementation research on shared decision making in primary care: inventory of intracluster correlation coefficients. Med Decis Making 2019 Aug;39(6):661-672. [doi: 10.1177/0272989X19866296] [Medline: 31423898]
41. Poole M, O'Farrell P. The assumptions of the linear regression model. In: Transactions of the Institute of British Geographers. London, UK: Academic Journal Offprint; Mar 1971:145.
42. Campbell MK, Fayers PM, Grimshaw JM. Determinants of the intracluster correlation coefficient in cluster randomized trials: the case of implementation research. Clin Trials 2005 Sep 03;2(2):99-107. [doi: 10.1191/1740774505cn071oa] [Medline: 16279131]
43. López Cabrera MV, Olivares Olivares SL, Heredia Escorza Y. Professional culture in medical schools: a medical educator interpretation. Med Sci Educ 2020 Mar 03;30(1):281-286 [FREE Full text] [doi: 10.1007/s40670-019-00896-x] [Medline: 34457668]
44. Al-Rumayyan A, Van Mook WNKA, Magzoub ME, Al-Eraky MM, Ferwana M, Khan MA, et al. Medical professionalism frameworks across non-Western cultures: A narrative overview. Medical Teacher 2017 Mar 05;39(sup1):S8-S14. [doi: 10.1080/0142159x.2016.1254740]
45. Wyatt TR, Rockich-Winston N, White D, Taylor TR. "Changing the narrative": a study on professional identity formation among Black/African American physicians in the U.S. Adv Health Sci Educ Theory Pract 2021 Mar 22;26(1):183-198. [doi: 10.1007/s10459-020-09978-7] [Medline: 32572728]
46. Boland L, Lawson ML, Graham ID, Légaré F, Dorrance K, Shephard A, et al. Post-training shared decision making barriers and facilitators for pediatric healthcare providers: a mixed-methods study. Acad Pediatr 2019;19(1):118-129. [doi: 10.1016/j.acap.2018.05.010] [Medline: 29860134]
47. Unni EJ, Lian N, Kuykendall W. Understanding community pharmacist perceptions and knowledge about HIV preexposure prophylaxis (PrEP) therapy in a Mountain West state. J Am Pharm Assoc (2003) 2016;56(5):527-532.e1. [doi: 10.1016/j.japh.2016.05.004] [Medline: 27594106]
48. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
49. 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 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
50. Lucero KS, Chen P. What do reinforcement and confidence have to do with it? A systematic pathway analysis of knowledge, competence, confidence, and intention to change. J Eur CME 2020 Oct 12;9(1):1834759 [FREE Full text] [doi: 10.1080/21614083.2020.1834759] [Medline: 33133769]
51. Godin G, Conner M, Sheeran P. Bridging the intention-behaviour 'gap': the role of moral norm. Br J Soc Psychol 2005 Dec;44(Pt 4):497-512. [doi: 10.1348/014466604X17452] [Medline: 16368016]
52. Vaidis D, Halimi-Falkowicz S. La théorie de la dissonance cognitive: une théorie âgée d'un demi-siècle. Revue électronique de Psychologie sociale 2007;1:9-18 [FREE Full text]
53. Hennekam S, Ladge J, Shymko Y. From zero to hero: An exploratory study examining sudden hero status among nonphysician health care workers during the COVID-19 pandemic. J Appl Psychol 2020 Oct;105(10):1088-1100. [doi: 10.1037/apl0000832] [Medline: 32897086]
54. Ibarra H. Provisional selves: experimenting with image and identity in professional adaptation. Administrative Science Quarterly 1999;44(4):764-791. [doi: 10.2307/2667055]
55. Campbell DT, Stanley JC. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally College Publishing Company; 1963.
56. Eccles MP, Hrisos S, Francis JJ, Stamp E, Johnston M, Hawthorne G, et al. Instrument development, data collection, and characteristics of practices, staff, and measures in the Improving Quality of Care in Diabetes (iQuaD) Study. Implement Sci 2011 Jun 09;6(1):61 [FREE Full text] [doi: 10.1186/1748-5908-6-61] [Medline: 21658211]

57. Eccles MP, Hrisos S, Francis JJ, Steen N, Bosch M, Johnston M. Can the collective intentions of individual professionals within healthcare teams predict the team's performance: developing methods and theory. Implement Sci 2009 May 05;4:24 [FREE Full text] [doi: 10.1186/1748-5908-4-24] [Medline: 19416543]
58. Maas CJM, Hox JJ. Sufficient sample sizes for multilevel modeling. Methodology 2005 Jan;1(3):86-92. [doi: 10.1027/1614-2241.1.3.86]
59. Mackenbach J. Roaming through methodology. XXVI. The ecological fallacy and its less well-known counterpart, the atomistic fallacy. Nederlands tijdschrift voor geneeskunde 2000;144(44):2097-2100. [doi: 10.4135/9781412983907.n612]
60. Webb TL, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull 2006 Mar;132(2):249-268. [doi: 10.1037/0033-2909.132.2.249] [Medline: 16536643]
61. Eccles MP, Hrisos S, Francis J, Kaner EF, Dickinson HO, Beyer F, et al. Do self- reported intentions predict clinicians' behaviour: a systematic review. Implement Sci 2006 Nov 21;1(1):28 [FREE Full text] [doi: 10.1186/1748-5908-1-28] [Medline: 17118180]
62. Trivasse H, Webb TL, Waller G. A meta-analysis of the effects of training clinicians in exposure therapy on knowledge, attitudes, intentions, and behavior. Clin Psychol Rev 2020 Aug;80:101887. [doi: 10.1016/j.cpr.2020.101887] [Medline: 32736152]
63. Toli A, Webb TL, Hardy GE. Does forming implementation intentions help people with mental health problems to achieve goals? A meta-analysis of experimental studies with clinical and analogue samples. Br J Clin Psychol 2016 Mar;55(1):69-90. [doi: 10.1111/bjc.12086] [Medline: 25965276]
64. Arnold Rehring SM, Steiner JF, Reifler LM, Glenn KA, Daley MF. Commitment to change statements and actual practice change after a continuing medical education intervention. J Contin Educ Health Prof 2021 Mar 16;41(2):145-152. [doi: 10.1097/ceh.0000000000000340]

Abbreviations

CPD: continuing professional development
FMSQ: Federation of Medical Specialists of Québec
ICC: intraclass correlation coefficient
STROBE: Strengthening in the Reporting of Observational Studies in Epidemiology

Edited by T Leung; submitted 15.10.21; peer-reviewed by P van der Keylen, DS Mulkalwar, A Joseph; comments to author 09.01.22; revised version received 02.03.22; accepted 26.04.22; published 02.06.22.

Please cite as:
Bergeron L, Décary S, Djade CD, Daniel SJ, Tremblay M, Rivest LP, Légaré F
Factors Associated With Specialists' Intention to Adopt New Behaviors After Taking Web-Based Continuing Professional Development Courses: Cross-sectional Study
JMIR Med Educ 2022;8(2):e34299
URL: https://mededu.jmir.org/2022/2/e34299
doi:10.2196/34299
PMID:35476039

Original Paper

Anesthesiologists With Advanced Degrees in Education: Qualitative Study of a Changing Paradigm

Anuj Aggarwal[1], MD; Olivia Hess[2], BSc; Justin L Lockman[3], MD, MSED; Lauren Smith[1], MD; Mitchell Stevens[4], PhD; Janine Bruce[5], PhD; Thomas Caruso[1], MD, MEd

[1]Department of Anesthesiology, Perioperative, and Pain Medicine, Stanford University School of Medicine, Stanford, CA, United States

[2]Stanford University School of Medicine, Stanford, CA, United States

[3]Department of Anesthesiology and Critical Care Medicine, Children's Hospital of Philadelphia, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, United States

[4]Graduate School of Education, Stanford University, Stanford, CA, United States

[5]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, United States

Corresponding Author:
Anuj Aggarwal, MD
Department of Anesthesiology, Perioperative, and Pain Medicine
Stanford University School of Medicine
300 N Pasteur Dr
Stanford, CA, 94305
United States
Phone: 1 650 723 6415
Email: akaggarw@stanford.edu

Abstract

Background: Anesthesiology education has undergone profound changes over the past century, from a pure clinical apprenticeship to novel comprehensive curricula based on andragogic learning theories. Combined with institutional and regulatory requirements, these new curricula have propagated professionalization of the clinician-educator role. A significant number of clinician-educator anesthesiologists, often with support from department chairs, pursue formal health professions education (HPE) training, yet there are no published data demonstrating the benefits or costs of these degrees to educational leaders.

Objective: This study aims to collect the experiences of anesthesiologists who have pursued HPE degrees to understand the advantages and costs of HPE degrees to anesthesiologists.

Methods: Investigators performed a qualitative study of anesthesiologists with HPE degrees working at academic medical centers. Interviews were thematically analyzed via an iterative process. They were coded using a team-based approach, and representative themes and exemplary quotations were identified.

Results: Seven anesthesiologists were interviewed, representing diverse geographic regions, subspecialties, and medical institutions. Analyses of interview transcripts resulted in the following 6 core themes: outcomes, extrinsic motivators, intrinsic motivators, investment, experience, and recommendations. The interviewees noted the advantages of HPE training for those wishing to pursue leadership or scholarship in medical education; however, they also noted the costs and investment of time in addition to preexisting commitments. The interviewees also highlighted the issues faculty and chairs might consider for the optimal timing of HPE training.

Conclusions: There are numerous professional and personal benefits to pursuing HPE degrees for faculty interested in education leadership or scholarship. Making an informed decision to pursue HPE training can be challenging when considering the competing pressures of clinical work and personal obligations. The experiences of the interviewed anesthesiologists offer direction to future anesthesiologists and chairs in their decision-making process of whether and when to pursue HPE training.

(JMIR Med Educ 2022;8(2):e38050) doi:10.2196/38050

KEYWORDS

academic medical centers; trends; medical education; medical; faculty; anesthesiologists; medical professionals; learning; institute; clinician; educator; experience; decision-making; training

Introduction

Since the advent of anesthesiology, a combination of societal, economic, and higher education movements has influenced the adoption of formal learning theories by anesthesiology educators [1]. Initially, anesthesiology education centered around a clinical apprenticeship. However, the advancement of surgical techniques and the drive for reliable anesthetic techniques resulted in the restructuring of anesthesiology programs over the second half of the twentieth century [2,3]. Although pedagogical learning theories originally dominated anesthesiology training, the end of the last century marked the introduction of andragogic and experiential learning theories that accompanied the expansion of formal curricula for anesthesiology trainees (Table 1) [4].

These new curricula, combined with expanding institutional and regulatory requirements, have propagated a professionalization of a clinician-educator role for anesthesiologists. Many academic physicians have transitioned from the traditional clinician-researcher-educator role, and instead pursue educational scholarship as the central facet of their academic careers [5]. In the United States, this evolution has been influenced by the Accreditation Council for Graduate Medical Education (ACGME), which states that graduate medical education program directors should have at least 3 years of educational or administrative experience [6]. Though this statement by the ACGME calls for a broad allowance of education and administrative experiences, department leadership have recognized the value of formal training in education. Over the past several decades, this environment has propagated the development of health professions education (HPE) programs that are specifically designed for physicians, including certificate, master, and doctoral programs. There are now more than 150 such programs worldwide, with significant variation in content delivery (online, asynchronous, hybrid, or in-person), duration, and curriculum [7-9].

Additionally, there has recently been an emphasis on blended training environments that incorporate nonclinical interests, physician wellness, and social justice [1]. This progression of learning paradigms is also contributing to the increased growth of HPE programs [8,10]. Despite the rising number of anesthesiologists seeking HPE, there are limited data demonstrating the benefits and costs of these programs [9]. Many anesthesiologists who are interested in education-oriented careers are without appropriate guidance about whether to pursue an HPE degree. This is further compounded by the variety of programs, ranging from certification programs at local institutions to formal degree–granting programs.

There is an emerging community of anesthesiologists who have attained HPE degrees, along with others who are in the process of completing HPE programs, which grant master's degrees or higher. This qualitative study investigates the experiences of anesthesiology educators who have completed HPE master's degrees. It seeks to understand the influence these degrees have had on their professional and personal advancement and demonstrate common, valuable elements for future anesthesiologists wishing to pursue HPE programs.

Table 1. Examples of andragogy and pedagogy learning styles.

Andragogy	Pedagogy
Self-directed learner	Teacher-dependent learner
Greater life experiences	Little to no life experience
Learning determined by social roles	Learning determined by teacher
Problem centered	Content centered
Intrinsic motivation	Extrinsic motivation

Methods

Study Design and Research Characteristics

We performed a prospective, semistructured interview qualitative study with thematic analysis. The study design and reporting adhere to the standards described by the Standard for Reporting Qualitative Research guidelines [11]. All participants were approached via direct, electronic solicitation and were informed that participation was voluntary. Study participation offered no direct benefit to the participants. The interviewers obtained verbal consent from all study participants.

Context

Investigators interviewed participating anesthesiologists who worked at academic medical centers representing different regions of the country and different anesthesiology subspecialties. All interviews were conducted remotely using video conference software (Zoom Video Communications) to maintain safe social distancing, as interviews were performed during the COVID-19 pandemic. The interviews occurred from August through October of 2020.

Sampling Strategy

The study team sought to collect a sample of up to 12 anesthesiologists or until thematic saturation occurred. Participants were identified based on national reputation and professional relationships representing different institutions and different subspecialities [12]. Nonprobability sampling is typical in qualitative research, where the goal is not to randomly select from the population, but rather to purposefully identify and select relevant individuals [13]. This homogenous group sampling provides a rich understanding of their experiences [13]. We selectively recruited academic anesthesiologists who had earned a master's degrees in HPE. During qualitative research, data collection and analysis often occur simultaneously [13]. Sample size is considered adequate when little additional information is emerging from the interviews, which began at the 6th interview, resulting in a total of 7 participants.

Ethics Approval

The Stanford University Institutional Review Board granted a waiver for this study (Protocol #57512).

Data Collection Methods

Prior to each interview, we collected demographic data via questionnaires administered electronically using Research Electronic Data Capture [14,15]. Interviews were audio recorded, and transcripts were subsequently generated using an automated audio transcription service and were reviewed for accuracy prior to analysis (Otter.ai).

Data Collection Instruments

Via the preinterview survey instrument, the participants provided us with information regarding the duration of the HPE program they completed, time since completion of the program, degree granted by the program, granting institution, the format of educational content delivery, and the tuition funding sources. Subsequent interviews used a semistructured interview guide, which remained unchanged (Multimedia Appendix 1). The interview guide consists of 7 open-ended questions accompanied by 8 optional probing questions and prompts to increase the amount of information provided by each participant for a given question (Multimedia Appendix 1). Two members of the study team (JB and TC) with expertise in qualitative methodology designed the interview guide. The questions addressed motivation, timing (stage of career), perceptions of the educational program, perception of barriers to completing the program, and summative or wholistic opinions of the program. To maintain consistency of the interview technique, the same study investigator (AA) conducted all interviews.

Data Analysis

Interview transcripts were analyzed using an iterative, grounded theory approach to identify common themes [16,17]. Each theme was defined using a representative quotation from an interview. In cases where investigators coding the transcripts disagreed about the categorization of themes, a third investigator served as arbitrator. Because we were more interested in the relative importance and connection of ideas compared to the frequency of ideas, we performed this inductive thematic analysis, not a content analysis [13]. After the analysis, we presented the results to 2 of the participants to member check, which provided confirmation that the thematic analysis aligned with the message they sought to impart [13].

Results

Participants

All anesthesiologists approached for the study agreed to participate, for a total of 7 physician participants representing 6 different academic medical centers and representing the ranks of assistant and associate professor (Table 2). By self-identified gender, 4 (57%) of the participants were female and 3 (43%) were male. In total, participant degrees included 4 different HPE degrees granted by 6 different academic institutions; all participants completed a master’s degree. Moreover, 2/7 (29%) participants reported completing HPE training prior to attending medical school, and the remaining after their anesthesiology graduate medical education training. The 2 participants who completed degrees prior to medical school were self-funded for tuition and expenses, and department resources funded tuition of the 5 who earned HPE degrees after joining a faculty as academic physicians. The most recent HPE graduate completed their program 5 years ago; the most distant graduate was 25 years ago. For participants who completed their HPE degrees after completing residency or fellowship training, the range of completion date was 5-9 years ago. The median time since the receipt of HPE among the 7 participants was 6 years. Time to programmatic completion ranged from 1 to 6 years, with a median of 2 years. HPE instructional formats included in-person, virtual, and hybrid. The interviews lasted between 18 and 28 minutes. After the 7 interviews with iterative analyses were conducted, little additional information emerged, and it became apparent that thematic saturation had occurred.

Table 2. Summary of participant demographics and experiences.

Variables	Values
Time since degree completion (years), mean (range)	10.4 (5-25)
Duration of degree program (years), mean (range)	2.71 (1-6)
Degrees completed by interviewees, n (%)	
Master of Education	4 (57)
Master of Education in the Health Professions	1 (14)
Master of Science in Health Professions Education	1 (14)
Master of Academic Medicine	1 (14)
Degree-granting institutions, n (%)	
Harvard	2 (29)
Johns Hopkins	1 (14)
Massachusetts General Hospital	1 (14)
University of Cincinnati	1 (14)
University of Houston	1 (14)
University of Southern California	1 (14)
Self-identified gender of interviewees, n (%)	
Female	4 (57)
Male	3 (43)
National regions represented by interviewees, n (%)	
West Coast	2 (29)
Northeast	3 (43)
South	1 (14)
Upper Midwest	1 (14)
Tuition funding source, n (%)	
Departmental	5 (71)
Self-funded	2 (29)

Thematic Results

Our analysis of interview transcripts identified the following 6 core themes: outcomes, extrinsic motivators, intrinsic motivators, investment, experience, and recommendations (Table 3).

Table 3. Representative statements from the thematic analysis.

Code	Definition	Representative statements
Outcomes	Intended and unintended impacts of the degree on one’s career, including how they have used the degree and practical day-to-day application of skills or knowledge	• “It’s definitely a factor to get you noticed by people like chairs when they have educational leadership roles to fill.” • “I think it has been a good opportunity for me to… further push those key projects…[and] get a better understanding of where the problems are in care delivery within my department.” • “I think it influenced much of what I did even beyond education, when I look back, what really happened in the years just after, it just gave me that desire to know my own self, to just go for it. And that was really great to develop more confidence in your own abilities outside of what you do.” • “I mean, like the job offers I’m getting, it’s insane.”
Extrinsic motivators	Reasons for an individual seeking a degree based on attaining a known, external reward	• “Pretty much… [the chair] told me I had to do it to become faculty.” • “I would say that [for] the department and my division, definitely, it was an expectation that I would pursue the degree.” • “… my career goals were to kind of move up [to] med ed administration and to publish in medical education.” • “My mentor… was a very key proponent in me getting my masters because… to continue to move up that that would be a skill set and a degree that would look good from an experience standpoint.”
Intrinsic motivators	Reasons for an individual seeking a degree for its own sake without an external reward, including emotions, values, and goals	• “I was just frustrated with myself. And I felt like I just needed a formalized process and I needed everything at once and I was tired of trying to find it on my own.” • “I felt like I needed to know the language and I needed to know the theory behind why things are done the way they're done in medical education. And so that prompted me to get my masters.” • “I really wanted advanced training and knowledge in education in general, which I thought would be helpful, just to understand more what's going on” • “I started to really become interested in studying educational processes, and team dynamics even, and the ways we think and how it influences the way we act and just everything like that.”
Investment	Positive and negative aspects of obtaining an HPE[a] degree, including personal or financial sacrifices, opportunity costs, and time commitment	• “The biggest stressor was that I had to negotiate with my family because of time.” • “I didn't jump into the program my first year as an attending even though I was advised to, because I felt like I really needed to lay my ground as a clinician. … We work a lot of days in a row. And that makes doing an online curriculum while you're a full-time employee very difficult…” • “When I enrolled in the program, I had the added pressure to really get through it as fast as possible… because there was this tension with my family, basically.” • “First of all, it's a time commitment. …If you just stay in your clinical practice, right, and you try to do things within the division or department, it's already very busy.”
Experience	Overall perspectives about the degree program, including opinions about the process of obtaining the degree (ie, satisfaction with the content covered, mode or format of delivery, and suggestions for improvement)	• “It reviews a lot of the scientific methodology that we all appreciate even in other aspects of research. There's an emphasis on leadership, which I really appreciated. I especially appreciated that understanding of ourselves. There was an emphasis on understanding your MBTI scores and what that meant, which really gets into where you understand your strengths, and what works well.” • “But what I wish the program did was potentially focus less on individualized projects and potentially allow more collaboration and group projects for your Capstone… It would be really interesting to use the program more to develop interprofessional projects than having everybody do one individual project.” • “I think what I really would have loved is if there was somebody in there who could help you either write a case report, you know, or help you with the research part as you're doing it, or help you write a grant.”

Code	Definition	Representative statements
Recommendations	Advice that the participant would offer to someone interested in pursuing an HPE degree regarding timing, factors to consider, and aspects of a program to look for or avoid	• "If you think, look, I love to teach… you don't need a master's degree to be a teacher of residents, right, anybody in an academic center is going to teach residents. But if you think you want to be involved in residency leadership or medical school leadership, if you see yourself as being a program director one day or you know, dean for curriculum of a medical school, that kind of thing, then I think it is a good step because as I said before, I think it will get you noticed when those kinds of opportunities come up." • "I think right out of training, you don't necessarily know which person you are. What I usually advise is to do some workshops, figure out if you just want to become a really good teacher. I think you don't need a master's to do that." • "I think if you want to study teaching and you want to have a foundation in adult learning theory and you want to be able to become an administrator or become a researcher in med[ical] ed[ucation] that I would advise the [HPE degree]." • "If you're going to use this degree, you're pretty much marrying yourself to academics. But then I would also say that I think that there's a lot of opportunities for innovation, and a lot of interesting ways to use the master of education. And I would also say that I would sort of make sure that I had an academic or administrative niche that, you know, you can really start applying the coursework early on. So that you know, like, you can sort of build your academic portfolio while you're working on the degree."

[a]HPE: health professions education.

Outcomes

The participants consistently highlighted intended and unintended consequences from formal HPE training. Among the intended outcomes, the participants noted that the training prepared them for educational leadership roles within their departments and enhanced recognition from other departmental leaders. They also reported being better equipped to augment their daily academic activities, including learner assessment, clinical teaching improvement, and production of scholarship. Unintended but positive outcomes included earlier promotion and offers from other institutions for specific education appointments. The participants also noted a theme of self-exploration and discovery from their HPE experiences.

Extrinsic Motivators

The participants highlighted 3 recurring extrinsic motivating factors for pursuing formal HPE training, which were as follows: expectations from departmental leadership, availability of financial support, and opportunities for career advancement. Moreover, 4/7 (57%) participants noted that the completion of formal HPE programs allowed them to fulfill departmental requirements for promotion and to work with specific mentors in a structured way. Those who completed HPE programs after beginning their careers noted that financial tuition support served as a positive motivator.

Intrinsic Motivators

Two common intrinsic motivators for pursuing formal HPE training emerged, which were (1) a passion for education and teaching and (2) personal insight about a lack of knowledge of medical education theory. The participants sought to better understand the latest methodologies of medical education and noted feeling this was needed for career advancement along a medical education path; 1 (14%) participant said, "I needed to know the language and…the theory behind why things are done the way they're done in medical education." Moreover, 4 (57%) participants stated that their intrinsic motivation to pursue HPE training increased after becoming faculty (as gaps in their knowledge of education theory became clearer) compared to their motivation prior to becoming faculty.

Experience

The interviewees highlighted 3 experiential components of the HPE programs, which were structure, coursework, and research. Regarding structure, the participants noted that those entering HPE training programs should reflect on their own personal learning style to direct whether in-person, virtual, or hybrid learning best suits them. Programmatic flexibility was also important to the participants; they noted that given the complex schedules, asynchronous coursework helped them integrate varying clinical and administrative duties. The 5 (71%) participants who completed HPE programs after starting their careers all continued to work either full- or part-time while obtaining the degrees. Some commented that a more structured format was important to keep them focused, whereas others enjoyed the flexibility of completing coursework at their own pace.

Regarding coursework, the participants noted that the content complemented their personal and academic interests and enhanced their leadership abilities; in particular, content on leadership and professional development was highly regarded. Interdisciplinary programs that included professionals from other specialties, such as nursing, were viewed positively. The participants voiced some frustration around mandatory coursework that they perceived to be irrelevant to their goals, and regarding coursework that required a significant time investment. Additionally, 5 (71%) participants reported that they expected more statistical training, and at times, lack of statistical expertise may have hindered the completion of research projects. Nonetheless, the participants did endorse that they gained a better understanding of medical education research methodologies through the programs.

Investment

Common among interviews was the invested time and opportunity cost of HPE training programs. The participants specifically highlighted the balance between family, work, and programmatic demands as potential barriers when choosing to pursue an HPE program; 5 (71%) participants reported weighing the trade-off between establishing themselves as clinicians in their departments and the time necessary to dedicate to the educational program. Several participants (4/5, 80%) chose to delay starting HPE programs for several years after becoming faculty in order to establish clinical skills first. The participants noted that these programs led to an increase in “off hour” work and required a significant investment of time, often more than anticipated. Those participants (5/5, 100%) who pursued degrees after joining an anesthesiology department all received financial support for tuition.

Recommendations

The participants offered several key recommendations for colleagues considering pursuing formal HPE training. The first focused on program structure, with encouragements for colleagues to carefully evaluate the balance between virtual and in-person formats, the flexibility of course sequence, and time to degree completion; 3 (43%) recommended that individuals seek programs associated with their home institution for optimal flexibility. Regarding earning a formal HPE degree versus a teaching certificate, the participants recommended that individuals reflect on their motivation for pursuing additional training. Exemplary questions would probe, “Why pursue further education?” “What is the long-term goal?” “How will the training be utilized personally and professionally?” “Is there aspiration for an education leadership position that might be enhanced and more likely achieved by the completion of an HPE degree?”

Beyond career advancement, the participants highlighted that a genuine interest in education itself was important to ensure that the coursework is enjoyable. They suggested taking advantage of workshops or shorter programs prior to HPE enrollment to better understand whether pursuing further training might be enjoyable. Many (6/8, 75%) participants noted that having an HPE degree is not a prerequisite to being a talented educator, but the knowledge obtained from an HPE degree can be useful for learning methodologies for academic scholarship. There were different opinions regarding how to manage time commitments between work and HPE training. A common theme was the importance of practicing self-compassion given the time commitments and added stress of pursuing HPE training.

Discussion

Principal Findings

This qualitative study explored opinions of anesthesiologist educators who have completed formal HPE programs to provide a resource for physicians considering similar degrees. Recurring themes included outcomes on career, extrinsic and intrinsic motivators, personal investment, experience of obtaining formalized training, and recommendations. The participants noted that formal HPE training prepared them for leadership roles and to effectively engage in medical education scholarship. Further, they noted that the degree attracted the attention of other anesthesiology leaders seeking individuals for education leadership roles. They reported significant time investment, opportunity cost, and tension between personal and family well-being while training. The participants had insightful advice to guide anesthesiologists who are considering pursuing this degree, including stepwise introduction into the field of medical education, understanding motivations toward pursuing an advanced degree, and the recognition that being an educator and part of the medical education community does not necessarily require an advanced education degree.

This study provides future anesthesiology education leaders with guidance from several formal educators from around the United States. Faculty at departments where there are no anesthesiologists who have completed HPE training lack the opportunity to gain insights from colleagues when considering such training. Given the time and resources required to complete the programs, this information may help these individuals balance multiple personal and professional factors when deciding whether to pursue formalized educational training.

We suspect that the recent increased interest of anesthesiologists in pursuing HPE has emerged partly due to increasing ACGME expectations for faculty to not only administer training programs but also be educational innovators. Twenty-first century program directors must be capable of designing and implementing curricula while evaluating trainees using novel tools. Most medical school or anesthesiology residency programs do not equip physicians with expertise in learning theory, curriculum design, or evaluation and assessment [18]. HPE programs help faculty grow professionally, acquire knowledge, and join a community of like-minded individuals [19]. Further, these programs increase the quality of educational research by both subjective and objective measures [20]. Given the potential benefits to faculty and to anesthesiology trainees, the experiences gathered via the interviewees offer insights, caution, and encouragement for those considering HPE training. These themes may be particularly useful for those who lack institutional mentorship in 1 of the more than 160 anesthesiology residencies in the United States alone.

Limitations

There were several limitations to this study. Given that this was a homogenous group sample, there is potential for sampling bias. In order to minimize bias, we sought anesthesiologists from different subspecialties, genders, and races. Further bias mitigation strategies included using 2 investigators who independently analyzed the interview transcripts and member checking after the analysis [13]. When considering gathering quantitative survey data from a larger cohort of anesthesiologists compared to this smaller sample of in-depth interviews, we chose the latter in order to gain deeper insight into each individual participant’s experience. A national survey of physicians with HPE degrees will further enrich the themes presented and is an area of future research. Further, the participants represented large academic medical centers, and the applicability of their experiences to medium or smaller sized anesthesiology departments or private practice groups is unclear. Lastly, since all participants completed their HPE programs in the past, recall bias may play a role in the answers the participants provided throughout the interviews.

Conclusion

Our work offers insights from anesthesiology educators who have completed formal HPE training programs and can serve as a starting point for conversations for anesthesiologists who are considering pursuing similar programs. Future inquiries include larger survey data as well as longitudinal studies to observe career trajectories of individuals who pursued HPE degrees. The results support the benefits of HPE degrees for those who seek careers in medical education, especially those dedicated to pursuing careers in education leadership and scholarship.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study interview questions.
[DOCX File , 35 KB - mededu_v8i2e38050_app1.docx]

References

1. Caruso TJ, Lawrence K, Domingue BW, Qian J, Armstrong-Carter E. From Socrates to Virtual Reality: A Historical Review of Learning Theories and Their Influence on the Training of Anesthesiologists. JEPM 2020 Aug 24;22(2):1-10. [doi: 10.46374/volxxii-issue2-caruso]
2. McGoldrick KE. The history of professionalism in anesthesiology. AMA J Ethics 2015 Mar 01;17(3):258-264 [FREE Full text] [doi: 10.1001/journalofethics.2015.17.3.mhst1-1503] [Medline: 25813594]
3. Hartzell JD. Adult learning theory in medical education. Am J Med 2007 Nov;120(11):e11; author reply e13. [doi: 10.1016/j.amjmed.2006.10.024] [Medline: 17976395]
4. Ahmad M, Tariq R. History and Evolution of Anesthesia Education in United States. J Anesth Clin Res 2017;08(06):1-9. [doi: 10.4172/2155-6148.1000734]
5. Artino AR, Cervero RM, DeZee K, Holmboe E, Durning S. Graduate Programs in Health Professions Education: Preparing Academic Leaders for Future Challenges. J Grad Med Educ 2018 Apr;10(2):119-122 [FREE Full text] [doi: 10.4300/JGME-D-18-00082.1] [Medline: 29686748]
6. Common Program Requirements Section III B. Accreditation Council for Graduate Medical Education. 2011. URL: https://www.acgme.org/globalassets/dh_dutyhourscommonpr07012007.pdf [accessed 2022-06-15]
7. Tekian A. Doctoral programs in health professions education. Medical Teacher 2013 Oct 28;36(1):73-81. [doi: 10.3109/0142159x.2013.847913]
8. Tekian A, Artino AR. AM Last Page. Academic Medicine 2013;88(9):1399. [doi: 10.1097/acm.0b013e31829decf6]
9. Tekian A, Harris I. Preparing health professions education leaders worldwide: A description of masters-level programs. Medical Teacher 2012 Jan 17;34(1):52-58. [doi: 10.3109/0142159x.2011.599895]
10. Tekian A, Roberts T, Batty HP, Cook DA, Norcini J. Preparing leaders in health professions education. Medical Teacher 2013 Oct 25;36(3):269-271. [doi: 10.3109/0142159x.2013.849332]
11. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for Reporting Qualitative Research. Academic Medicine 2014;89(9):1245-1251. [doi: 10.1097/acm.0000000000000388]
12. Patton MQ. Qualitative Research and Evaluation Methods, 3rd Edition. Thousand Oaks, California, USA: Sage; 2002.
13. Jowsey T, Deng C, Weller J. General-purpose thematic analysis: a useful qualitative method for anaesthesia research. BJA Educ 2021 Dec;21(12):472-478. [doi: 10.1016/j.bjae.2021.07.006] [Medline: 34840819]
14. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
15. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
16. Watling CJ, Lingard L. Grounded theory in medical education research: AMEE Guide No. 70. Medical Teacher 2012 Aug 22;34(10):850-861. [doi: 10.3109/0142159x.2012.704439]
17. Bryant A, Charmaz K. Grounded theory research: methods and practices. In: The Sage Handbook of Grounded Theory. Thousand Oaks, California, USA: Sage; 2007:1-28.

18. Srinivasan M, Li S, Meyers F, Pratt DD, Collins JB, Braddock C, et al. "Teaching as a Competency": competencies for medical educators. Acad Med 2011 Oct;86(10):1211-1220. [doi: 10.1097/ACM.0b013e31822c5b9a] [Medline: 21869655]
19. Steinert Y, Naismith L, Mann K. Faculty development initiatives designed to promote leadership in medical education. A BEME systematic review: BEME Guide No. 19. Medical Teacher 2012 May 11;34(6):483-503. [doi: 10.3109/0142159x.2012.680937]
20. Ahmed R, Farooq A, Storie D, Hartling L, Oswald A. Building capacity for education research among clinical educators in the health professions: A BEME (Best Evidence Medical Education) Systematic Review of the outcomes of interventions: BEME Guide No. 34. Medical Teacher 2015 Nov 26;38(2):123-136. [doi: 10.3109/0142159x.2015.1112893]

Abbreviations

ACGME: Accreditation Council for Graduate Medical Education
HPE: health professions education

Edited by N Zary; submitted 16.03.22; peer-reviewed by A Szabova, M Barad; comments to author 28.04.22; revised version received 05.05.22; accepted 12.06.22; published 30.06.22.

Please cite as:
Aggarwal A, Hess O, Lockman JL, Smith L, Stevens M, Bruce J, Caruso T
Anesthesiologists With Advanced Degrees in Education: Qualitative Study of a Changing Paradigm
JMIR Med Educ 2022;8(2):e38050
URL: https://mededu.jmir.org/2022/2/e38050
doi:10.2196/38050
PMID:35771619

Original Paper

Readiness to Embrace Artificial Intelligence Among Medical Doctors and Students: Questionnaire-Based Study

Thomas Boillat[1], MSc, DPhil; Faisal A Nawaz[2]; Homero Rivas[1], MD, MBA

[1]Design Lab, College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates
[2]College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates

Corresponding Author:
Thomas Boillat, MSc, DPhil
Design Lab, College of Medicine
Mohammed Bin Rashid University of Medicine and Health Sciences
Healthcare City 14
Dubai
United Arab Emirates
Phone: 971 43838759
Email: Thomas.boillat@mbru.ac.ae

Abstract

Background: Similar to understanding how blood pressure is measured by a sphygmomanometer, physicians will soon have to understand how an artificial intelligence–based application has come to the conclusion that a patient has hypertension, diabetes, or cancer. Although there are an increasing number of use cases where artificial intelligence is or can be applied to improve medical outcomes, the extent to which medical doctors and students are ready to work and leverage this paradigm is unclear.

Objective: This research aims to capture medical students' and doctors' level of familiarity toward artificial intelligence in medicine as well as their challenges, barriers, and potential risks linked to the democratization of this new paradigm.

Methods: A web-based questionnaire comprising five dimensions—demographics, concepts and definitions, training and education, implementation, and risks—was systematically designed from a literature search. It was completed by 207 participants in total, of which 105 (50.7%) medical doctors and 102 (49.3%) medical students trained in all continents, with most of them in Europe, the Middle East, Asia, and North America.

Results: The results revealed no significant difference in the familiarity of artificial intelligence between medical doctors and students (P=.91), except that medical students perceived artificial intelligence in medicine to lead to higher risks for patients and the field of medicine in general (P<.001). We also identified a rather low level of familiarity with artificial intelligence (medical students=2.11/5; medical doctors=2.06/5) as well as a low attendance to education or training. Only 2.9% (3/105) of medical doctors attended a course on artificial intelligence within the previous year, compared with 9.8% (10/102) of medical students. The complexity of the field of medicine was considered one of the biggest challenges (medical doctors=3.5/5; medical students=3.8/5), whereas the reduction of physicians' skills was the most important risk (medical doctors=3.3; medical students=3.6; P=.03).

Conclusions: The question is not whether artificial intelligence will be used in medicine, but when it will become a standard practice for optimizing health care. The low level of familiarity with artificial intelligence identified in this study calls for the implementation of specific education and training in medical schools and hospitals to ensure that medical professionals can leverage this new paradigm and improve health outcomes.

(JMIR Med Educ 2022;8(2):e34973) doi:10.2196/34973

KEYWORDS

artificial intelligence in medicine; health care; questionnaire; medical doctors; medical students

Introduction

Background

Both public news and scientific articles widely argue that artificial intelligence will eventually disrupt medicine and the way physicians and medical professionals will be practicing in the future [1,2]. There has been impactful research that demonstrates the potential of artificial intelligence in medicine (AIM), for instance, to classify images such as x-rays [3]. Artificial intelligence is being evaluated not only for image processing and analysis but also for prognosis [4-6], treatment [7-9], and patient monitoring [10,11] among other uses. In addition, artificial intelligence algorithms have also been implemented in many consumer health products such as wearables and mobile devices [12]. From a medical perspective, it means that artificial intelligence–based algorithms are already giving recommendations to both patients and physicians and taking decisions on their behalf. It is therefore critical that physicians understand how this approach works and for software vendors and hospitals to identify what physician needs are to facilitate its implementation. So far, the evidence has not been very reassuring. When asked, "How familiar are you with artificial intelligence?" only 6% (out of a sample of 669 participants) of physicians and physicians in training in Seoul answered positively [13]. In another recent study, French medical experts reported that artificial intelligence is a "fuzzy notion" [2]. To evaluate the amount of empirical evidence collected regarding medical doctors' (MDs) and medical students' (MSs) level of understanding toward AIM, we conducted a systematic literature research. Of the 96 articles collected from Scopus, we identified only 9 (9%) studies (Multimedia Appendix 1 [2,11,13-19]) that surveyed medical professionals, the other ones being either out of scope or literature reviews. From existing empirical research, we identified the following. First, most studies surveyed medical professionals from either 1 university or 1 country. Second, one-third of the studies focused on the use of artificial intelligence in radiology. Third, none of the existing studies aimed to assess the level of understanding toward AIM.

Objectives

Owing to the importance of the topic, with this research, we intend to close this gap by surveying MDs and MSs from around the world on AIM topics that are the most discussed in the current literature. On the basis of the literature search, our questionnaire comprises the following sections: (1) the level of familiarity with AIM, (2) education and training related to AIM, (3) challenges and barriers linked to the implementation of artificial intelligence in clinical settings, and (4) risks linked to AIM.

Methods

Data were collected by means of a web-based questionnaire built following the guidelines developed by Burgess [20] as well as the 7-step process by Fowler [21].

Step 1—Define Your Research Aims

On the basis of the existing literature, we identified limited empirical data regarding MSs' and physicians' level of understanding toward AIM, their participation to AIM education, and challenges and barriers related to AIM implementation as well as potential risks linked to the democratization of AIM in clinical settings.

Step 2—Identify the Population and Sample

We were particularly interested in comparing MSs as a population (eg, aged 18-25 years) who has grown with technology and practicing physicians (eg, aged 30-60 years) who have clinical experience but might have been less exposed to technology. Mindful that the place of study and employment has a direct link with the knowledge and expertise that one acquires, we targeted the 6 continents to have a broad representation of the population under investigation. Participants were recruited by means of individual emails and posts from the authors' (TB, FAN, and HR) LinkedIn and Twitter profile feeds. This technique was used to avoid having participants from the same medical schools or hospitals, potentially having received the same education or training and thus creating biases in the data. To detect potential biases, we used the graduation year and name of the university to identify participants from the same cohort. When we found participants from the same school and graduation year, we randomly chose 5 of them. On the basis of the goal of this research, which investigates differences between two independent populations, MDs and MSs, our data sample should not be smaller than 176. This number was calculated based on a medium effect size (Cohen *d*) of 0.5, referring to limited existing empirical evidence [22]. Power was set to 0.95 with an allocation ratio of 1.

Step 3—Decide How to Collect Replies

Data were centralized in the university's platform after being collected by means of web-based questionnaire (Microsoft Forms).

Step 4—Design Your Questionnaire

From our review of previous work, we identified that none of the existing questionnaires were built following a systematic approach. In most studies, AIM factors were chosen based on research motivation. To systematically cover the most relevant AIM factors, we conducted a systematic literature search following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [23]. To this end, we performed a title and keyword search on the Scopus database using the keywords *"artificial intelligence"* AND *medicine* OR *"machine learning"* AND *medicine* as well as *"artificial intelligence"* AND *healthcare* OR *"machine learning"* AND *healthcare*. We did not perform an abstract search because of the abundance of unrelated articles. The search resulted in 837 papers. After being reviewed by 2 independent researchers for consistency, 9.3% (78/837) of the studies were retained for our qualitative analysis; of the 78 studies, only 9 (12%) used questionnaires. In total, 244 sections and 405 subsections were extracted. The latter were clustered by 2 independent researchers based on their similarity. From the 11 clusters, we created four different groups: (1) concepts and definitions, (2) training and education, (3) implementation, and (4) risks. A 5-point Likert scale was used for most questions, and drop-down menus were used for questions requiring categorical answers. More specifically, questions in the concepts and definitions factor displayed the following scales: (1) I have never heard of it, (2) I have heard of it a few times, (3) I understand it, (4) I can potentially explain it, and (5) I can confidently explain it. We defined the AIM's level of familiarity by calculating the mean across the different factors (questions 1.1-1.10 of Table 1). Questions in the training and education factor displayed the following scales: (1) strongly disagree, (2) disagree, (3) neutral, (4) agree, and (5) strongly agree. Questions in the implementation and risks factors displayed a scale similar to that in the training and education factors, but with an added option (0), I do not know. Finally, the clinical experience (only MDs) was derived from each age group as follows: 20 to 29 years=1, 30 to 39 years=2, 40 to 49 years=3, 50 to 59 years=4, and 60 to 69 years=5. The questionnaire can be accessed via Multimedia Appendix 2.

Table 1. Mean (SD) and *P* value for each factor.

Factors		MD[a], mean (SD)	MS[b], mean (SD)	*P* value
1. Familiarity with AIM[c]				
1.1	AI[d]	3.1 (1.0)	3.3 (1.0)	.24
1.2	ML[e]	2.7 (1.1)	2.8 (1.2)	.46
1.3	Supervised ML	2.0 (1.2)	2.1 (1.2)	.77
1.4	Unsupervised ML	1.9 (1.1)	2.0 (1.2)	.87
1.5	Deep learning	2.2 (1.1)	2.4 (1.2)	.18
1.6	Neural networks	2.2 (1.1)	2.5 (1.2)	.14
1.7	Fuzzy logic	1.6 (0.9)	1.5 (0.9)	.48
1.8	Support vector machine	1.5 (0.9)	1.4 (0.8)	.37
1.9	Overfitting or underfitting	1.6 (1.1)	1.5 (1.0)	.83
1.10	Feature selection	1.7 (1.1)	1.8 (1.1)	.64
2. Education and training				
2.1	Last time an AIM course was attended	1.4 (1.0)	1.98 (1.5)	.006[f]
2.2	Better understand the main concepts of artificial intelligence	4.0 (1.0)	4.3 (0.8)	.08
2.3	Explore the opportunities offered by artificial intelligence in general	4.1 (1.1)	4.2 (0.9)	.36
2.4	Explore the opportunities offered by AIM and your field	4.1 (1.1)	4.3 (0.8)	.14
2.5	Know more of existing commercial solutions	3.8 (1.1)	4.0 (0.9)	.23
2.6	Create my own artificial intelligence algorithm or applications	3.8 (1.1)	3.7 (1.1)	.40
3. Challenges to AIM's implementation				
3.1	Outcomes of artificial intelligence algorithms are difficult to trace or understand (the black box syndrome)	2.8 (1.7)	2.9 (1.7)	.67
3.2	The complexity of the field of medicine	3.5 (1.5)	3.8 (1.3)	.12
4.3	The availability of high-quality data samples	3.7 (1.4)	3.3 (1.7)	.75
3.4	The artificial intelligence's level of autonomy (what artificial intelligence should and should not do)	3.7 (1.4)	3.7 (1.4)	.96
3.5	The costs associated with the implementation of artificial intelligence	3.4 (1.6)	3.75 (1.4)	.16
3.6	Data privacy or confidentiality	3.7 (1.5)	3.7 (1.5)	.79
4. Barriers to AIM's implementation				
4.1	The availability of comparison studies	3.7 (1.4)	3.3 (1.7)	.13
4.2	The safe use of artificial intelligence	3.9 (1.3)	4.0 (1.4)	.93
4.3	Build trust between humans and artificial intelligence	3.7 (1.5)	3.7 (1.5)	.88
4.4	Availability of regulations and legislation	3.7 (1.6)	3.8 (1.6)	.75
4.5	The top management's level of understanding	3.8 (1.5)	3.6 (1.6)	.45
5. Risks linked to AIM's implementation				
5.1	Dehumanization of health care	3.3 (1.2)	3.5 (1.1)	.12
5.2	Reduction in physicians' skills (eg, physicians might execute fewer types of tasks)	3.3 (1.2)	3.6 (1.0)	.03[f]
5.3	Artificial intelligence will eventually harm patients	2.3 (0.9)	2.8 (1.0)	<.001[f]
5.4	Physicians may become redundant	2.6 (1.1)	3.0 (1.1)	.008[f]

[a]MD: medical doctor.

[b]MS: medical student.

[c]AIM: artificial intelligence in medicine.

[d]AI: artificial intelligence.

[e]ML: machine learning.

[f]Significant difference.

Step 5—Run a Pilot Survey

The questionnaire was completed by 15 MSs and 17 physicians from six different regions (Asia, Oceania, North America, the Middle East, Europe, and Eastern Europe). Cronbach α coefficient values of internal reliability reached .85, above the accepted .70 threshold [24]. When unpacked, the four quantitative parts (ie, concepts and definitions, training and education, implementation, and risks) respectively reached the following coefficient of internal reliability: (1) .91, (2) .94, (3) .81, and (4) .81. We used principal component analysis to examine the factor structure of the questionnaire. Kaiser-Meyer-Olkin (KMO) factor adequacy showed no correlation across the four factors (concepts and definitions, training and education, implementation, and risks). However, we did find the following correlations: (1) KMO=0.72, (2) KMO=0.75, (3) KMO=0.58, and (4) KMO=0.62. Following Kaiser and Rice [25], values above 0.5 are considered acceptable.

Step 6—Conduct the Main Survey

Participants used the link displayed in LinkedIn and Twitter posts to open the questionnaire. The landing page displayed the consent form including the objective and nature of the research, the risks and benefits, compensation and costs, confidentiality, participation (including rights to withdraw), contact information, and instruction. Only after choosing “I accept,” were the participants redirected to the questionnaire. The recruitment and questionnaire were open from August to December 2020.

Step 7—Data Analysis

We used descriptive statistics to describe and compare the demographics as well as the distributions of MDs and MSs within the four different factors (concepts and definitions, training and education, implementation, and risks). We then tested the descriptive statistics between MDs and MSs for significant differences using unpaired 2-tailed *t* tests (95% CI). We also built a linear regression model to explore factors associated with the risks brought by AIM (the risks factor). $P<.05$ is considered statistically significant. The outlined methods were carried out in accordance with relevant guidelines and regulations. We relied on the *required* functionality of our survey tool to ensure that participants did not miss any questions. As a result, no missing data were observed.

Ethics Approval

This study was approved by the Mohammed Rashid University of Medicine and Health Sciences’ Institutional Review Board Committee under MBRU-IRB-2020-024, and informed consent was obtained from all participants. The CHERRIES (Checklist for Reporting the Results of Internet E-Surveys) for the distributed survey is included as Multimedia Appendix 3 [26].

Results

Demographics

A total of 207 completed questionnaires were received. Among these 207 questionnaires, 105 (50.7%) were practicing physicians holding a medical degree and 102 (49.3%) were MSs. The repartition between men and women is somewhat even, as shown in Table 2. Although most of the participants were based in the Middle East (100/207, 48.3%), only 24.3% (51/207) of them were trained or are receiving their medical education in the Middle East (the list of institutions is available in Multimedia Appendix 4). Europe and Asia followed, with 24.8% (52/207) of the participants having received or are receiving their education in Europe and 15.8% (33/207) in Asia, whereas 17.4% (36/207) of the participants were based in Europe and 18.8% (39/207) in Asia. The distribution of participants can be found in Multimedia Appendix 4. The average time to complete the questionnaire was 12 minutes.

Table 2. Participants' demographics (N=207).

Demographics	MD[a], n (%)	MS[b], n (%)	Total, n (%)
Participants	105 (50.1)	102 (49.9)	207 (100.0)
Sex			
Men	62 (59.1)	43 (40.9)	105 (50.1)
Women	43 (42.1)	59 (57.9)	102 (49.9)
Age (years)			
<20	0 (0)	19 (18.6)	19 (9.2)
20-29	18 (17.1)	82 (80.4)	100 (48.3)
30-39	27 (25.7)	1 (0.9)	28 (13.3)
40-49	26 (24.8)	0 (0)	26 (12.6)
50-59	26 (24.8)	0 (0)	26 (12.6)
60-69	8 (7.6)	0 (0)	8 (3.9)
>70	0 (0)	0 (0)	0 (0)
Where the highest medical degree was obtained			
Asia	14 (13.3)	19 (18.6)	33 (15.9)
Africa	4 (3.8)	3 (2.9)	7 (3.4)
Central America	0 (0)	0 (0)	0 (0)
North America	11 (10.5)	10 (9.8)	21 (10.1)
South America	1 (0.9)	0 (0)	1 (0.5)
Europe	29 (27.6)	23 (22.6)	52 (25.1)
Eastern Europe	0 (0)	1 (0.9)	1 (0.5)
Middle East	9 (8.6)	42 (41.2)	51 (24.6)
Oceania	1 (0.9)	2 (1.9)	3 (1.4)
Where the participants are based			
Asia	15 (14.3)	24 (23.5)	39 (18.8)
Africa	1 (0.9)	4 (3.9)	5 (2.4)
Central America	2 (1.9)	0 (0)	2 (0.9)
North America	11 (10)	9 (8)	20 (9)
South America	0 (0)	0 (0)	0 (0)
Europe	15 (14.1)	21 (20.6)	36 (17.4)
Eastern Europe	0 (0)	1 (0.9)	1 (0.5)
Middle East	59 (56.2)	41 (40.2)	100 (48.3)
Oceania	2 (1.9)	2 (1.9)	4 (1.9)

[a]MD: medical doctor.

[b]MS: medical student.

Main Outcomes

As shown in Table 1, artificial intelligence (1.1) is the only concept that most participants understand with a mean of 3.27 (SD 1) for MSs and 3.11 (SD 1) for MDs. It is followed by machine learning (1.2), neural networks (1.6), and deep learning (1.5). Supervised and unsupervised machine learning (1.3 and 1.4), which are two concepts widely used in medicine, did not score very high.

The concept of overfitting and underfitting (1.9), which is one of the core principles in artificial intelligence, obtained among the lowest scores. In addition to questions 1.7, 1.8, and 1.9, MSs showed a better level of understanding than MDs as displayed in Figure 1. However, statistical comparisons between the 2 populations revealed no significant difference across the artificial intelligence concepts, as shown in Table 1.

We asked, "When was the last time you attended a course on AIM?" (2.1), a large majority of both MDs and MSs have never attended a course on AIM, whereas slightly more MSs have done so this year (ie, in 2020) or last year (Figure 2). Tests of statistical significance showed a difference between the 2 populations as shown in Table 1.

Figure 1. Familiarity with artificial intelligence in medicine (AIM)—comparison between medical doctor (MD) and medical student (MS; y-axis: means and SDs). ML: machine learning.

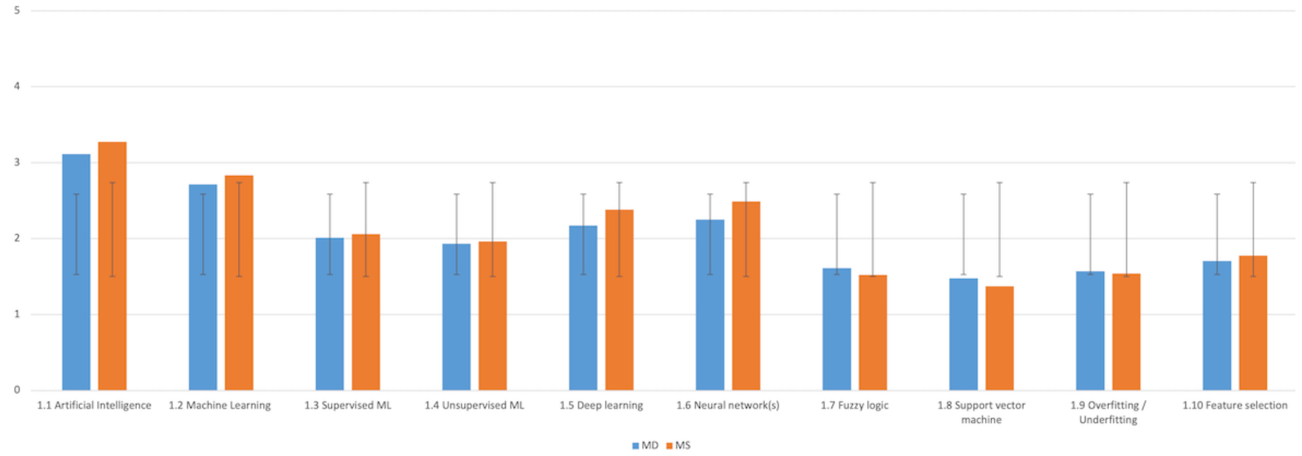
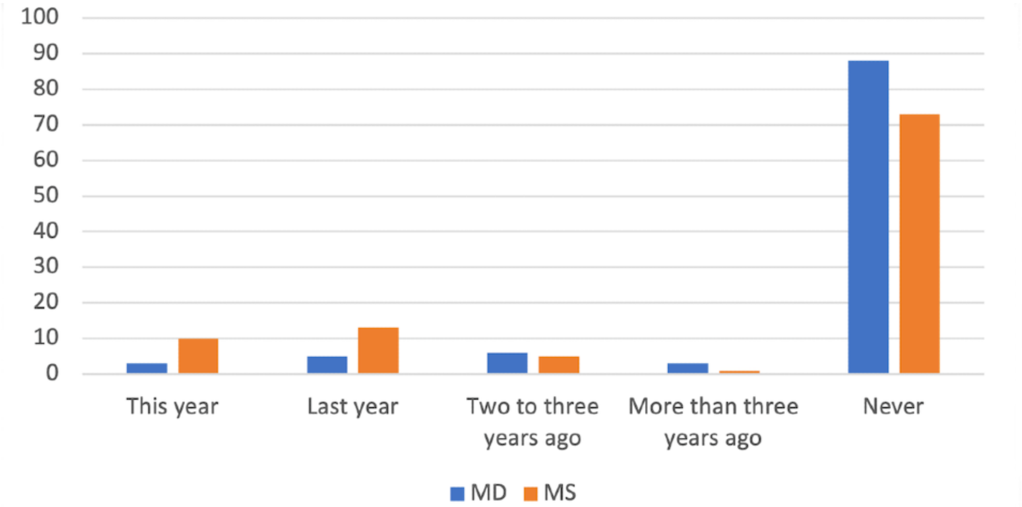


Figure 2. Last time that medical doctor (MD) and medical student (MS) attended a course on artificial intelligence in medicine (AIM; y-axis: percentages).



For both MDs and MSs, the priority is to further explore opportunities offered by artificial intelligence in their own field and in general and to better understand the main concept of artificial intelligence. Despite MDs having clinical expertise and, thus, a better idea of potential opportunity, there was no significant difference with MSs. However, the fact that MDs are more eager to learn how to create their own artificial intelligence algorithms or applications might confirm that they see more clinical potential than MSs, as displayed in Figure 3.

From an MD perspective, challenges linked to data privacy and confidentiality are the biggest challenges, followed by the availability of high-quality data samples and the artificial intelligence's level of autonomy, as shown in Figure 4. From an MS viewpoint, the complexity of the field of medicine is the biggest challenge, which could be explained by their limited expertise and clinical exposure. This is followed by the cost associated with the implementation of artificial intelligence as well as the artificial intelligence's level of autonomy and challenges related to data privacy and confidentiality. Challenges caused by the black box syndrome drew the least attention. There was no statistical difference between MDs and MSs across the different challenges.

In addition to challenges linked to the implementation of AIM, we also identified in the literature some barriers that can prevent the implementation of AIM, as displayed in Figure 5. From both the MD and MS perspectives, the safe use of artificial intelligence is the most important, followed by the availability of regulations and legislation as well as trust that must be built between human and artificial intelligence. The top management's level of understanding is the only one that MDs rated higher than MSs. For these factors, too, we did not find a significant difference between MDs and MSs.

Among the risks linked to the use of artificial intelligence in clinical settings, the potential reduction in physicians' skills was rated the highest by both MDs and MSs, as shown in Figure 6. With a score of 3.29 and 3.62, given by MDs and MSs, respectively, the test for statistical difference was significantly positive. The second highest score went to the risk linked to the dehumanization of health care with 3.27 for MDs and 3.52 for MSs. MSs are also more concerned than MDs that physicians may become redundant and that artificial intelligence will eventually harm patients, and both showed statistical differences between the 2 groups (*P*=.008 and *P*<.001, respectively).

The questionnaire ended with the following question: Can you imagine working with an artificial intelligence algorithm as a colleague? Most participants answered “yes” as shown in Figure 7.

We were interested to know more about the underlying reasons behind this choice and thus asked the participants to motivate their answer. Textbox 1 presents the extract of the collected responses.

We were interested in investigating the relationships between the level of familiarity with AIM, clinical experience, and the perception of risks. As shown in Table 3, there are significant negative correlations between the level of familiarity with AIM and the risk of dehumanization of health care, reduction in physicians’ skills, and risk that physicians may become redundant. In other words, the more MDs and MSs know about AIM, the less they perceive these factors as risks. No significant difference was identified between the level of AIM familiarity and the risk to eventually harm patients. Similarly, there are significant negative correlations between clinical experience and the risk that artificial intelligence will eventually harm patients and that physicians become redundant. No significant difference was identified with the risk that artificial intelligence will dehumanize health care or reduce physicians’ skills.

Figure 3. Reasons to attend a course on artificial intelligence in medicine (AIM)—comparison between medical doctor (MD) and medical student (MS; y-axis: means and SDs).

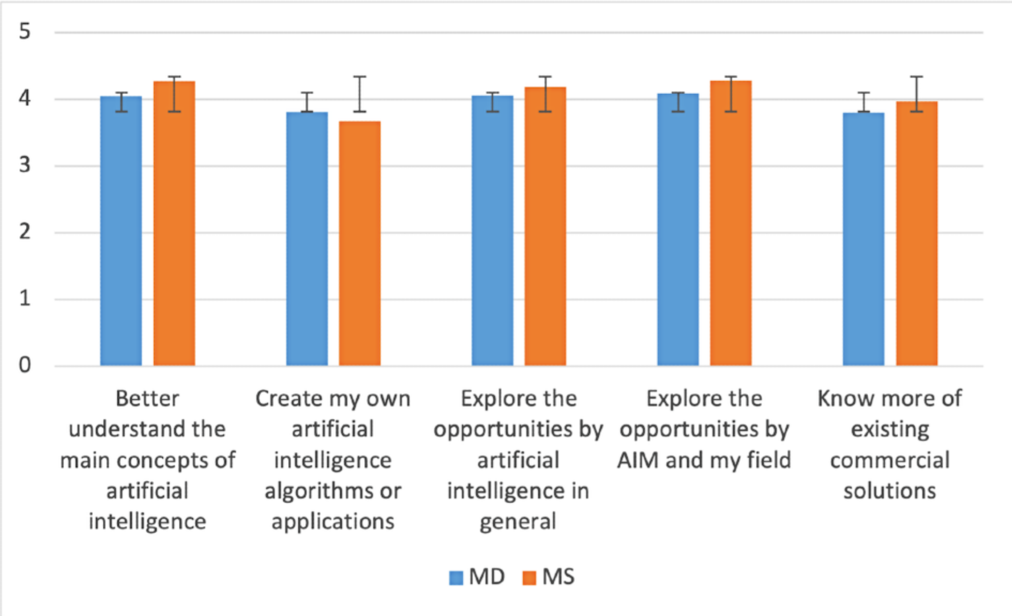


Figure 4. Challenges to artificial intelligence in medicine’s (AIM) implementation—comparison between medical doctor (MD) and medical student (MS; y-axis: means and SDs).

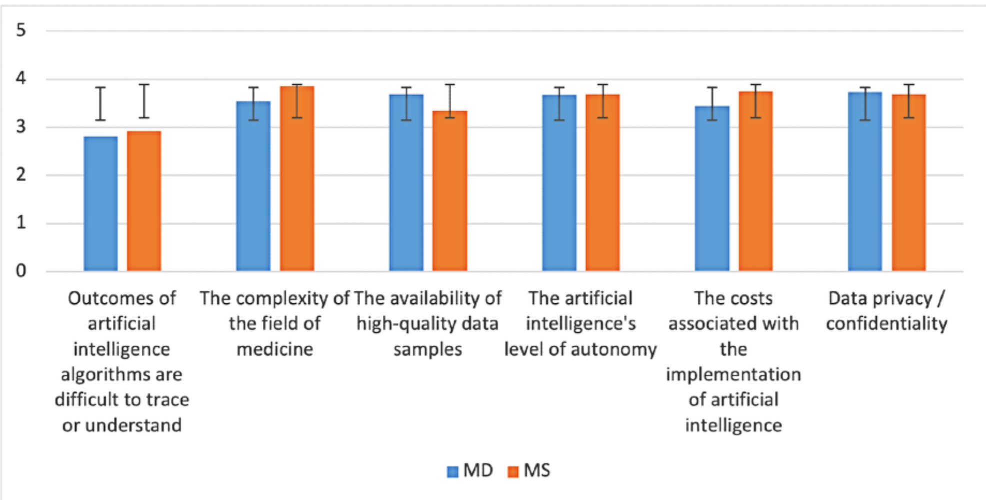


Figure 5. Barriers to artificial intelligence in medicine's (AIM) implementation—comparison between medical doctor (MD) and medical student (MS; y-axis: means and SDs).

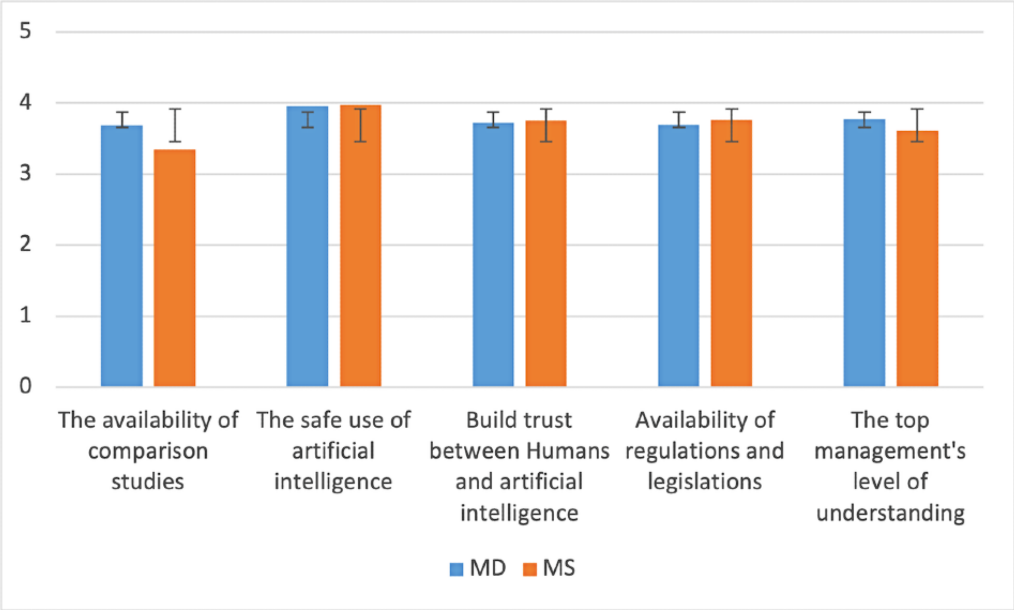


Figure 6. Risks linked to artificial intelligence in medicine's (AIM) implementation—comparison between medical doctor (MD) and medical student (MS; y-axis: means and SDs).

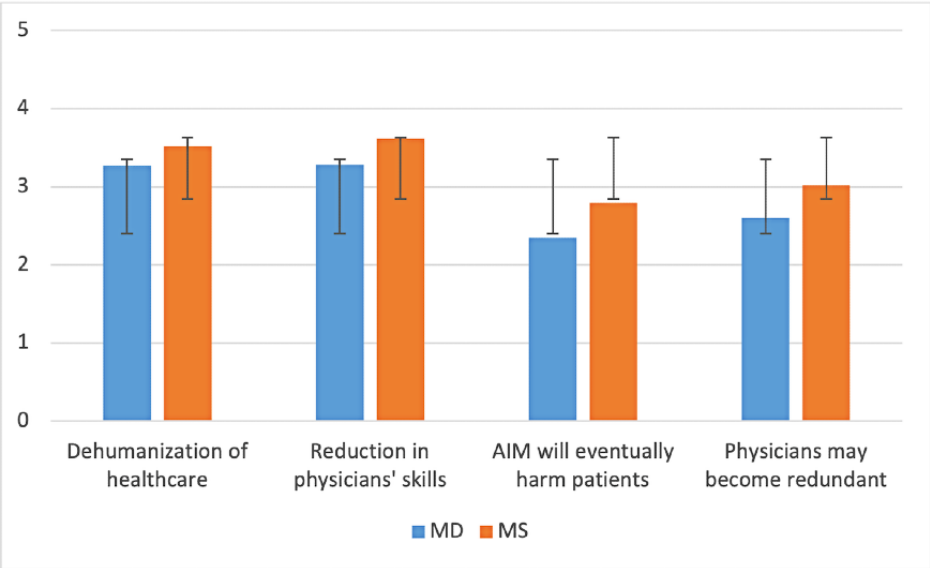
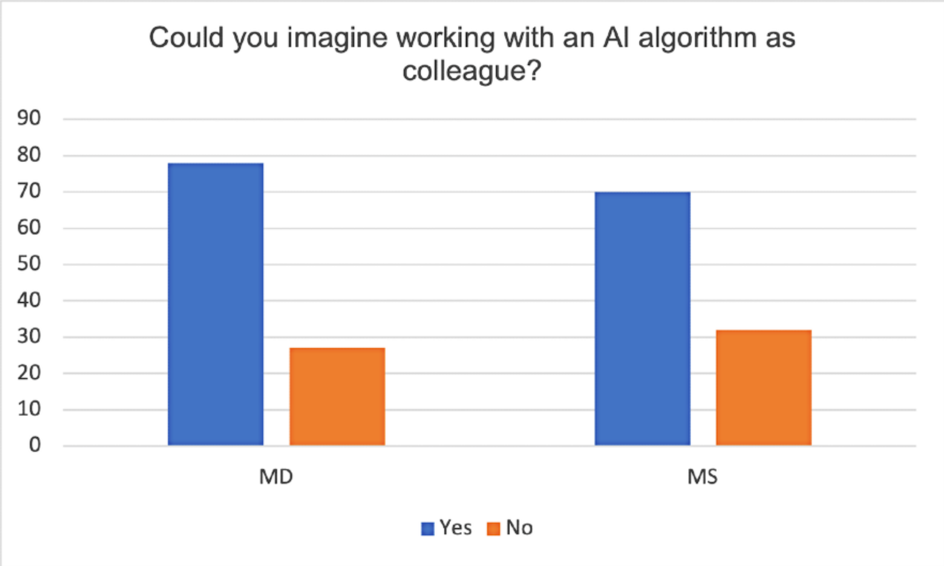


Figure 7. Working with an artificial intelligence (AI) algorithm—results (y-axis: percentages).



Textbox 1. Why could you (not) work with artificial intelligence as a colleague? Answers from participants.

Yes

- "That is the future - safer, secure, less emotionally driven, more reliable."
- "Ease the work, accurate diagnosis, improve patient care and reduce workload."
- "It would be very efficient and helpful as information will be processed and delivered instantly with less room for error."
- "Because an artificial intelligence will help in reducing the human errors such as near misses or misdiagnosis. It will learn the more it sees and will adapt to the patient presentation just as we medical students do."
- "The speed of development is exponential and the current status is quite impressive."

No

- "Physicians are being undermined and eventually replaced by mid-level providers and artificial intelligence."
- "Because although we already deal with 'algorithms' that have the potential to become artificial intelligence algorithms in our academic learning. I have not encountered many physicians who adopt that way of linear thinking in their practice. To them, intuition plays a bigger role in clinical judgment."
- "I think artificial intelligence should only be bossed around and not seen as a colleague who can think by himself because artificial intelligence cannot have moral or emotional values from itself but from a human boss who manages or controls it."
- "Artificial intelligence is OK for hypothesis generation, e.g., suggesting rare diagnoses which may be missed, but cannot replace the dynamic interaction with knowledgeable colleagues."
- "I don't think artificial intelligence will be able to communicate like my colleagues in my lifetime."

Table 3. Factors associated with AIM[a] risks (significance level *P*>.05).

Associated factors	Estimate (SE; SD)	*t* test (*df*)	*P* value
Dehumanization of health care			
Familiarity with AIM	−0.28393 (0.09757; 1.146)	−2.91 (205)	.004[b]
Clinical experience	−0.08585 (0.0608; 1.164)	−1.415 (205)	.16
Reduction in physicians' skills			
Familiarity with AIM	−0.27568 (0.09252; 1.087)	−2.98 (205)	.003[b]
Clinical experience	−0.10175 (0.05743; 1.102)	−1.772 (205)	.08
Artificial intelligence will eventually harm patients			
Familiarity with AIM	−0.0949 (0.08102; 0.952)	−1.171 (205)	.24
Clinical experience	−0.12819 (0.04897; 0.940)	−2.618 (205)	.009[b]
Physicians may become redundant			
Familiarity with AIM	−0.21163 (0.09575; 1.125)	−2.21 (205)	.28[b]
Clinical experience	−0.14515 (0.05846; 1.122)	−2.483 (205)	.01[b]

[a]AIM: artificial intelligence in medicine.

[b]Significant difference.

Discussion

Principal Findings

This research focused on assessing the level of understanding of AIM of MDs and MSs by means of a web-based questionnaire. It aims to complement the limited number of empirical studies on this key topic. When asked about artificial intelligence fundamentals, the participants provided somehow inconsistent answers. If most MDs and MSs understand artificial intelligence as a concept, it is unclear why they have only heard of overfitting and underfitting a few times, although these 2 concepts are key to understanding the outcomes of an artificial intelligence algorithm and their impact [27]. Similarly, the concepts of supervised and unsupervised algorithms did not reach a high level of familiarity for either MDs or MSs, whereas deep learning and neural networks, the 2 most used types of algorithms in supervised settings, received a higher score.

It was reassuring to find a strong positive correlation between deep learning (1.5) and overfitting or underfitting (1.9) as well as between neural networks (1.6) and overfitting or underfitting. This means that those who have a good level of understanding of deep learning and natural networks also have a good understanding of overfitting or underfitting. When analyzed at an aggregated level (questions 1.1–1.10), our results did not reveal any significant difference between MDs and MSs, which was unexpected because of the high level of curiosity of the younger population when it comes to technology and innovation. Globally, this low level of familiarity with artificial intelligence is not surprising when looking at the low number of MDs or MSs who had attended a course on artificial intelligence (Figure 2). Our analysis also showed that participants who attended a course on artificial intelligence have a statistically significant level of familiarity with artificial intelligence (*P*<.001). According to a recent study, a large majority of MSs argued that artificial intelligence should be part of medical training [17], although very few medical schools offer such programs [28].

When it came to the challenges linked to the implementation of AIM, we did not expect to observe a statistically significant difference between the 2 groups. We expected that clinical experience and an understanding of clinic organization would play a role in evaluating potential challenges. It can also be explained by the low level of artificial intelligence familiarity, which can limit MDs in understanding where AIM could bring new opportunities. It was also not expected that the *black box* syndrome would not be perceived as a bigger challenge (MDs=2.82; MSs=2.92). Such a lack of transparency is exactly what medicine does not want to see and has been identified as high risk by many scholars and practitioners [29-33]. These results also contradict the high importance that both MDs and MSs put in building trust between artificial intelligence and humans, which is very challenging owing to the lack of algorithms' transparency. Both the MDs and MSs also showed concerns with the safe use of artificial intelligence and the existence of regulations and legislation. Some efforts are being made with, for instance, the *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning-Based Software as a Medical Device*. The draft of this document, published by the US Food and Drug Administration, seeks feedback from experts [34]. It was also very interesting to discover that MSs are genuinely more risk averse than MDs. They fear that artificial intelligence might reduce physicians' skills, eventually harm patients, and make physicians redundant. These results can be partially explained by the correlation between risks and years of experience. As shown in Table 3, the more time spent in clinics, the lower the perceived risks caused by artificial intelligence. When analyzing the reasons why MDs and MSs could be willing to work with artificial intelligence as a colleague, it appears that the opportunities offered by artificial intelligence to improve patient care and reduce human errors are the most prevalent. Conversely, participants who said that they could not work with an artificial intelligence algorithm did not necessarily disagree with using artificial intelligence but rather with seeing artificial intelligence as a colleague, as they argued that technology can have neither the same emotion as humans nor the same way of thinking and interaction.

Limitations

Some limitations with our research and its data sample in particular should be considered. First, our recruitment technique limited the reach of the questionnaire to participants as part of the network of the authors. Consequently, some regions of the world are underrepresented. In addition, it did not allow us to systematically calculate the response rate. Second, although the sample size was statistically sufficient for our research goals, it did not allow us to further investigate the differences across variables such as regions, age groups, education (eg, undergraduate programs vs postgraduate programs), or medical specializations.

Comparison With Previous Work

This research differentiates itself from existing studies [8,17,35] through its approach and the diversity of its data sample. By building our questionnaire based on a literature search, we ensured that the most common AIM topics are included in our questionnaire. This approach is unique among similar studies that rather selected their AIM scope based on unknown criteria. As a result, our questionnaire is also more specific compared with existing research. For instance, in their study, Oh et al [13] asked, “Do you agree that you have good familiarity with artificial intelligence? [Strongly] agree, neutral, [strongly] disagree.” Instead, we decoupled the same question in 10 specific subtopics from machine learning to deep learning. In addition to a thinner abstraction level, it allowed us to identify some inconsistencies in some answers where some participants were supposedly able to confidently explain machine learning, but they had never heard of unsupervised algorithms, which is very unlikely. In addition, unlike most of existing work, we combined quantitative with qualitative data, which allowed us to know the *why*. When it comes to our data sample, its characteristics are also unique. Most specifically, our data sample is more diverse than those in existing research, with participants having studied (or studying) in 128 different universities across 6 continents. In contrast, in a study by Santos et al [17], the 263 answers were collected from 3 universities only. The likelihood that participants received the same education is rather high, bringing potential biases in the data. For these reasons, we argue that our data sample and this research provide a relevant representation of the population.

Practical Implications

Although more and more medical applications embed artificial intelligence–based algorithms or agents, it is key for software developers to consider the physicians’ low level of familiarity toward artificial intelligence. When a radiologist asks a colleague his or her opinion about a patient’s x-ray, for instance, it is assumed that both went to medical schools and are physicians and had gone through a specific radiology training, regardless of where they come from. However, when the colleague is an artificial intelligence algorithm, things change drastically. In order for physicians to leverage the use of artificial intelligence–based applications, we argue that software developers should consider the following elements:

- Provide general information on how the artificial intelligence–based algorithm or software was built. Some topics would include information about the process as well as the types of data used and the amount of data used during the training and testing phases. It will allow physicians to gain understanding and trust.
- Integrate different user (physician) profiles with a dynamic level of guidance, according to the level of familiarity toward artificial intelligence. A physician with a low level of familiarity will require more information about the process by which the software treats the data. In contrast, a physician who is familiar with the topic only requires key information such as the confidence level.
- Describe the path that has led to each outcome or decision along with the level of confidence. It will allow the physician to understand the reasoning and the extent to which the outcome can supports his or her decision.
- Let the physician take the final decision, although the software provides the impact of this decision from a medical perspective. The documentation of the decision will then be used to improve the algorithm’s accuracy.

Conclusions

On the basis of the number of current clinical trials leveraging artificial intelligence [36], the question is not whether artificial intelligence will be implemented in clinical settings but rather when it will become a standard in health care optimization. In the near future, practicing physicians will need to be equipped with the appropriate knowledge and skills to determine whether the artificial intelligence–based suggested diagnosis or treatment is appropriate. Thus, it is critical that physicians have a good understanding of the key concepts behind artificial intelligence. We believe that changes should first come from medical schools that should integrate AIM into their curriculum to both explain the origins and fundamentals of AIM and integrate AIM research throughout clinical topics from pathology to surgery, internal medicine, emergency medicine, and psychiatry, to name a few. By examining the individual components of AIM, our study informs existing research that highlights the needs to define what AIM content should be taught in undergraduate medical education [37]. This, in turn, requires university faculty to train and adapt their teaching material to this dynamic paradigm. By educating the physicians of tomorrow, they will act as drivers of change in their future placements.

At the same time, hospitals and clinics must emphasize on the importance of AIM and provide mandatory training for their medical professionals by means of continuing medical education or continuing professional development. To standardize and encourage both medical schools and hospitals to train their (future) physicians, governments can also play a key role by providing clear regulations, guidelines, and resources. Some countries such as the United Arab Emirates have already implemented national programs [38] to help all sectors integrate and regulate artificial intelligence. Therefore, we foresee future research focusing on assessing the outcomes of existing interventions (eg, lectures, modules, and training programs) in view of supporting medical schools, hospitals, and governments with the implementation of educational programs toward equipping medical professionals with relevant artificial intelligence skills.

Acknowledgments

The authors are very grateful to Ivan James Prithishkumar, Laila Zarnegar, Catherine Kellett, and Stefan Du Plessis, who helped distribute the questionnaire.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Existing questionnaire-based research.
[DOCX File , 26 KB - mededu_v8i2e34973_app1.docx]

Multimedia Appendix 2
Questionnaire.
[DOCX File , 24 KB - mededu_v8i2e34973_app2.docx]

Multimedia Appendix 3
CHERRIES (Checklist for Reporting the Results of Internet E-Surveys) checklist.
[DOCX File , 18 KB - mededu_v8i2e34973_app3.docx]

Multimedia Appendix 4
Institutions participating.
[DOCX File , 33 KB - mededu_v8i2e34973_app4.docx]

References

1. Walsh F. AI 'outperforms' doctors diagnosing breast cancer. BBC News. 2020. URL: https://www.bbc.com/news/health-50857759 [accessed 2022-03-14]
2. Laï MC, Brian M, Mamzer MF. Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. J Transl Med 2020 Jan 09;18(1):14 [FREE Full text] [doi: 10.1186/s12967-019-02204-y] [Medline: 31918710]
3. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv 2017 [FREE Full text]
4. Alloghani M, Al-Jumeily D, Aljaaf A, Khalaf M, Mustafina J, Tan S. The application of artificial intelligence technology in healthcare: a systematic review. In: Proceedings of the International Conference on Applied Computing to Support Industry: Innovation and Technology. 2019 Presented at: International Conference on Applied Computing to Support Industry: Innovation and Technology; September 15-16, 2019; Ramadi, Iraq p. 249-261. [doi: 10.1007/978-3-030-38752-5_20]
5. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019 Apr 04;380(14):1347-1358. [doi: 10.1056/nejmra1814259]
6. Ramesh A, Kambhampati C, Monson J, Drew P. Artificial intelligence in medicine. Ann R Coll Surg Engl 2004 Sep 01;86(5):334-338 [FREE Full text] [doi: 10.1308/147870804290] [Medline: 15333167]
7. Li D, Kulasegaram K, Hodges BD. Why we needn’t fear the machines: opportunities for medicine in a machine learning world. Acad Med 2019;94(5):623-625. [doi: 10.1097/acm.0000000000002661]
8. Xiang Y, Zhao L, Liu Z, Wu X, Chen J, Long E, et al. Implementation of artificial intelligence in medicine: status analysis and development suggestions. Artif Intell Med 2020 Jan;102:101780. [doi: 10.1016/j.artmed.2019.101780] [Medline: 31980086]
9. Yeasmin S. Benefits of artificial intelligence in medicine. In: Proceedings of the 2nd International Conference on Computer Applications & Information Security (ICCAIS). 2019 Presented at: 2nd International Conference on Computer Applications & Information Security (ICCAIS); May 1-3, 2019; Riyadh, Saudi Arabia p. 2-6. [doi: 10.1109/cais.2019.8769557]
10. Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism 2017 Apr;69S:36-40. [doi: 10.1016/j.metabol.2017.01.011] [Medline: 28126242]
11. Komorowski M, Celi LA. Will artificial intelligence contribute to overuse in healthcare?*. Critical Care Medicine 2017;45(5):912-913. [doi: 10.1097/ccm.0000000000002351]

12. Turakhia MP, Desai M, Hedlin H, Rajmane A, Talati N, Ferris T, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: the Apple Heart study. Am Heart J 2019 Jan;207:66-75 [FREE Full text] [doi: 10.1016/j.ahj.2018.09.002] [Medline: 30392584]
13. Oh S, Kim JH, Choi SW, Lee HJ, Hong J, Kwon SH. Physician confidence in artificial intelligence: an online mobile survey. J Med Internet Res 2019 Mar 25;21(3):e12422 [FREE Full text] [doi: 10.2196/12422] [Medline: 30907742]
14. Bisdas S, Topriceanu C, Zakrzewska Z, Irimia A, Shakallis L, Subhash J, et al. Artificial intelligence in medicine: a multinational multi-center survey on the medical and dental students' perception. Front Public Health 2021;9:795284 [FREE Full text] [doi: 10.3389/fpubh.2021.795284] [Medline: 35004598]
15. Galán GC, Portero FS. Medical students' perceptions of the impact of artificial intelligence in radiology. Radiologia (Engl Ed) 2021 Apr 29. [doi: 10.1016/j.rx.2021.03.006] [Medline: 33934846]
16. Castagno S, Khalifa M. Perceptions of artificial intelligence among healthcare staff: a qualitative survey study. Front Artif Intell 2020;3:578983 [FREE Full text] [doi: 10.3389/frai.2020.578983] [Medline: 33733219]
17. Santos DP, Giese D, Brodehl S, Chon SH, Staab W, Kleinert R, et al. Medical students' attitude towards artificial intelligence: a multicentre survey. Eur Radiol 2019 Apr;29(4):1640-1646. [doi: 10.1007/s00330-018-5601-1] [Medline: 29980928]
18. Kasetti P, Botchu R. The impact of artificial intelligence in radiology: as perceived by medical students. Russian Electr J Radiol 2020;10(4):179-185. [doi: 10.21569/2222-7415-2020- 10-4-179-185]
19. Park CJ, Yi PH, Siegel EL. Medical student perspectives on the impact of artificial intelligence on the practice of medicine. Curr Probl Diagn Radiol 2021;50(5):614-619. [doi: 10.1067/j.cpradiol.2020.06.011] [Medline: 32680632]
20. Burgess T. A general introduction to the design of questionnaires for survey research. Information Systems Services, University of Leeds, Leeds. 2001. URL: http://users.sussex.ac.uk/~grahamh/RM1web/Burgess_2001_Survey_design.pdf [accessed 2022-03-14]
21. Fowler Jr JF. Survey Research Methods. Thousand Oaks, CA: Sage Publications; 2013.
22. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Edition. New York: Routledge; 1988.
23. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
24. Nunnally J. Introduction to Psychological Measurement. New York: McGraw-Hill; 1978.
25. Kaiser H, Rice J. Little Jiffy, Mark IV. Educational and Psychological Measurement 1974;34(1):111-117. [doi: 10.1177/001316447403400115]
26. 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 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
27. Yang YJ, Bang CS. Application of artificial intelligence in gastroenterology. World J Gastroenterol 2019 Apr 14;25(14):1666-1683 [FREE Full text] [doi: 10.3748/wjg.v25.i14.1666] [Medline: 31011253]
28. Paranjape K, Schinkel M, Panday R, Car J, Nanayakkara P. Introducing artificial intelligence training in medical education. JMIR Med Educ 2019 Dec 03;5(2):e16048 [FREE Full text] [doi: 10.2196/16048] [Medline: 31793895]
29. Kulikowski C. Artificial intelligence in medicine: a personal retrospective on its emergence and early function. In: Proceedings of ACM conference on History of medical informatics. 1987 Presented at: ACM conference on History of medical informatics; November 5 - 6, 1987; Bethesda Maryland USA p. 199. [doi: 10.1145/41526.41544]
30. Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. J R Soc Med 2019 Jan 03;112(1):22-28 [FREE Full text] [doi: 10.1177/0141076818815510] [Medline: 30507284]
31. Shortliffe EH. Artificial intelligence in medicine: weighing the accomplishments, hype, and promise. Yearb Med Inform 2019 Aug 25;28(1):257-262 [FREE Full text] [doi: 10.1055/s-0039-1677891] [Medline: 31022745]
32. Sidey-Gibbons JA, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol 2019 Mar 19;19(1):64 [FREE Full text] [doi: 10.1186/s12874-019-0681-4] [Medline: 30890124]
33. Sun TQ, Medaglia R. Mapping the challenges of Artificial Intelligence in the public sector: evidence from public healthcare. Govern Inform Q 2019 Apr;36(2):368-383. [doi: 10.1016/j.giq.2018.09.008]
34. Diamond M. Proposed regulatory framework for modifications to artificial intelligence/machine learning-based software as a medical device (SAMD). U.S. Food and Drug Administration. 2020. URL: https://www.fda.gov/media/135713/download [accessed 2022-03-14]
35. Fan W, Liu J, Zhu S, Pardalos PM. Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). Ann Oper Res 2018 Mar 19;294(1-2):567-592. [doi: 10.1007/s10479-018-2818-y]
36. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med 2020 Sep 11;3(1):118 [FREE Full text] [doi: 10.1038/s41746-020-00324-0] [Medline: 32984550]
37. Lee J, Wu A, Li D, Kulasegaram K. Artificial intelligence in undergraduate medical education: a scoping review. Acad Med 2021 Nov 01;96(11S):62-70. [doi: 10.1097/ACM.0000000000004291] [Medline: 34348374]
38. National program for artificial intelligence: empowering the next wave of (coding) talents in the UAE. Artificial Intelligence Office, UAE. 2021. URL: https://ai.gov.ae/ [accessed 2021-11-15]

Abbreviations

AIM: artificial intelligence in medicine
KMO: Kaiser-Meyer-Olkin
MD: medical doctor
MS: medical student

Edited by N Zary; submitted 15.11.21; peer-reviewed by K Wac, N Thalange, MF Alwashmi; comments to author 17.01.22; revised version received 14.02.22; accepted 17.02.22; published 12.04.22.

Please cite as:
Boillat T, Nawaz FA, Rivas H
Readiness to Embrace Artificial Intelligence Among Medical Doctors and Students: Questionnaire-Based Study
JMIR Med Educ 2022;8(2):e34973
URL: https://mededu.jmir.org/2022/2/e34973
doi:10.2196/34973
PMID:35412463

Original Paper

Evaluation of Online Near-Peer Teaching for Penultimate-Year Objective Structured Clinical Examinations in the COVID-19 Era: Longitudinal Study

Savan Shah[1], BSc, MBChB

Barking Havering and Redbridge University Hospitals NHS Trust, Queen's Hospital, London, United Kingdom

Corresponding Author:
Savan Shah, BSc, MBChB
Barking Havering and Redbridge University Hospitals NHS Trust
Queen's Hospital
2 Parham Drive
London, IG2 6LZ
United Kingdom
Phone: 44 7415057781
Email: savan.shah2012@gmail.com

Abstract

Background: The benefits of near-peer learning are well established in several aspects of undergraduate medical education including preparing students for Objective Structured Clinical Examinations (OSCEs). The COVID-19 pandemic has resulted in a paradigm shift to predominantly online teaching.

Objective: This study aims to demonstrate the feasibility and benefits of an exclusively online near-peer OSCE teaching program in a time of significant face-to-face and senior-led teaching shortage.

Methods: A teaching program was delivered to penultimate-year students by final-year students at Manchester Medical School. Program development involved compiling a list of salient topics and seeking senior faculty approval. Teachers and students were recruited on Facebook. In total, 22 sessions and 42 talks were attended by 72 students and taught by 13 teachers over a 3-month period. Data collection involved anonymous weekly questionnaires and 2 separate anonymous student and teacher postcourse questionnaires including both quantitative and qualitative components.

Results: On a scale of 1-10, students rated the quality of the program highly (mean 9.30, SD 1.15) and felt the sessions were highly useful in guiding their revision (mean 8.95, SD 0.94). There was a significant increase in perceived confidence ratings after delivery of the program ($P<.001$). Teachers felt the program helped them better understand and retain the subject material taught (mean 9.36, SD 0.81) and develop skills to become effective clinical teachers (mean 9.27, SD 0.79).

Conclusions: This is the first study demonstrating the efficacy of a near-peer OSCE teaching program delivered exclusively online. This provides an exemplary framework for how similar programs should be encouraged given their efficacy and logistical viability in supplementing the undergraduate curriculum.

(JMIR Med Educ 2022;8(2):e37872) doi:10.2196/37872

KEYWORDS

near-peer teaching; peer-assisted learning; Objective Structured Clinical Examination; OSCE; online teaching; COVID-19; medical education; learning; medical school; near-peer teacher; NPT; near-peer learner; NPL

Introduction

"Near-peer teaching" refers to a way of teaching where the teacher is a trainee who is at least 1 year senior to the student and on the same level of the medical education spectrum [1]. Advocates fundamentally claim that its effectiveness stems from the social and cognitive congruence between near-peer learners (NPLs) and near-peer teachers (NPTs) as they are of similar ages, and therefore, share similar social roles (social congruence) and knowledge base (cognitive congruence) [2]. This cognitive congruence equips NPTs with unique insights and greater appreciation of the knowledge held by NPLs, subsequently enabling them to tailor the teaching to an appropriate level.

Near-peer teaching has grown in popularity in recent years within medical schools as a means of supplementing the formal curriculum. This paradigm shift was perhaps to be expected, given the sheer volume of literature advocating its benefits and effectiveness. For instance, Rodrigues et al [3] report on the efficacy of a near-peer teaching scheme delivered by junior doctors to medical students in a randomized controlled trial. The students ultimately made significantly fewer prescribing errors than the control group. In line with the social congruence theory, Leeper et al [4] argue that near-peer teaching creates a safer learning environment whereby students are more open to making mistakes and learning from them, while Topping [5] proposes that NPTs serve as influential role models in a phenomenon referred to as peer modeling. This helps students navigate the “hidden curriculum”—a set of unwritten rules students should follow to excel. The 19th century French moralist Joseph Joubert famously stated, “to teach is to learn twice,” eloquently conveying one of the key benefits of near-peer teaching. Elaborating on this, a randomized trial by Bargh and Schul [6] demonstrated that students who were asked to study a text with the task of teaching other students about it scored higher in an unexpected written test than those students from the control group who were asked to study for a test on their own. This suggests that teaching and its preceding preparation serves as a powerful drive for learning in a way that is distinct from preparing for an assessment. Others have taken a more pragmatic approach in accounting for the popularity of near-peer teaching, claiming it alleviates teaching pressures on faculty as medical school class sizes grow with a rising demand for doctors globally [7]. Therefore, it is reasonable to postulate that such near-peer programs will continue to grow in popularity.

Since the World Health Organization declared COVID-19 a global pandemic, widespread global lockdown restrictions resulted in a shift to online learning and cancellation of most face-to-face teaching and assessments [8]. Even prior to the pandemic, online learning had become an increasingly valued component of the undergraduate curriculum with multiple studies highlighting its merits. For instance, a recent meta-analysis found that online learning results in significantly better knowledge and skills outcomes based on posttest scores compared to traditional “offline” classroom teaching [9]. Another contemporaneous review of virtual teaching during the pandemic reports on the value of peer learning in this setting [10]. It highlights that peer learning reduces learner stress, helps develop resilience, and provides a platform for critical thinking and collaboration. Additional practical perceived benefits have been reported by medical students during the pandemic, including an absence of travel to attend sessions, flexibility to learn at their own pace, and opportunities to ask questions anonymously, subsequently encouraging wider engagement [11]. Nevertheless, numerous shortcomings of online learning have also been reported, including distraction from the physical environment, participants speaking over each other, and less of a perceived obligation to participate [12].

In this paper, I aim to outline how an online near-peer Objective Structured Clinical Examination (OSCE) teaching program can enhance the preparedness, knowledge, and skills of both the student and the teacher. In doing so, I aim to demonstrate that such endeavors are practical, reproducible, easy to implement, and valuable adjuncts to the medical undergraduate curriculum.

Methods

Ethics Approval

Consultation with the University of Manchester’s ethics decision tool [13] revealed that no formal external ethical approval was required for this study. All administered questionnaires outlined how the data gathered would be used and handled, followed by an opportunity to consent to the terms.

Program Development and Delivery

Manchester Medical Society page facilitated advertisement of the program to prospective NPTs and NPLs. NPTs expressing interest were added to a group chat, while NPLs followed a link to join a Facebook page. An instructional sheet was sent to all the NPTs. It incorporated reminders to include disclaimers, keep the presentations under 30 minutes, have them completed by a particular date before being sent for quality assurance by the head of clinical teaching, and include an OSCE practice scenario at the end of their presentations to contextualize the taught content. This ensured session standardization and OSCE relevance.

The fourth-year syllabus at Manchester Medical School is split into 2 distinct overarching themes: “families and children” (F&C) and “mind and movement” (M&M). The programs, therefore, consisted of 2 separate weekly, hour-long sessions on Zoom, 1 for each theme. Each session was split into 2 separate 30-minute lessons delivered by 2 NPTs. A total of 42 talks were delivered over 3 months, covering a range of topics within each theme (Table 1). Presentations were reviewed and returned each weekend by the head of clinical teaching so that timely revisions could be made. Formal approval of the teaching program involved drafting a proposal and liaising with multiple members of the senior faculty at Manchester Medical School. Examples of 2 presentations can be seen in Multimedia Appendices 1 and 2.

Table 1. List of topics covered over the course of the program.

Week number	F&C[a] topics	M&M[b] topics
Week 1	• Developmental milestones • Gynecology history taking and menorrhagia	• Delirium • Mental state examination
Week 2	• Breast medicine: triple assessment • Pediatric examinations	• Weakness • Loss of vision
Week 3	• Psychiatric history taking • Common ear presentations and examination	• Pediatric history taking • Dermatology: common presentations and their assessment
Week 4	• Oncological emergencies • Infectious disease: a brief overview	• Parkinson disease and examination • Osteoarthritis, rheumatoid arthritis, and hand examination
Week 5	• Cancer red flags • Pediatric respiratory presentations and clinical assessment	• Prescribing • Stroke medicine
Week 6	• Cervical health • HIV	• Ethics and law • Gout
Week 7	• Infertility • Skin cancers	• Vasculitis • Psychopharmacology
Week 8	• Pregnant abdomen examination and complications in pregnancy • Pediatric gastroenterology and abdominal examination	• Falls • Common fractures
Week 9	• Bleeding in early pregnancy • Ethics and law	• Alcohol dependence • Pediatric orthopedics and examination
Week 10	• Sexually transmitted infections and PV discharge • Nonaccidental injury and safeguarding	• Dementia • Cancelled
Week 11	• OSCE[c] example stations • Cancelled	• SBAR[d] handover • OSCE example stations

[a]F&C: families and children.

[b]M&M: mind and movement.

[c]OSCE: Objective Structured Clinical Examination.

[d]SBAR: Situation, Background, Assessment, Recommendation.

Feedback Questionnaires and Analysis

Anonymized weekly feedback questionnaires, 1 for the M&M and 1 for the F&C theme, were distributed using Google Forms following each session. This functioned as a means of tailoring future sessions toward the needs and wishes of the NPLs.

In addition, 2 separate postcourse questionnaires were distributed to NPLs and NPTs at the end of the program (Multimedia Appendices 3 and 4). The NPL postcourse questionnaire covered a range of questions relating to the perceived overall quality of the sessions, usefulness of the program, usefulness in guiding revision, confidence ratings prior to and after the program (out of 10), and the likelihood of organizing a similar teaching program. A free-text section allowed NPLs to relay any positive comments and advice for future improvement. The NPT postcourse evaluative questionnaire comprised of multiple statements with which the NPTs indicated their level of agreement on a 10-point Likert scale. They included topics relating to enjoyment of the program, its effect on teaching skills, and its benefits for teacher learning.

The questionnaire concluded with 2 free-text sections inquiring about the greatest perceived benefit of engaging in the program and providing space for further comments.

A paired 2-tailed *t* test analysis compared confidence ratings of NPLs prior to and after the course. The remaining items in the NPL and NPT postcourse questionnaires were analyzed by mean and SD.

Two NPTs conducted qualitative analysis of the free-text sections in the NPL and NPT postcourse questionnaires. This involved identifying and agreeing on common themes and categorizing each response appropriately.

Results

Program Development and Delivery

The 22 sessions (11 for F&C and M&M each) were attended by a total of 72 different NPLs and taught by 13 NPTs. The F&C weekly session attendance ranged from 8-26 NPLs (mean 14.64, SD 4.41). The M&M weekly session attendance ranged from 11-26 NPLs (mean 14.45, SD 4.63). Overall weekly attendance across the 2 themes ranged from 21-49 NPLs (mean 29.09, SD 8.64).

Feedback Questionnaires and Analysis

In total, 37 NPL postcourse evaluative questionnaires were completed equating to a 51.39% feedback response rate. A paired 2-tailed *t* test analysis demonstrated a statistically significant increase in mean confidence ratings following delivery of the program (t_{36}=–13.71, P<.001). Quality of teaching, usefulness of the OSCE scenarios, and usefulness in guiding revision were all rated highly (Table 2).

There was considerable variation in likelihood of organizing a similar program, with the majority taking a neutral stance (n=23), and the remaining indicating it is either somewhat likely or very likely (n=13), or it is somewhat unlikely (n=1). Of 37 students, 23 (62.16%) left comments in the free-text sections. Qualitative analysis revealed 8 broad themes with some overlap (Table 3).

Of 13 NPTs, 11 (84.62%) completed the NPT postcourse questionnaire. Respondents unanimously agreed that engaging with the program helped them better understand the material taught (mean 9.45, SD 0.86), facilitated long-term retention of the content (mean 9.36, SD 0.81), helped them develop their teaching skills (mean 9.00, SD 1), was rewarding and motivational in engaging in more teaching in the future (mean 9.27, SD 0.79), and was highly enjoyable (mean 9.30, SD 1.15) (Table 4).

Thematic analysis of the 2 free-text sections mostly reinforced findings from these rating questions. When asked what the NPTs felt the greatest benefit of the program was, 3 broad themes emerged, with some overlap (Table 5).

The "any positive comments/areas for improvement" section comprised of 3 comments expressing gratitude for organizing the program. A summary of the key findings from both questionnaires can be seen in Table 6.

Table 2. Findings from rating questions in the near-peer learner postcourse questionnaires.

Questions	Responses, mean (SD)
How confident did you feel about your OSCE[a] examinations prior to these teaching sessions?	4.51 (1.41)
How confident do you feel about your OSCE examinations now?	8.24 (0.93)
How would you rate the overall quality of the sessions?	9.30 (1.15)
How would you rate the usefulness of the OSCE practice scenarios at the end of each session in consolidating your learning?	8.92 (0.95)
How useful did you find the sessions in guiding your revision?	8.95 (0.94)

[a]OSCE: Objective Structured Clinical Examination.

Table 3. Qualitative analysis of free-text sections in near-peer learner postcourse questionnaires in response to the question "Any positive comments/areas for improvement?"

Theme	Examples of comments	Responses, n
Thanks, or expressions of gratitude	"Thank you so much to all the speakers and organizers for the effort through the semester!"	8
Praising the quality, organization, or structure of the teaching	"Really liked how the sessions were structured with a [sic] OSCE scenario at the end to consolidate the topic being taught. Really great stuff."	7
Helped in getting ahead with revision or placement	"Attending the sessions helped me get a head start on the placement I hadn't attended yet as everything was pitched at the correct level and at a good pace."	4
Advice to include more multiple-choice questions	"The only thing I could suggest to improve if [sic] more MCQ's[a] but I do understand that it was a more OSCE focused program which you guys hit the nail on the head. Thanks!"	3
Praising the top tips provided	"Really well organized and taught. You really provided some great OSCE tips throughout!"	3
Beneficial, given limited clinical exposure due to COVID-19	"The sessions have been particularly valuable with the reduced clinical opportunities we've had at placement with COVID."	2
Request for more sessions	"Would love some more sessions in the new year–found this really helpful in consolidating the TCD[b] cases."	2
Connection or technical issues	"Although the technical issues did sometimes interfere the sessions have been a great addition to the online cases."	2

[a]MCQ: multiple-choice question.

[b]TCD: Themed Case Discussion.

Table 4. Findings from rating questions in the near-peer teacher postcourse questionnaires.

Questions or statements	Responses, mean (SD)
Getting involved with the teaching has helped me better understand the subject material I taught.	9.45 (0.68)
Getting involved in the teaching has helped with my long-term retention of the subject material I taught.	9.36 (0.81)
I have found this teaching experience to be rewarding, and it has motivated me to get involved in more teaching in the future.	9.00 (1.00)
As a result of this experience, I have developed teaching skills that I will use in the future as a doctor.	9.27 (0.79)
On a scale of 1-10, how enjoyable did you find this teaching experience?	9.30 (1.15)

Table 5. Qualitative analysis of the free-text sections in the near-peer teacher postcourse questionnaires in response to the question "What do you think has been the greatest benefit of teaching in this program?"

Theme	Examples of comments	Responses, n
Beneficial to learning or guiding revision	"I think teaching is a very effective way of deepening your own understanding of a topic as well as identifying any gaps in your knowledge. If you are able to explain a complex topic as well as answer specific questions, this shows you have a very good level of knowledge. Therefore, this teaching program allows the tutors to expand on their understanding."	6
Enhancing or developing teaching skills	"Helpful feedback form students which I have reflected on to improve my teaching skills."	6
Increased self-confidence in teaching ability	"Over the course of the program, I've felt more confident and less nervous whilst teaching and ended up enjoying it a lot more than I would have expected!"	2

Table 6. Summary of the key findings from near-peer learner and near-peer teacher questionnaires.

Items in the questionnaires	Key findings
Near-peer learner postcourse questionnaire	
Change in perceived confidence to sit the OSCE[a] examination	• Precourse: mean 4.51, SD 1.41 • Postcourse: mean 8.24, SD 0.93 • *P*<.001
"Any positive comments/areas for improvement?"	• Thanks/expression of gratitude (n=8) • Praising the quality/organization/structure of teaching (n=7)
Near-peer teacher postcourse questionnaire	
Getting involved with the teaching has helped me better understand the subject material I taught	• Mean 9.45, SD 0.68
"What do you think has been the greatest benefit of teaching in this program?"	• Beneficial to learning/revision (n=6) • Enhancing/developing teaching skills (n=6)

[a]OSCE: Objective Structured Clinical Examination.

Discussion

Principal Findings

In line with the existing body of literature, this study clearly demonstrates the value of undergraduate near-peer programs for both the student and the teacher. Crucially, it demonstrates a significant increase in NPLs' perceived confidence to sit their OSCE examinations. Simulated OSCE scenarios and the setup of the sessions in helping guide revisions explain this finding. The study also demonstrates numerous benefits obtained by NPTs including enhanced comprehension and retention of the taught materials, development of teaching skills, and increased motivation in engaging with future teaching endeavors.

Evaluation of Findings and Comparison to Prior Work

A variety of reasons can account for the significant increase (*P*<.001) in NPLs' confidence in the OSCE performance. Evaluation of the free-text sections in the weekly questionnaires helped tailor future sessions suitably. For instance, the overwhelmingly positive reaction to the OSCE scenarios prompted organization of 2 sessions focusing exclusively on such scenarios for the final week of the program. The perceived benefit of these scenarios was mirrored in the NPL postcourse questionnaires' average rating of usefulness of the scenarios (mean 8.92, SD 0.95); it also supports the notion that simulation exercises in medical education lead to improvements in knowledge, confidence, and procedural performance upon retesting [14].

Rashid et al [15] adopted a similar methodology in 2011 to evaluate the success of their near-peer program for final year OSCE examinations and similarly found their program to have a positive influence on OSCE preparedness. However, unlike prior comparable studies, this study ensured that all taught materials were quality assured by a relevant senior faculty—the head of clinical teaching and examinations at Manchester Medical School. The exceptionally high ratings of the overall quality (mean 9.30, SD 1.15) are likely attributable to session standardization, facilitated by quality assurance and NPTs' instruction sheets. This highlights the importance of quality assurance and selection of appropriate individuals to ensure the relevance and accuracy of teaching. Qualitative analysis further consolidates this, with 7 students praising the overall quality, organization, or structure of the sessions. The anecdotal "top tips" provided throughout the sessions were also praised (n=3), further corroborating the notion of "cognitive congruence" and its intrinsic benefit in near-peer teaching [2].

NPTs also obtained multiple benefits, with engagement shown to enable better understanding (mean 9.45, SD 0.86) and long-term retention (mean 9.36, SD 0.81) of the taught material. These findings are supported by multiple prior studies [6,16] and analysis of the free-text sections, where a handful of NPTs (n=6) commented that teaching was beneficial to their learning and revision. Dandavino et al [17] postulate that other than learning, teaching provides the additional benefit of covert familiarization with teaching and learning principles, yielding a more effective learner and teacher. Additionally, the reported positive impact on teaching skills and motivation to teach as a doctor coincides with the United Kingdom's General Medical Council's "good medical practice" guidelines, which state that doctors "should be prepared to contribute to teaching and training doctors and students" [18]. The free-text sections complemented this notion, with NPTs outlining how teaching in an informal environment and responding to feedback helped enhance their teaching skills (n=6) and confidence in their teaching ability (n=2).

Limitations and Future Directions

First, the data collected and interpreted from the questionnaires in this study mostly deal with perceptions and lack objective measures. Future endeavors will look to collect data from tangible objective outcomes (ie, achieved OSCE marks of NPLs). Second, connectivity issues occasionally interfered with the delivery of teaching and our ability to share media and video clips—a drawback also identified by Mageswaran and Ismail [19], and 2 of the students in this study (Table 3). This can be avoided by asking all teachers to host sessions from a location where a stable connection can be ensured and to record sessions for students to play back in their own time if their connection fails during the live session. Third, although this study argues the case for near-peer learning, it fails to compare it to the alternative (ie, teaching by experienced senior faculty). Nevertheless, numerous prior studies have established that students taught by peers perform just as well as those taught by experienced teachers [3,20]. For instance, a randomized controlled trial by Tolsgaard et al [20] in 2007 found that students taught catheterization and intravenous access by student teachers performed just as well as those taught by associated professors in a postcourse assessment.

Conclusions

Numerous studies have established the value of the near-peer model in undergraduate medical education, including facilitation of OSCE preparedness [21,22]. However, this is the first study to evaluate an OSCE program delivered exclusively online. Given that the majority of curricular undergraduate clinical teaching is delivered by senior clinicians, there is scope for implementation of online OSCE-focused programs similar to the one described in this study. Such programs should involve careful review of weekly feedback to fine-tune sessions, practice scenarios to consolidate learning, and session standardization including quality assurance by relevant senior faculty as well as instructional sheets for teachers. Although online teaching is not a substitute to learning in the clinical setting, it is certainly a worthy adjunct that is economically and logistically viable and highly effective in preparing both students and teachers for clinical examinations and future practice. Peers embarking on similar endeavors will uncover novel ideas for delivering similar online programs in the future.

Acknowledgments

The author would like to thank Nick Smith for assisting with quality assurance of the teaching materials and all the student teachers that made the program possible.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Pediatric milestones presentation.
[PPTX File , 561 KB - mededu_v8i2e37872_app1.pptx]

Multimedia Appendix 2
Falls presentation.
[PPTX File , 1095 KB - mededu_v8i2e37872_app2.pptx]

Multimedia Appendix 3

Near-peer teacher postcourse questionnaire.

[XLSX File (Microsoft Excel File), 9 KB - mededu_v8i2e37872_app3.xlsx]

Multimedia Appendix 4

Near-peer learner postcourse questionnaire.

[XLSX File (Microsoft Excel File), 9 KB - mededu_v8i2e37872_app4.xlsx]

References

1. Bulte C, Betts A, Garner K, Durning S. Student teaching: views of student near-peer teachers and learners. Med Teach 2007 Sep 03;29(6):583-590. [doi: 10.1080/01421590701583824] [Medline: 17922356]
2. Lockspeiser TM, O'Sullivan P, Teherani A, Muller J. Understanding the experience of being taught by peers: the value of social and cognitive congruence. Adv Health Sci Educ Theory Pract 2008 Aug;13(3):361-372. [doi: 10.1007/s10459-006-9049-8] [Medline: 17124627]
3. Rodrigues J, Sengupta A, Mitchell A, Kane C, Kane C, Maxwell S, et al. The Southeast Scotland Foundation Doctor Teaching Programme--is "near-peer" teaching feasible, efficacious and sustainable on a regional scale? Med Teach 2009 Feb;31(2):e51-e57. [doi: 10.1080/01421590802520915] [Medline: 19330665]
4. Leeper H, Chang E, Cotter G, MacIntosh P, Scott F, Apantaku L, et al. A student-designed and student-led sexual-history-taking module for second-year medical students. Teach Learn Med 2007 Jun 19;19(3):293-301. [doi: 10.1080/10401330701366770] [Medline: 17594226]
5. Topping KJ. Peer Assisted Learning: A Practical Guide for Teachers. Cambridge, MA: Brookline books; 2001.
6. Bargh J, Schul Y. On the cognitive benefits of teaching. J Educ Psychol 1980;72(5):593-604 [FREE Full text] [doi: 10.1037/0022-0663.72.5.593]
7. Ten Cate O, Durning S. Peer teaching in medical education: twelve reasons to move from theory to practice. Med Teach 2007 Sep 03;29(6):591-599. [doi: 10.1080/01421590701606799] [Medline: 17922354]
8. Sandhu P, de Wolf M. The impact of COVID-19 on the undergraduate medical curriculum. Med Educ Online 2020 Dec 13;25(1):1764740 [FREE Full text] [doi: 10.1080/10872981.2020.1764740] [Medline: 32400298]
9. Pei L, Wu H. Does online learning work better than offline learning in undergraduate medical education? A systematic review and meta-analysis. Med Educ Online 2019 Dec;24(1):1666538 [FREE Full text] [doi: 10.1080/10872981.2019.1666538] [Medline: 31526248]
10. Wilcha R. Effectiveness of virtual medical teaching during the COVID-19 crisis: systematic review. JMIR Med Educ 2020 Nov 18;6(2):e20963 [FREE Full text] [doi: 10.2196/20963] [Medline: 33106227]
11. Dost S, Hossain A, Shehab M, Abdelwahed A, Al-Nusair L. Perceptions of medical students towards online teaching during the COVID-19 pandemic: a national cross-sectional survey of 2721 UK medical students. BMJ Open 2020 Nov 05;10(11):e042378 [FREE Full text] [doi: 10.1136/bmjopen-2020-042378] [Medline: 33154063]
12. Nozari A, Mukerji S, Lok L, Gu Q, Buhl L, Jain S, et al. Perception of web-based didactic activities during the covid-19 pandemic among anesthesia residents: pilot questionnaire study. JMIR Med Educ 2022 Mar 31;8(1):e31080 [FREE Full text] [doi: 10.2196/31080] [Medline: 35275840]
13. University ethical approval. The University of Manchester. URL: https://www.manchester.ac.uk/research/environment/governance/ethics/approval/ [accessed 2022-05-17]
14. Okuda Y, Bryson EO, DeMaria S, Jacobson L, Quinones J, Shen B, et al. The utility of simulation in medical education: what is the evidence? Mt Sinai J Med 2009 Aug;76(4):330-343. [doi: 10.1002/msj.20127] [Medline: 19642147]
15. Rashid MS, Sobowale O, Gore D. A near-peer teaching program designed, developed and delivered exclusively by recent medical graduates for final year medical students sitting the final objective structured clinical examination (OSCE). BMC Med Educ 2011 Mar 17;11:11 [FREE Full text] [doi: 10.1186/1472-6920-11-11] [Medline: 21414226]
16. Hall S, Harrison CH, Stephens J, Andrade MG, Seaby EG, Parton W, et al. The benefits of being a near-peer teacher. Clin Teach 2018 Oct 23;15(5):403-407. [doi: 10.1111/tct.12784] [Medline: 29573152]
17. Dandavino M, Snell L, Wiseman J. Why medical students should learn how to teach. Med Teach 2007 Sep 03;29(6):558-565. [doi: 10.1080/01421590701477449] [Medline: 17922358]
18. Good medical practice: the duties of a doctor registered with the General Medical Council. General Medical Council. 2013. URL: https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/good-medical-practice/duties-of-a-doctor [accessed 2022-05-17]
19. Mageswaran N, Ismail NAS. Preparing medical students for the final examinations during the COVID-19 crisis: a bumpy ride to the finishing line. JMIR Med Educ 2022 Feb 03;8(1):e31392 [FREE Full text] [doi: 10.2196/31392] [Medline: 35084354]
20. Tolsgaard MG, Gustafsson A, Rasmussen MB, Høiby P, Müller CG, Ringsted C. Student teachers can be as good as associate professors in teaching clinical skills. Med Teach 2007 Sep;29(6):553-557. [doi: 10.1080/01421590701682550] [Medline: 17978968]

21. Taylor D, Quick S. Students' perceptions of a near-peer Objective Structured Clinical Examination (OSCE) in medical imaging. Radiography (Lond) 2020 Feb;26(1):42-48. [doi: 10.1016/j.radi.2019.06.009] [Medline: 31902454]
22. Fulford L, Gunn V, Davies G, Evans C, Raza T, Vassallo M. Near peer integrated teaching for final year medical students. Perspect Med Educ 2016 Apr;5(2):129-132 [FREE Full text] [doi: 10.1007/s40037-016-0255-7] [Medline: 26908257]

Abbreviations

F&C: families and children
M&M: mind and movement
NPL: near-peer learner
NPT: near-peer teacher
OSCE: Objective Structured Clinical Examination

Edited by T Leung; submitted 09.03.22; peer-reviewed by A Sehmbi, B Mahamud; comments to author 04.05.22; revised version received 08.05.22; accepted 09.05.22; published 26.05.22.

Please cite as:
Shah S
Evaluation of Online Near-Peer Teaching for Penultimate-Year Objective Structured Clinical Examinations in the COVID-19 Era: Longitudinal Study
JMIR Med Educ 2022;8(2):e37872
URL: https://mededu.jmir.org/2022/2/e37872
doi:10.2196/37872
PMID:35617013

Original Paper

Harnessing Natural Language Processing to Support Decisions Around Workplace-Based Assessment: Machine Learning Study of Competency-Based Medical Education

Yusuf Yilmaz[1,2,3,4], MSc, PhD; Alma Jurado Nunez[5], MBBS, MSc; Ali Ariaeinejad[5], MSc; Mark Lee[1], BHSc, DipCLS; Jonathan Sherbino[1,6,7], MD, MEd; Teresa M Chan[1,3,6,7], MD, MHPE

[1]McMaster Education Research, Innovation, and Theory Program, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada
[2]Department of Medical Education, Ege University, Izmir, Turkey
[3]Program for Faculty Development, Office of Continuing Professional Development, McMaster University, Hamilton, ON, Canada
[4]Department of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada
[5]Department of Medicine and Masters in eHealth Program, McMaster University, Hamilton, ON, Canada
[6]Division of Emergency Medicine, Department of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada
[7]Division of Education and Innovation, Department of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

Corresponding Author:
Teresa M Chan, MD, MHPE
Division of Emergency Medicine, Department of Medicine
Faculty of Health Sciences
McMaster University
McMaster Clinics, Room 255
237 Barton St E
Hamilton, ON, L8L 2X2
Canada
Phone: 1 905 525 9140
Email: teresa.chan@medportal.ca

Abstract

Background: Residents receive a numeric performance rating (eg, 1-7 scoring scale) along with a narrative (ie, qualitative) feedback based on their performance in each workplace-based assessment (WBA). Aggregated qualitative data from WBA can be overwhelming to process and fairly adjudicate as part of a global decision about learner competence. Current approaches with qualitative data require a human rater to maintain attention and appropriately weigh various data inputs within the constraints of working memory before rendering a global judgment of performance.

Objective: This study explores natural language processing (NLP) and machine learning (ML) applications for identifying trainees at risk using a large WBA narrative comment data set associated with numerical ratings.

Methods: NLP was performed retrospectively on a complete data set of narrative comments (ie, text-based feedback to residents based on their performance on a task) derived from WBAs completed by faculty members from multiple hospitals associated with a single, large, residency program at McMaster University, Canada. Narrative comments were vectorized to quantitative ratings using the bag-of-n-grams technique with 3 input types: unigram, bigrams, and trigrams. Supervised ML models using linear regression were trained with the quantitative ratings, performed binary classification, and output a prediction of whether a resident fell into the category of at risk or not at risk. Sensitivity, specificity, and accuracy metrics are reported.

Results: The database comprised 7199 unique direct observation assessments, containing both narrative comments and a rating between 3 and 7 in imbalanced distribution (scores 3-5: 726 ratings; and scores 6-7: 4871 ratings). A total of 141 unique raters from 5 different hospitals and 45 unique residents participated over the course of 5 academic years. When comparing the 3 different input types for diagnosing if a trainee would be rated low (ie, 1-5) or high (ie, 6 or 7), our accuracy for trigrams was 87%, bigrams 86%, and unigrams 82%. We also found that all 3 input types had better prediction accuracy when using a bimodal cut (eg, lower or higher) compared with predicting performance along the full 7-point rating scale (50%-52%).

Conclusions: The ML models can accurately identify underperforming residents via narrative comments provided for WBAs. The words generated in WBAs can be a worthy data set to augment human decisions for educators tasked with processing large volumes of narrative assessments.

(JMIR Med Educ 2022;8(2):e30537) doi:10.2196/30537

KEYWORDS

natural language processing; machine learning algorithms; competency-based medical education; assessment; medical education; medical residents; machine learning; work performance; prediction models

Introduction

Workplace-based assessments (WBAs) are a key source of data about the competence of health professions learners [1-9]. Even in the busiest of environments, clinical teachers engage in direct observation, feedback, and assessment of trainees [10]. The data gathered in these busy environments often consist of both quantitative (numerical scores, typically associated with a scoring rubric, such as an entrustment scale) and qualitative (free-form narrative comments) data [8].

Throughout training, WBA programs can acquire hundreds of data points about a single trainee, which translate into hundreds of scores and thousands of words [3]. While quantitative scores can be aggregated and analyzed using several statistical methods [11,12], qualitative data often require an educator (eg, program director [PD], competence committee [CC] member, learner supervisor) to internally organize and make meaning of the data. With the rapid and expansive generation of narrative comments typical of a robust and active WBA system, the cognitive task load can overwhelm administrators. This becomes even more problematic when aggregated narrative data inform progress decisions for advancement in training.

Machine learning (ML) algorithms and natural language processing (NLP) have been demonstrated in other industries and in general health care to provide near real-time data analysis of large complex qualitative data sets. Adopting these techniques in medical education may thus be useful [11,13,14]. Early work in using ML algorithms (MLAs) to enhance human review of the quantitative learner assessment data generated by WBAs has been reported [15]. However, as the systematic review by Dias et al [14] pointed out, much of the work reported to date is around feasibility.

For machine-assisted qualitative data aggregation or analysis, the field is sparse. Some qualitative data sets have shown potential in assisting faculty in identifying those trainees who are at risk [16]. Early research suggests that keyword-specific algorithms may assist human review of qualitative data from WBAs [17]. A recent systematic review of NLP within medical education showed that the majority of the research to date examines clinical notes generated by the trainee, rather than assessment data generated about the trainee [13].

Narrative data have been shown to be both reliable and useful [18-20]. Not only are written comments deemed reliable for third-party readers to interpret the progression of trainees [18], but also the learners often cite that they value these commentaries above scores or numbers [20,21]. Qualitative assessments contain both clarifying and qualifying data about the numerical scores. To be clear, qualitative data can still be biased [11,22]. Assessors have multiple competing interests, clouding their ability to focus on the assessment task [10]. Cognitive load for raters embedded in the workplace may also lead to limitations in the types of ratings they generate [23,24]. Moreover, individual faculty members may have social biases that manifest in their comments [25,26].

However, the operational challenge unique to qualitative data compared with quantitative data is the aggregation of multiple narrative assessments into a global judgment. The difficulty of this task requires approaches akin to the ones used with inductive research methods—multiple reviewers, all providing their own interpretations of the data, and working together to generate a common interpretation. To navigate this challenge, many assessment systems use CCs, which harness the power of group dynamics to arrive at decisions about complex data sets [27-30]. These committees function similar to promotion and tenure committees or juries, and are often charged with aggregating, reviewing, and interpreting multiple sources of data to arrive at decisions about trainee performance [31-33]. While this type of approach is a systematic and robust method, it neglects the operational challenges of navigating the large volume of data created by programmatic assessments using only human-based systems.

There is potential for harnessing NLP and ML for the purposes of automating the first analysis of narrative data from WBAs to generate red flags of underperforming learners. This automated, early warning system could facilitate the more nuanced human review of the same data of the identified individual, allowing educators to focus their efforts and offload the overwhelming cognitive load to more efficient NLP and MLA processes. While this technology has potential to support a potential automated process and to create an early warning system, this paper acts as proof of concept and presents an approach as to how we can utilize NLP and ML to automatize the assessment process to offload a system for busy clinical teachers. To do that, the MLA should be trained with existing data so that future WBA data can be analyzed automatically. The purpose of this study is to explore NLP and MLA applications for identifying trainees at risk using a large WBA narrative comment data set associated with numerical ratings.

Methods

Study Context

This study retrospectively analyzed all WBA data from September 2012 to July 2018 of emergency medicine residents completed by faculty members from a large, multihospital residency training program at McMaster University, Canada. This clinical setting has between 6 and 10 trainees within a 5-year specialist training program for emergency medicine; therefore, at any given time there are roughly 40 trainees in the program, but only 6 new trainees enter the system each year. The health system is also nontrainee dependent (ie, staffed entirely by attending physicians, who function independently without the assistance of trainees or midlevel providers), which means there are more than double the amount of faculty members than there are trainee physicians affiliated with the program. As such, while trainees always have a supervising attending physician who is their teacher/assessor for the shift [10], not all shifts staffed by an attending physician will have a trainee.

The McMaster Modular Assessment Program (McMAP) is a programmatic assessment system with 76 WBA instruments grouped by junior, intermediate, and senior level, and mapped to the CanMEDS (The Canadian Medical Education Directives for Specialists) roles [3]. We descriptively explain those competencies in Table 1 and provide the number of assessments for each competency. However, we focus on each WBA in our analysis. One WBA is completed during each emergency department shift. Free-form narrative comments and a behaviorally anchored 7-point score are captured for each WBA. A full WBA example form is presented in Multimedia Appendix 1.

Analysis

A descriptive analysis of numerical scores and word frequencies was used to explore data and identify missing data. Demographics were analyzed using descriptive statistics in SPSS version 26 (IBM Inc.) [34]. Mean, SD, and frequencies were some of the descriptive statistics used. Missing data exploration was carried out on the data set to find ratings without comments and removed from the ML and NLP analyses. We used MATLAB R2019b and its libraries including "Statistics and Machine Learning Toolbox" and "Text Analytics Toolbox" to conduct analysis on MLA and NLP [35].

Two approaches were developed to stratify the data by quantitative rating. First, we used the 7-point scale ratings in the original form. To improve our ML models, we decided to collapse the ratings into a binary division. We chose this approach because many CCs promote a resident based on achieving a threshold (eg, a rating score 6.25 in our local setting for these WBAs, based on local standard-setting protocols) [3,36]. Thus, ratings from 1 to 5 were collapsed as a low score and ratings from 6 to 7 were collapsed as a high score.

Natural Language Processing and Machine Learning Analysis

NLP and a supervised ML analysis were run sequentially to identify patterns and results. Our model takes the input of a written feedback review for a resident's performance on a given day and tokenizes it to uni/bi/trigrams. Then, a linear regression ML model predicts the output for 2 different classifications: *at-risk* resident or *not-at-risk* resident.

Step 1: Preprocessing

Preprocessing steps are described in Textbox 1.

Textbox 1. Preprocessing steps for narrative comments.

1. Missing data: Assessment with no rating and comment was removed from machine learning algorithm analysis.
2. Tokenization: Each word was converted into a single-word format.
3. Part of speech: This function assigns a label to a word, such as verb, noun, proposition, number, punctuation.
4. Removal of stop words: To reduce noise in the data set, we removed stop words such as *a*, *and*, and *the*.
5. Lemmatization: Each word was converted into its root form (eg, *discharging* converted to *discharge*).
6. Removing punctuation: Punctuation was erased from the data set.
7. Removing infrequent words: Words with a frequency of 2 or fewer across the data set were removed.
8. Exclude empty assessment: Any blank narrative assessment fields were removed.

Step 2: Vectoring

After preprocessing, we used bag-of-words vectorizing. We generated unigrams (single, decontextualized words), bigrams (adjacent word couplets), and trigrams (adjacent word triplets) for input into the ML models.

Step 3: Machine Learning Analysis

Overview

Bag-of-words vectorizing for narrative data was used for the MLA stage. This technique takes each word within the comment and inputs each word into the MLA. Data were partitioned using a "holdout" technique with a 0.1 coefficient, meaning 10% of the data were randomly assigned with a nonstratified technique into a test data set, and the remaining data were selected for the training. ML analysis evaluated using tenfold cross-validation. More of the MLA explanation can be found in Multimedia Appendix 2.

Derivation Phase: Training of the Machine Learning Algorithm

The data were partitioned into a training and a testing data set. A supervised classification model, which used word frequency counts from the bag-of-words model as a predictor, was created and trained. The classification accuracy is the proportion of the labels that the model predicts correctly.

The supervised ML method used a linear learner model to train the data and to predict the test data set. Supervised learning can train a model when there are input data associated with a label as an outcome [37]. Our method is Error-Correcting Output Codes (ECOC), which uses K(K – 1)/2 binary support vector machine models, which means each classification group needs to be compared against the others. We did this by using the one-versus-one coding design, where K is the number of unique classification labels.

We trained the ECOC method composed of default classification models using the following parameters: *Learners* and *Linear.* The support vector machine used word frequency counts from the bag-of-words model as a predictor.

Validation Phase: Testing of the Machine Learning Algorithm

The last step was predicting the labels of the test data using the trained model and calculating the classification accuracy. Please see Multimedia Appendix 2 for further details on the training and testing phases.

Ethical Consideration

The Hamilton Integrated Research Ethics Board granted ethics exemption for this study under Tri-Council Policy Statement 2 (TCPS2) as this was deemed a quality improvement initiative.

Results

The initial database consisted of 7199 assessments, of which 5597 contained faculty comments for trainee performance. There were 141 unique raters from 5 different hospitals; 68% (n=96) of them were male. The database had a total of 45 unique residents; 56% (n=25) were male. Table 1 presents the overall demographics related to the assessments.

Rating distributions of the assessment ranged between 3 and 7. The frequencies for ratings 7, 6, 5, 4, and 3 were 2713/7199 (37.69%), 2158/7199 (29.98%), 635/7199 (8.82%), 79/7199 (1.10%), and 12/7199 (0.17%), respectively. We excluded a total of 1638 items because there were missing data (eg, the task rating did not have a meaningful comment associated or vice versa). The test set consisted of 484 high ratings and 72 low ratings.

In line with our previous work [15], we dichotomized our task rating scores: all scores of 5 and below were considered *at risk* and all scores of 6 and 7 were considered *not at risk.*

There were 94,016 words in the narrative comments. Assessments ranged from 1 to 155 words with a mean of 16.91 (SD 13.8). Figure 1 shows the frequencies of word counts across assessments by rating scale. Each rating scale is represented with a color in Figure 1 and seemed to have a similar trend in each rating scale regardless of the number of ratings.

Multimedia Appendix 3 depicts word clouds with size-based weightings of unigrams, bigrams, and trigrams grouped by higher-score (6 or 7) and lower-score (≤5) associated phrases. Bigram analysis showed more promising weighted phrases such as *good approach* or *excellent management.* The trigram analysis highlights key phrases that allow a human reader to begin contextualizing the assessment such as *rapport patient family* or *excellent communication skill.* There are more diverse phrases in the trigrams associated with lower scores rather than those associated with lower scores.

Table 2 presents the MLA results for accurately identifying residents who were deemed *at risk.* Accuracy was higher using a binary division of the rating scale labeling. Trigrams provided the most accurate results. The MLA demonstrated excellent sensitivity for identifying residents who achieved competence (6 or 7 on the rating scale). Unigrams had the highest sensitivity rates. The specificity was poor. More details on the analysis output (ie, confusion matrix and area under the curve graphs) can be found in Multimedia Appendix 4.

Table 1. Assessment distribution across the data set (N=7199).

Distribution	Frequency, n (%)
Postgraduate year	
1	1017 (14.13)
2	3139 (43.60)
3	405 (5.63)
4	1585 (22.02)
5	1053 (14.63)
Categories of work-based assessments	
Junior modules (PGY[a]1 and 2)	
Medical expert and scholar	1220 (16.95)
Advocacy and management	882 (12.25)
Communication and collaboration	1606 (22.31)
Professional and communicator	828 (11.50)
Pediatric emergency medicine	881 (12.24)
Senior modules (PGY3-5)	
Leadership and team management	582 (8.08)
Quality decision making	805 (11.18)
Teaching and scholarship	395 (5.49)
Numerical rating scores	
Missing matching qualitative comment	1602 (22.25)
3	12 (0.17)
4	79 (1.10)
5	635 (8.82)
6	2158 (29.98)
7	2713 (37.69)
Binary classification	
Missing matching qualitative comment	1602 (22.25)
1-5	726 (10.08)
6-7	4871 (67.66)

[a]PGY: postgraduate year.

Figure 1. Distribution of word counts in comments within the assessments, split by the resultant scores.

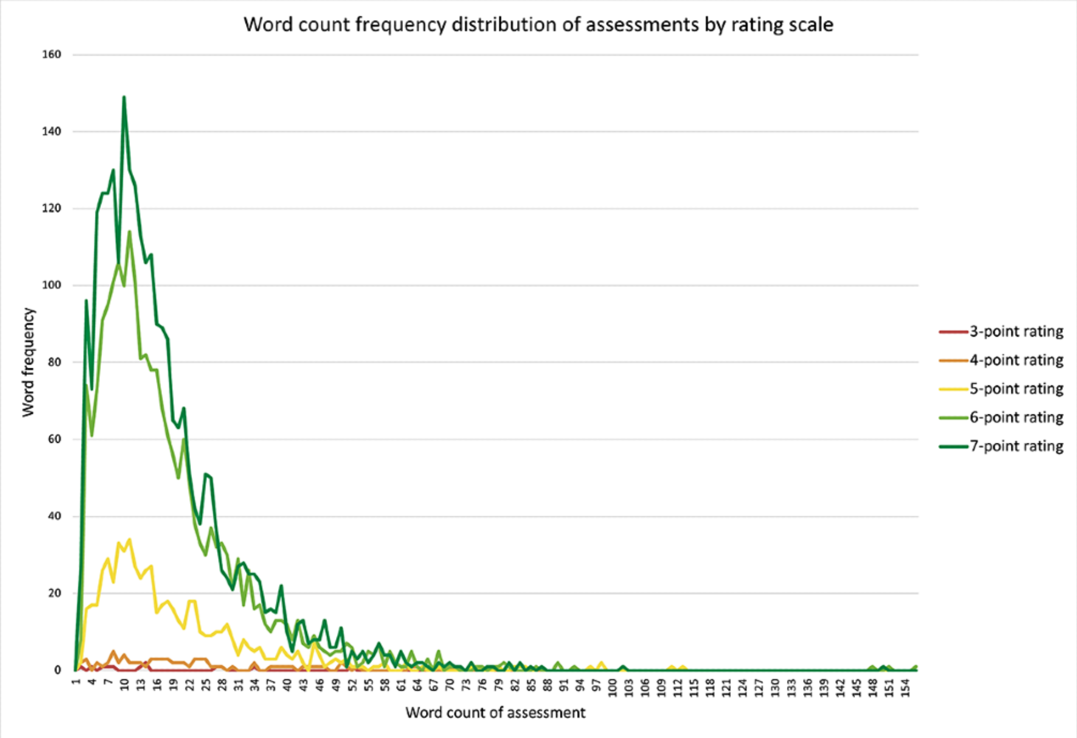


Table 2. Machine learning results on workplace-based assessments.

n-Gram input type	Sensitivity (%)		Specificity (%)		Accuracy of machine learning and natural language processing algorithms for diagnosing trainees at risk (%)	
	1-7 scale	Binary	1-7 scale	Binary	1-7 scale	Binary
Unigrams	82.0	92.2	5.1	13.9	51.8	82.0
Bigrams	81.1	87.3	4.1	21.4	50.4	85.6
Trigrams	79.6	87.0	Insufficient data to calculate	4.2	49.8	86.9

Discussion

Principal Findings

While NLP ML analyses have had many applications in health services (eg, interpreting large volumes of tweets or other data sets) [38,39], they are yet to be regularly used within the domain of aggregating and interpreting trainee-level data. This study demonstrates that an automated NLP ML analysis can identify resident performance that achieves competence on a direct-observation WBA using narrative comments.

While dichotomizing our 7-point assessment scale improved performance, the data set was not large enough to draw conclusions for specificity measures, due to a lack of true negatives within our trigram data. Specifically, our present ML model could not identify trainees who are failing with trigrams. The reason for low specificity was that our data had far fewer assessments on the lower end of the scale, especially for trigrams. While MLA can support the decision-making process, trainees who are at risk should be approached cautiously, triangulating data using human raters before decision making. However, the sensitivity of our algorithms suggests that we can harness the power of the NLP MLA to *rule out* the trainees who are not deemed at risk of meeting a performance standard. Human meta-raters could be most effectively deployed, then, to read those who have been flagged as possibly being at risk, and to make determinations of whether someone was truly at risk (eg, true positive) versus unduly flagged (eg, false positive). Moreover, McMAP is one of the first programmatic assessment systems in residency education [1,3,21]. It preceded a national shift to competency-based medical education by 6 years. There is no comparable pilot with similar accumulated data set yet since other programs began their system in 2017.

Based on this study, it is clear that larger data sets from amalgamated sources of common WBAs may hold the key to increasing the sampling (and therefore, the accuracy, sensitivity, and specificity) of our proposed algorithms. Early work within our specialty has shown that this may be possible [40], because all of specialist emergency medicine training have recently moved into a harmonized assessment system within Canada [41]. Finding ways to aggregate a nation's worth of WBA across a specialty and multiple sites will undoubtedly afford us enough data to power NLP MLAs that can be helpful for faculty decision makers and decrease the workload introduced by robust WBA programs.

This automated process could obviate the need for a manual review of all qualitative phrases. While the specificity of the automated process is quite poor to identify residents who have achieved competence in the task, this process allows our CCs and PDs to continuously monitor their trainees' performance. This will allow for an automated process to accurately identify trainees who may potentially require assistance or remediation. As our sensitivity ranged from 87% to 92.2%, we suggest that with higher stakes, summative decisions will still require human oversight and review to ensure that those who might be misclassified by the algorithm as requiring assistance (or needing more time) can be identified.

Comparison With the Prior Work

An exploratory study of residents' perception of WBA found that residents deemed feedback more valuable than numeric scores and acknowledged their skepticism on faculty comprehension of rating tasks [21]. Credibility is essential for feedback to be actionable. The factors that contribute to feedback credibility are the closeness of the relationship between supervisor and trainee, the consistency between the narrative and numeric score, and the quality of the narrative and a system that fosters a feedback culture [21,42]. This study demonstrates that ML and NLP can provide additional information on the evidence that supports results in WBA.

To complete a direct observation assessment, faculty undergo cognitive processes that involve observation, processing, and integration within the short time frame dedicated to the assessment [43,44]. When observing, the raters select the learner behaviors that are relevant to the assessment. These attributes may or may not be described in the narrative portion of the assessment. Processing involves the recollection of behaviors, matching behavior to a specific set or a subset of competencies, synthesizing the information collected, and integrating all the information into a narrative or numeric score [44]. Processing also responds to the individual conception of competency, context-specific settings, references to the highest and lowest performance witnessed by the rater, and emotions [43]. Intrarater reliability and consistency between narratives and numeric scores depend on the aforesaid cognitive process [45].

The interpretation of narrative comments is a complex task because words can be vague or have nonliteral meanings [18,46]. Raters and trainees decipher the alternative meanings of words using contextual information and experience. The precision of a narrative, the strength of the adjectives used, or specific references to competency domains are some of the elements to be considered when interpreting the hidden code [18,46]. As writing style differs between raters, the code is not universal and it can be mistakenly interpreted (eg, including areas of improvement in a narrative assessment might be considered positive or negative depending on the individual).

The traditional quantitative assessment paradigm leads learners and faculty to focus on numbers, and partially explains the complexity of the faculty task of "converting" or transferring their perception of competence into a 7-point scale. In fact, rater bias may be a result of the complexity of the unconscious action required to complete complex assessment tasks to assign scores to observations (very blunt, nonrich category) to a rater's judgment.

While not realized within our study, NLP analyses have been shown to provide information on the quality, usefulness, and relevance of narrative assessment [47-49]. Moreover, it can generate insights about identity of raters, their cognitive process, potential biases, and personality traits. For instance, the use of determiners, prepositions, and pronouns have been identified as features for gender discrimination [50] and relevant linguistic differences have been found in narratives from male and female faculties [51]. While human meta-raters (ie,. those who read others' comments) require more context about the feedback (eg, raters, audience, intent) [46], ML analysis can overcome the issues around context by increasing n-grams to match the scores based on qualitative data.

Strengths and Limitations

This study is a worked example that is based on real trainee data and frontline faculty assessors in the context of WBA. With a diverse team of educators, computer scientists, and clinicians, we have been able to move the mark toward solving a problem that many medical educators are facing around qualitative comments.

However, our study has also some limitations. Residency training selects for highly qualified and high-performing learners. As a result, assessments have a positive yield that creates a right-skewed data distribution, where residents tend to have higher ratings rather than low ratings. Our data were no different. The range restriction of our data has impacted our results.

Our data set was not sufficient to create a validation set. In the future, with more data, we will likely move toward having an 80% derivation, with 10% testing and 10% validation profile for our data partitioning. We acknowledge that there are limitations of the output of the model, but unfortunately, we are limited to the results we could obtain with these data. This early work will allow us to approximate sample sizes and to further the field toward an eventuality where the technology we currently have can be properly harvested in this area. We anticipate, based on our early work, that we will need data sets that are amalgamated by a country's worth of data to create the accuracy and precision required to truly make this a reality. With a larger data set we might have been able to complete more cross-validation procedures [52-56]. Human factors was another limitation in our study. Faculty members sometimes do not provide written comment with their ratings. Our study context is in an emergency department where there is not always time to provide any comment at all. We labeled them as missing in our study because we could not use them for NLP. Finally, our data set shows that greater pooling of data will be required by training programs (possibly across multiple centers or across a nation) to ensure that we have the depth of data to gain insights using NLP MLA technologies to advise CCs and PDs about trainees at risk. While there are some who might want to see a dichotomy between algorithms and humans, our team proposes that we should aspire for human-augmented decision making (eg, decision support), as opposed to assuming that MLAs might replace our training committees and faculty.

Future Directions

Using N-grams with different scales showed a great promise on the retrospective data. These results beg for confirmation in a prospective study. While we used our WBA based on 2 different scales, we highly anticipate that this result will show a similar pattern in entrustment scales. Therefore, future research should focus on entrustment scales.

Next, greater data sets will be required to adequately harness the power of NLP and MLA technologies to assist faculty members or trainees in terms of decision making around academic or clinical progress. There have been some great strides recently made in creating amalgamated trainee assessment data sets for nationalized program evaluation [40], but full data pooling and sharing will be required to adequately generate the insights that are required using these technologies. Greater attention must be paid to create harmonized data standards and safe reporting protocols so that we can pool both quantitative and qualitative data required to capitalize on the technologies that currently exist, and are used regularly in other sectors.

Finally, NLP and ML must be tested against the current reference standard of CC-driven insights so that we can decide whether ML results are truly useful to augment faculty decision making and help improve the decision-making process.

Conclusions

Our early data show promise that NLP with ML analysis of narrative assessment data may eventually serve as a decision-support system for CC, PDs, and other education decision makers. NLP and ML analyses have the potential to reduce the workload of large narrative data sets.

Acknowledgments

This study was supported by the 2020 Canadian Association of Emergency Physicians (CAEP) Emergency Medicine Advancement Fund. YY is the recipient of the The Scientific and Technological Research Council of Turkey (Türkiye Bilimsel ve Teknolojik Araştırma Kurumu, TUBITAK) Postdoctoral Fellowship grant.

Conflicts of Interest

TMC reports a honorarium from McMaster University for her education research work with the McMaster Education Research, Innovation, and Theory (MERIT) group and administrative stipend for her role of Associate Dean via the McMaster Faculty of Health Sciences Office of Continuing Professional Development. She also discloses that she has received various unrelated research grants, teaching honoraria, and speakership fees from academic institutions (Baylor University/Texas Children's Hospital, Catholic University of Korea, Harvard Medical School, International Association of Medical Sciences Educators, Northern Ontario School of Medicine, University of British Columbia, University of Northern British Columbia), nonprofit organizations (Physician Services Incorporated Foundation), physician organizations (Association of American Medical Colleges, Canadian Association of Emergency Physicians, Society of Academic Emergency Medicine, the Royal College of Physicians and Surgeons of Canada), and governmental sources (Government of Ontario, Virtual Learning Strategy eCampus Ontario program).

Multimedia Appendix 1
Sample McMAP task in knowledge translation.
[DOCX File , 17 KB - mededu_v8i2e30537_app1.docx]

Multimedia Appendix 2
Supplemental material for methods.
[DOCX File , 15 KB - mededu_v8i2e30537_app2.docx]

Multimedia Appendix 3
The different word cloud representations of the types of phrases found within comments associated with higher score (6 or 7 out of 7) and lower score comments (5 or less out of 7). While some similar word and phrases are found in the unigrams, bigrams, and trigrams, there is more homogeneity in the phrases within the higher score-associated trigrams and more diversity in the lower score-associated trigrams.
[PNG File , 1518 KB - mededu_v8i2e30537_app3.png]

Multimedia Appendix 4
Confusion matrix for binary class.
[DOCX File , 199 KB - mededu_v8i2e30537_app4.docx]

References

1. Acai A, Li S, Sherbino J, Chan TM. Attending Emergency Physicians' Perceptions of a Programmatic Workplace-Based Assessment System: The McMaster Modular Assessment Program (McMAP). Teach Learn Med 2019;31(4):434-444. [doi: 10.1080/10401334.2019.1574581] [Medline: 30835560]
2. Babu KS, Htike MM, Cleak VE. Workplace-based assessments in Wessex: the first 6 months. Psychiatr. Bull 2018 Jan 02;33(12):474-478. [doi: 10.1192/pb.bp.108.022889]
3. Chan T, Sherbino J. The McMaster Modular Assessment Program (McMAP). Academic Medicine 2015;90(7):900-905. [doi: 10.1097/acm.0000000000000707]
4. Gaunt A, Patel A, Rusius V, Royle TJ, Markham DH, Pawlikowska T. 'Playing the game': How do surgical trainees seek feedback using workplace-based assessment? Med Educ 2017 Sep;51(9):953-962. [doi: 10.1111/medu.13380] [Medline: 28833426]
5. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. Med Teach 2010;32(8):676-682. [doi: 10.3109/0142159X.2010.500704] [Medline: 20662580]
6. Massie J, Ali JM. Workplace-based assessment: a review of user perceptions and strategies to address the identified shortcomings. Adv in Health Sci Educ 2015 May 24;21(2):455-473. [doi: 10.1007/s10459-015-9614-0]
7. Moonen-van Loon JMW, Overeem K, Donkers HHLM, van der Vleuten CPM, Driessen EW. Composite reliability of a workplace-based assessment toolbox for postgraduate medical education. Adv Health Sci Educ Theory Pract 2013 Dec;18(5):1087-1102. [doi: 10.1007/s10459-013-9450-z] [Medline: 23494202]
8. Sebok-Syer SS, Klinger DA, Sherbino J, Chan TM. Mixed Messages or Miscommunication? Investigating the Relationship Between Assessors' Workplace-Based Assessment Scores and Written Comments. Acad Med 2017 Dec;92(12):1774-1779. [doi: 10.1097/ACM.0000000000001743] [Medline: 28562452]
9. Simmons M. How are workplace-based assessments viewed by UK psychiatry trainees. Psychiatria Danubina 2013;Suppl 2:S182-S184. [doi: 10.24869/psyd]
10. Li S, Acai A, Sherbino J, Chan TM. The Teacher, the Assessor, and the Patient Protector: A Conceptual Model Describing How Context Interfaces With the Supervisory Roles of Academic Emergency Physicians. AEM Education and Training 2020 Jan 26;5(1):52-62. [doi: 10.1002/aet2.10431]
11. Chan T, Sebok-Syer S, Thoma B, Wise A, Sherbino J, Pusic M. Learning Analytics in Medical Education Assessment: The Past, the Present, and the Future. AEM Education and Training 2018 Mar 22;2(2):178-187. [doi: 10.1002/aet2.10087]
12. Friedman KA, Raimo J, Spielmann K, Chaudhry S. Resident dashboards: helping your clinical competency committee visualize trainees' key performance indicators. Med Educ Online 2016;21:29838 [FREE Full text] [doi: 10.3402/meo.v21.29838] [Medline: 27037226]
13. Chary M, Parikh S, Manini AF, Boyer EW, Radeos M. A Review of Natural Language Processing in Medical Education. West J Emerg Med 2019 Jan;20(1):78-86 [FREE Full text] [doi: 10.5811/westjem.2018.11.39725] [Medline: 30643605]
14. Dias RD, Gupta A, Yule SJ. Using Machine Learning to Assess Physician Competence. Academic Medicine 2019;94(3):427-439. [doi: 10.1097/acm.0000000000002414]
15. Ariaeinejad A, Samavi R, Chan T, Doyle T. A Performance Predictive Model for Emergency Medicine Residents. In: CASCON '17: Proceedings of the 27th Annual International Conference on Computer Science and Software Engineering. Riverton, NJ: IBM Corp; 2017 Presented at: 27th Annual International Conference on Computer Science and Software Engineering; Nov 6-8, 2017; Markham, ON URL: https://dl.acm.org/doi/abs/10.5555/3172795.3172800
16. Ross S, Binczyk NM, Hamza DM, Schipper S, Humphries P, Nichols D, et al. Association of a Competency-Based Assessment System With Identification of and Support for Medical Residents in Difficulty. JAMA Netw Open 2018 Nov 02;1(7):e184581 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.4581] [Medline: 30646360]
17. Tremblay G, Carmichael P, Maziade J, Grégoire M. Detection of Residents With Progress Issues Using a Keyword-Specific Algorithm. J Grad Med Educ 2019 Dec;11(6):656-662 [FREE Full text] [doi: 10.4300/JGME-D-19-00386.1] [Medline: 31871565]
18. Ginsburg S, Eva K, Regehr G. Do In-Training Evaluation Reports Deserve Their Bad Reputations? A Study of the Reliability and Predictive Ability of ITER Scores and Narrative Comments. Academic Medicine 2013;88(10):1539-1544. [doi: 10.1097/acm.0b013e3182a36c3d]
19. Ginsburg S, van der Vleuten C, Eva KW, Lingard L. Hedging to save face: a linguistic analysis of written comments on in-training evaluation reports. Adv in Health Sci Educ 2015 Jul 17;21(1):175-188. [doi: 10.1007/s10459-015-9622-0]
20. Ginsburg S, van der Vleuten CPM, Eva KW. The Hidden Value of Narrative Comments for Assessment: A Quantitative Reliability Analysis of Qualitative Data. Acad Med 2017;92(11):1617-1621. [doi: 10.1097/ACM.0000000000001669] [Medline: 28403004]
21. Li S, Sherbino J, Chan TM. McMaster Modular Assessment Program (McMAP) Through the Years: Residents' Experience With an Evolving Feedback Culture Over a 3-year Period. AEM Education and Training 2017 Jan 19;1(1):5-14. [doi: 10.1002/aet2.10009]
22. Cheung W, Dudek N, Wood T, Frank J. Daily Encounter Cards-Evaluating the Quality of Documented Assessments. Journal of Graduate Medical Education 2016;8(4):601-604. [doi: 10.4300/jgme-d-15-00505.1]

23. Pattani R, Ginsburg S, Mascarenhas Johnson A, Moore JE, Jassemi S, Straus SE. Organizational Factors Contributing to Incivility at an Academic Medical Center and Systems-Based Solutions. Academic Medicine 2018;93(10):1569-1575. [doi: 10.1097/acm.0000000000002310]
24. Tavares W, Ginsburg S, Eva KW. Selecting and Simplifying: Rater Performance and Behavior When Considering Multiple Competencies. Teach Learn Med 2016;28(1):41-51. [doi: 10.1080/10401334.2015.1107489] [Medline: 26787084]
25. Dayal A, O'Connor DM, Qadri U, Arora VM. Comparison of Male vs Female Resident Milestone Evaluations by Faculty During Emergency Medicine Residency Training. JAMA Intern Med 2017 May 01;177(5):651-657. [doi: 10.1001/jamainternmed.2016.9616]
26. Mueller A, Jenkins T, Osborne M, Dayal A, O'Connor D, Arora V. Gender Differences in Attending Physicians? Feedback to Residents: A Qualitative Analysis. Journal of Graduate Medical Education 2017;9(5):585. [doi: 10.4300/jgme-d-17-00126.1]
27. Chahine S, Cristancho S, Padgett J, Lingard L. How do small groups make decisions? Perspect Med Educ 2017 May 22;6(3):192-198. [doi: 10.1007/s40037-017-0357-x]
28. Hauer KE, Chesluk B, Iobst W, Holmboe E, Baron RB, Boscardin CK, et al. Reviewing Residents' Competence. Academic Medicine 2015;90(8):1084-1092. [doi: 10.1097/acm.0000000000000736]
29. Hauer K, Cate OT, Boscardin C, Iobst W, Holmboe E, Chesluk B, et al. Ensuring Resident Competence: A Narrative Review of the Literature on Group Decision Making to Inform the Work of Clinical Competency Committees. Journal of Graduate Medical Education 2016;8(2):156-164. [doi: 10.4300/jgme-d-15-00144.1]
30. Kinnear B, Warm EJ, Hauer KE. Twelve tips to maximize the value of a clinical competency committee in postgraduate medical education. Medical Teacher 2018 Jun 26;40(11):1110-1115. [doi: 10.1080/0142159x.2018.1474191]
31. Ekpenyong A, Baker E, Harris I, Tekian A, Abrams R, Reddy S, et al. How do clinical competency committees use different sources of data to assess residents' performance on the internal medicine milestones?A mixed methods pilot study. Med Teach 2017 Oct;39(10):1074-1083. [doi: 10.1080/0142159X.2017.1353070] [Medline: 28738746]
32. Schumacher D, King B, Barnes M, Elliott S, Gibbs K, McGreevy J, et al. Members of the APPD LEARN CCC Study Group. Influence of Clinical Competency Committee Review Process on Summative Resident Assessment Decisions. Journal of Graduate Medical Education 2018 Aug;10(4):437. [doi: 10.4300/jgme-d-17-00762.1]
33. Schumacher DJ, Michelson C, Poynter S, Barnes MM, Li ST, Burman N, APPD LEARN CCC Study Group, et al. Thresholds and interpretations: How clinical competency committees identify pediatric residents with performance concerns. Med Teach 2018 Jan;40(1):70-79. [doi: 10.1080/0142159X.2017.1394576] [Medline: 29345207]
34. IBM. IBM SPSS Statistics for Windows, Version 26. Armonk, NY: IBM Corp; 2019.
35. R2019b - Updates to the MATLAB and Simulink product families Internet. MathWorks. URL: https://www.mathworks.com/products/new_products/release2019b.html [accessed 2022-05-12]
36. Chan TM, Sherbino J, Mercuri M. Nuance and Noise: Lessons Learned From Longitudinal Aggregated Assessment Data. J Grad Med Educ 2017 Dec;9(6):724-729 [FREE Full text] [doi: 10.4300/JGME-D-17-00086.1] [Medline: 29270262]
37. Liu Y, Chen PC, Krause J, Peng L. How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature. JAMA 2019 Nov 12;322(18):1806-1816. [doi: 10.1001/jama.2019.16489] [Medline: 31714992]
38. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
39. Dimitrov D, Baran E, Fafalios P, Yu R, Zhu X, Zloch M, et al. TweetsCOV19 - A Knowledge Base of Semantically Annotated Tweets about the COVID-19 Pandemic. In: CIKM '20: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. NY, NY, USA: Association for Computing Machinery; 2020 Presented at: 29th ACM International Conference on Information & Knowledge Management; October 19-23, 2020; Virtual Event, Hosted in Ireland p. 2991-2998 URL: https://dl.acm.org/doi/abs/10.1145/3340531.3412765 [doi: 10.1145/3340531.3412765]
40. Thoma B, Hall A, Clark K, Meshkat N, Cheung W, Desaulniers P, et al. Evaluation of a National Competency-Based Assessment System in Emergency Medicine: A CanDREAM Study. Journal of Graduate Medical Education 2020 Aug;12(4):425-434. [doi: 10.4300/jgme-d-19-00803.1]
41. Sherbino J, Bandiera G, Doyle K, Frank JR, Holroyd BR, Jones G, et al. The competency-based medical education evolution of Canadian emergency medicine specialist training. CJEM 2019 Sep 30;22(1):95-102. [doi: 10.1017/cem.2019.417]
42. Govaerts M. Workplace-Based Assessment and Assessment for Learning: Threats to Validity. Journal of Graduate Medical Education 2015;7(2):265-267. [doi: 10.4300/jgme-d-15-00101.1]
43. Gauthier G, St-Onge C, Tavares W. Rater cognition: review and integration of research findings. Med Educ 2016 Apr 13;50(5):511-522. [doi: 10.1111/medu.12973]
44. Hanson JL, Rosenberg AA, Lane JL. Narrative descriptions should replace grades and numerical ratings for clinical performance in medical education in the United States. Front. Psychol 2013;4:668. [doi: 10.3389/fpsyg.2013.00668]
45. Gingerich A, Kogan J, Yeates P, Govaerts M, Holmboe E. Seeing the 'black box' differently: assessor cognition from three research perspectives. Med Educ 2014 Oct 12;48(11):1055-1068. [doi: 10.1111/medu.12546]
46. Ginsburg S, Kogan J, Gingerich A, Lynch M, Watling C. Taken out of Context: Hazards in the Interpretation of Written Assessment Comments. Academic Medicine 2020:2. [doi: 10.1097/acm.0000000000003047]

47. Chen J, Druhl E, Polepalli RB, Houston TK, Brandt CA, Zulman DM, et al. A Natural Language Processing System That Links Medical Terms in Electronic Health Record Notes to Lay Definitions: System Development Using Physician Reviews. J Med Internet Res 2018 Jan 22;20(1):e26 [FREE Full text] [doi: 10.2196/jmir.8669] [Medline: 29358159]
48. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? Journal of Biomedical Informatics 2009 Oct;42(5):760-772. [doi: 10.1016/j.jbi.2009.08.007]
49. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc 2011 Sep;18(5):544-551 [FREE Full text] [doi: 10.1136/amiajnl-2011-000464] [Medline: 21846786]
50. Argamon S, Koppel M, Pennebaker JW, Schler J. Automatically profiling the author of an anonymous text. Commun. ACM 2009 Feb;52(2):119-123. [doi: 10.1145/1461928.1461959]
51. Heath JK, Weissman GE, Clancy CB, Shou H, Farrar JT, Dine CJ. Assessment of Gender-Based Linguistic Differences in Physician Trainee Evaluations of Medical Faculty Using Automated Text Mining. JAMA Netw Open 2019 May 10;2(5):e193520. [doi: 10.1001/jamanetworkopen.2019.3520]
52. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge, Massachusetts: The MIT Press; 2016.
53. Althnian A, AlSaeed D, Al-Baity H, Samha A, Dris AB, Alzakari N, et al. Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain. Applied Sciences 2021 Jan 15;11(2):796. [doi: 10.3390/app11020796]
54. Sordo M, Zeng Q. On Sample Size and Classification Accuracy: A Performance Comparison. In: International Symposium on Biological and Medical Data Analysis. Berlin, Heidelberg: Springer; 2005:193-201.
55. Prusa J, Khoshgoftaar T, Seliya N. The Effect of Dataset Size on Training Tweet Sentiment Classifiers. In: Proceedings of the 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA).: IEEE; 2015 Presented at: IEEE 14th International Conference on Machine Learning and Applications (ICMLA); December 9-11, 2015; Miami, Florida, USA p. 96-102. [doi: 10.1109/icmla.2015.22]
56. Rahman MS, Sultana M. Performance of Firth-and logF-type penalized methods in risk prediction for small or sparse binary data. BMC Med Res Methodol 2017 Feb 23;17(1):33 [FREE Full text] [doi: 10.1186/s12874-017-0313-9] [Medline: 28231767]

Abbreviations

CanMEDS: The Canadian Medical Education Directives for Specialists
CC: competence committee
ECOC: Error-Correcting Output Codes
McMAP: McMaster Modular Assessment Program
ML: machine learning
MLA: machine learning algorithm
NLP: natural language processing
PD: program director
WBA: workplace-based assessment

Edited by T Leung; submitted 19.05.21; peer-reviewed by T Vukušić Rukavina, M Elbattah, F Rudzicz; comments to author 18.08.21; revised version received 05.12.21; accepted 30.04.22; published 27.05.22.

Please cite as:
Yilmaz Y, Jurado Nunez A, Ariaeinejad A, Lee M, Sherbino J, Chan TM
Harnessing Natural Language Processing to Support Decisions Around Workplace-Based Assessment: Machine Learning Study of Competency-Based Medical Education
JMIR Med Educ 2022;8(2):e30537
URL: https://mededu.jmir.org/2022/2/e30537
doi:10.2196/30537
PMID:35622398

Original Paper

Critical Comparison of the Quality and Content of Integrated Vascular Surgery, Thoracic Surgery, and Interventional Radiology Residency Training Program Websites: Qualitative Study

Katherine Jensen[1], BSc; Qi Yan[1], MD; Mark G Davies[1], MHPE, MD, DPhil

Division of Vascular and Endovascular Surgery, University of Texas Health at San Antonio, San Antonio, TX, United States

Corresponding Author:
Mark G Davies, MHPE, MD, DPhil
Division of Vascular and Endovascular Surgery
University of Texas Health at San Antonio
7703 Floyd Curl Drive
MC7741
San Antonio, TX, 78229
United States
Phone: 1 210 567 5715
Fax: 1 210 567 1762
Email: Daviesm@uthscsa.edu

Abstract

Background: With the move to virtual interviewing, residency websites are an important recruitment resource, introducing applicants to programs across the country and allowing for comparison. Recruitment is highly competitive from a common potential pool between vascular surgery, thoracic surgery, and interventional radiology with the ratio of applicants to positions being highest in interventional radiology, followed by thoracic surgery and lastly vascular surgery, as reported by the National Resident Matching Program.

Objective: The aim of this study is to evaluate the accessibility and availability of online content for those integrated residency programs.

Methods: A list of accredited vascular surgery, thoracic surgery, and interventional radiology residencies was obtained from the Accreditation Council for Graduate Medical Education (ACGME) database. Program websites were evaluated by trained independent reviewers (n=2) for content items pertaining to program recruitment and education (scored absent or present). Statistical analysis was performed in R software.

Results: Of ACGME-accredited programs, 56 of 61 (92%) vascular surgery, 27 of 27 (100%) thoracic surgery, and 74 of 85 (87%) interventional radiology programs had functional websites (P=.12). Vascular surgery websites contained a median of 26 (IQR 20-32) content items, thoracic surgery websites contained a median of 27 (IQR 21-32) content items, and interventional radiology websites contained a median of 23 (IQR 18-27) content items. Two content items considered highly influential to applicant program decisions are procedural experience and faculty mentorship, which were reported at 32% (18/56) and 11% (6/56) for vascular surgery, 19% (5/27) and 11% (3/27) for thoracic surgery, and 50% (37/74) and 15% (11/74) for interventional radiology (P=.008 and P=.75), respectively. Key deficits were work hours, debt management, and curriculum for interventional radiology; resident profiles, sample contracts, and research interests in vascular surgery; and operative experiences and the program director’s contact and message for thoracic surgery. Interventional radiology deficits were work hours, and thoracic surgery deficits were procedural experience. Both interventional radiology and thoracic surgery websites lacked information on evaluation criteria and faculty mentorship.

Conclusions: This study has uncovered key differences in the availability of online content for residencies recruiting from the same pool of applicants. Thoracic surgery has the most information, followed by vascular surgery, with interventional radiology reporting the least content. In the era of virtual interviewing from the same potential pool of applicants, programs should review and revise their web presence with the aim to increase the availability of online content to attract valuable candidates.

(JMIR Med Educ 2022;8(2):e35074) doi:10.2196/35074

KEYWORDS

training; recruitment; website; content; quality; vascular surgery; thoracic surgery; interventional radiology; radiology; surgery; web-based; web resource; surgeon; comparison; residency; integrated program

Introduction

The role of vascular surgeons in the medical environment has changed considerably with the increasing use of endovascular approaches for treatment of vascular lesions [1]. By 2026, it is predicted that 75% to 95% of overall vascular lesions (aneurysms, stenosis, occlusive disease, traumatic vascular lesions, etc) will be treated endovascularly [1,2]. Vascular surgery, as always, will continue to compete in recruitment with cardiac surgery for procedural domain, but with the increasing use of endovascular approaches, it faces additional recruitment competition from interventional radiology [2]. Due to the overlap in patient populations, professional interests, skills, and treatments performed by vascular surgeons, thoracic surgeons, and interventional radiologists, these specialties appeal to a common potential applicant pool, and recruitment is highly competitive among these training programs.

Candidates for residency programs increasingly use the internet to research potential programs for application [3-6]. Online information has been analyzed for a range of residency and fellowship programs, including orthopedic surgery, plastic and reconstructive surgery, emergency medicine, cardiothoracic surgery, neurosurgery, otolaryngology, trauma surgery, surgical critical care, acute care surgery, microsurgery, interventional radiology, and vascular surgery [3-5,7-27]. Studies have individually analyzed the availability of online content for integrated vascular surgery [27], thoracic surgery [12], and interventional radiology [15,25] training program websites, but to our knowledge, no study has compared the accessibility and availability of online content across these training paradigms. Given the importance of online resources in recruiting prospective applicants and the current mandates to move to virtual interviewing, we sought to assess the current state of integrated vascular surgery, thoracic surgery, and interventional radiology training program websites. The purpose of this study is to evaluate the presence, accessibility, and comprehensiveness of integrated vascular surgery, thoracic surgery, and interventional radiology training program websites.

Methods

Study Design

A comprehensive list of accredited integrated vascular surgery, thoracic surgery, and interventional radiology residencies was obtained from the Accreditation Council for Graduate Medical Education (ACGME) database. Programs participating in the 2020 National Resident Matching Program (NRMP) were eligible for study inclusion. Following identification of all programs with websites, programs were accessed and evaluated by two independent reviewers (one medical student and one resident) for availability of recruitment and educational content items. The websites were viewed independently by each reviewer. The program search and review was performed in November 2019.

Research Question

Are there key differences in the three specialty program websites for integrated residencies that could potentially impair recruitment efforts in the virtual environment?

Accessibility of Websites

Accessibility of websites was determined by surveying the ACGME database for the total number of programs listed and the presence or absence of website links. Links, if they were provided, were characterized as either functional or nonfunctional. Functional links led to a website. Nonfunctional links led to an error page. Functional links were then evaluated as being either *direct* (landing directly on the program webpage) or *indirect* (landing on a different page such as the departmental website, requiring further action by the reviewer to access the specific program webpage if possible).

Availability of Content

Websites for integrated vascular surgery, thoracic surgery, and interventional radiology residency programs were analyzed for availability of information used to inform and recruit prospective applicants. Content items on recruitment and education (listed in Textbox 1) were selected based on ACGME program requirements as well as previously published literature reviewing the online content of residency and fellowship programs [5,14,19,20]. Content on the training program websites was counted as present if it was present on the main training program webpage or it was accessible via a direct link provided on the main training program webpage.

Textbox 1. Content features included in evaluation of integrated vascular surgery, integrated thoracic surgery, and integrated interventional radiology training program websites.

Program recruitment (n=41)

- Program description
- Number incoming positions available
- Faculty listing
- Faculty education and training history
- Faculty profile (descriptive)
- Faculty publications
- Faculty contact information
- Current residents
- Resident education history
- Resident profiles
- Resident contact information
- Alumni listing
- Alumni education history
- Alumni contact information
- Alumni career placement
- Board examination performance
- Program chair message
- Program director message
- Program director contact
- Administrative/coordinator contact
- Facility description
- Application requirements
- Selection process
- Interview dates
- Interview day details
- Electronic Residency Application Service (ERAS) link
- If present, is ERAS link functional?
- Call requirement
- Contract
- Salary
- Work hours
- Benefits
- Vacation
- City information
- Domestic considerations
- Well-being strategies
- Debt management
- Meal allowance
- Educational fund
- Parking
- Visa

Program education (n=16)

- Rotation schedule
- Didactic instruction
- Research requirements
- Research interests (department/faculty)
- Operative experience
- Journal club
- Conference schedule
- National/regional meetings attended
- Evaluation criteria
- Faculty mentorship
- National organization link
- Curriculum
- Company link
- Elective rotation
- Simulation training
- Vascular lab

Program Recruitment and Education

Websites were evaluated for content relevant to program recruitment and education. Program recruitment information included faculty listings, faculty and departmental research interests, alumni career placements, and information on current residents. Recruitment information regarding the application and interview process as well as general resident quality of life metrics were also evaluated (see Textbox 1). Program education content addressed operative and didactic training. It also covered resident research opportunities. Overall, 41 program recruitment and 16 program education content items were evaluated.

Rater Training and Consistency

Each website was accessed and evaluated by two reviewers (one medical student and one resident) for availability of content items as well as quality of websites (determined as a function of four dimensions: content, design, organization, and user friendliness). Each reviewer was trained by examining an optimal website, an average website, and a below average website with the senior author. Disputed assessments were resolved by consensus following discussion with the senior author. Reviewers were not blinded.

Overall, there was considerable interrater reliability with 81% agreement (κ=0.74).

Data Analysis

Intergroup analysis of continuous variables was performed using ANOVA. Categorical variables were compared using chi-square analysis. Statistical significance was defined as $P<.05$. Percent agreement and kappa statistics were calculated for interrater reliability. Statistical analysis was performed using statistical software R version 4.0.0 (R Foundation for Statistical Computing).

Ethical Approval

All data reviewed was open to the public, and there was no contact with fellowship staff; thus, no institutional review board review, ethics approval, or informed consent was necessary.

Results

Accessibility of Websites

Of the programs included in this analysis, 87% (53/61) of the vascular surgery, 89% (24/27) of the thoracic surgery, and 95% (81/85) of the interventional radiology programs provided a link to their program webpage on the ACGME webpage ($P=.18$). Of those programs that provided links, the majority of the links were functional with no difference between the specialties ($P=.24$). However, few links landed directly on the program webpage. Less than one-third of the programs with functional links provided links that landed directly on the program webpage ($P=.52$). Overall, 56 of 61 (92%) vascular surgery, 27 of 27 (100%) thoracic surgery, and 74 of 85 (87%) interventional radiology programs had a dedicated webpage (Table 1).

Table 1. Accessibility of integrated vascular surgery, integrated thoracic surgery, and integrated interventional radiology training program websites from the Accreditation Council for Graduate Medical Education webpage.

	Vascular surgery	Thoracic surgery	Interventional radiology	*P* value
Programs	61	27	85	N/A[a]
Providing website links[b], n (%)	53 (87)	24 (89)	81 (95)	.18
Functioning links, n (%)	47 (89)	21 (88)	74 (91)	.24
Direct links, n (%)	17 (32)	7 (33)	17 (23)	.52

[a]N/A: not applicable.

[b]Accreditation Council for Graduate Medical Education links were accessed November 2019.

Availability of Content

Content was assessed in two domains: recruitment and education. Of the 57 recruitment and educational content items included in this analysis, vascular surgery program webpages contained a median of 26 (IQR 20-32) content items, thoracic surgery program webpages contained a median of 27 (IQR 21-32) content items, and interventional radiology program webpages contained a median of 23 (IQR 18-27) content items. Of the 41 recruitment content items included in this analysis, vascular surgery program webpages contained a median of 19.5 (IQR 15-24) content items, thoracic surgery program webpages contained a median of 20 (IQR 16-24) content items, and interventional radiology program webpages contained a median of 18 (IQR 15-21) content items. Of the 16 education content items included in this analysis, vascular surgery program webpages contained a median of 7 (IQR 4-9) content items, thoracic surgery program webpages contained a median of 6 (IQR 4-7) content items, and interventional radiology program webpages contained a median of 4 (IQR 3-7) content items (Table 2).

Table 2. Availability of content on US integrated vascular surgery, integrated thoracic surgery, and integrated interventional radiology training program websites.

	Vascular surgery (n=56), n (%)	Thoracic surgery (n=27), n (%)	Interventional radiology (n=74), n (%)	*P* value
Program recruitment				
Program description	56 (100)	26 (96)	69 (93)	.14
Faculty listing	53 (95)	27 (100)	73 (99)	.23
Faculty education (training history)	52 (93)	25 (93)	68 (92)	.98
Admin/coordinator contact	53 (95)	22 (82)	63 (85)	.14
Faculty profile (descriptive)	46 (82)	19 (70)	68 (92)	*.02* [a]
Application requirements	42 (75)	22 (82)	50 (68)	.34
ERAS[b] link	35 (63)	16 (59)	47 (64)	.93
If present, is ERAS link functional?	35 (63)	16 (59)	47 (64)	.93
Benefits	39 (70)	19 (70)	51 (69)	.99
Facility description	50 (89)	22 (82)	49 (66)	*.007*
Number of incoming positions available	44 (79)	20 (74)	45 (61)	.08
Salary	34 (61)	17 (63)	48 (65)	.89
Current residents	40 (71)	20 (74)	47 (64)	.49
Vacation policy	32 (57)	18 (67)	46 (62)	.69
Program director contact	24 (43)	8 (30)	43 (58)	*.03*
Faculty publications	24 (43)	19 (70)	40 (54)	.06
Well-being strategies	26 (46)	17 (63)	37 (50)	.36
City information	33 (59)	16 (59)	35 (47)	.34
Educational fund	19 (34)	12 (44)	35 (47)	.30
Resident education history	35 (63)	17 (63)	34 (46)	.11
Parking	17 (30)	9 (33)	34 (46)	.16
Domestic considerations	32 (57)	16 (59)	32 (43)	.19
Visa	22 (39)	12 (44)	32 (43)	.87
Interview dates	32 (57)	13 (48)	29 (39)	.13
Faculty contact information	17 (30)	9 (33)	27 (37)	.76
Meal allowance	16 (29)	11 (41)	27 (37)	.48
Sample contract	9 (16)	7 (26)	21 (28)	.25
Call requirement	22 (39)	16 (59)	21 (28)	*.02*
Program director message	17 (30)	3 (11)	18 (24)	.16
Resident profiles	10 (18)	7 (26)	18 (24)	.60
Interview details	10 (18)	5 (19)	18 (24)	.63
Alumni listing	20 (36)	7 (26)	17 (23)	.27
Alumni career placement	15 (27)	6 (22)	17 (23)	.85
Selection process	4 (7)	2 (7)	13 (18)	.14
Debt management	22 (39)	11 (41)	8 (11)	*<.001*
Resident contact information	5 (9)	3 (11)	6 (8)	.90
Work hours	12 (21)	12 (44)	6 (8)	*<.001*
Alumni education history	7 (13)	1 (4)	2 (3)	.06
Alumni contact information	1 (2)	0 (0)	1 (1)	.79

	Vascular surgery (n=56), n (%)	Thoracic surgery (n=27), n (%)	Interventional radiology (n=74), n (%)	*P* value
Program chair message	5 (9)	2 (7)	1 (1)	.13
Board examination performance	4 (7)	0 (0)	1 (1)	.10
Program education				
Rotation schedule	47 (84)	19 (70)	45 (61)	*.02*
Research interests (department)	22 (39)	16 (59)	49 (66)	*.008*
Didactic instruction	39 (70)	17 (63)	42 (57)	.32
Research requirements	41 (73)	19 (70)	39 (53)	*.04*
Operative experience	18 (32)	5 (19)	37 (50)	*.008*
Journal club	34 (61)	11 (41)	25 (34)	*.008*
Meetings attended	26 (46)	11 (41)	23 (31)	.20
Elective rotation	25 (45)	12 (44)	20 (27)	.07
Conference schedule	21 (38)	6 (22)	18 (24)	.19
Curriculum	21 (38)	10 (37)	12 (16)	*.01*
Faculty mentorship	6 (11)	3 (11)	11 (15)	.75
Vascular lab	36 (64)	2 (7)	10 (14)	*<.001*
National organization link	4 (7)	7 (26)	8 (11)	*.04*
Evaluation criteria	9 (16)	6 (22)	5 (7)	.08
Simulation training	19 (34)	9 (33)	3 (4)	*<.001*
Company link	2 (4)	1 (4)	0 (0)	.26

[a]Italics indicate significant values.

[b]ERAS: Electronic Residency Application Service.

Vascular Surgery

For program recruitment, almost all programs provided information on program description, faculty listing, faculty education, administrator or coordinator contact information, facility description, descriptive faculty profiles, and the number of incoming positions. The majority of programs provided information on application requirements, a functional link to the Electronic Residency Application Service (ERAS), benefits, salary, current residents, city information, resident education history, domestic considerations, vacation policy, and interview dates. Less than one-half of programs provided information on program director contact information, faculty publications, well-being strategies, faculty contact information, program director message, alumni listing, alumni career placement, educational fund, parking, nonnational visa information, meal allowance, call requirement, alumni career placement, and debt management. Fewer than one-quarter of the programs provided information on sample contracts, resident profiles, interview details, work hours, and alumni education history. Almost no programs provided information on their selection process, resident contact information, program director message, board examination performance, and alumni contact information (Table 2).

For program education, almost all programs provided information on rotation schedule (84%; 47/56). The majority of programs provided information on didactic instruction, research requirements, journal club, and vascular lab training (Registered Physician in Vascular Interpretation [RPVI]). Less than one-half of programs provided information on departmental research interests, operative experience, meetings attended, elective rotations, conference schedule, curriculum, and simulation training. Fewer than one-quarter of the programs provided information on evaluation criteria and faculty mentorship. Almost no programs provided information on national organizational links and cardiovascular product company links (Table 2).

Thoracic Surgery

For program recruitment, almost all programs provided information on program description, faculty listing, faculty education, administrator or coordinator contact information, facility description, and application requirements. The majority of programs provided information on descriptive faculty profiles, a functional link to ERAS, benefits, the number of incoming positions available, salary, current residents, faculty publications, well-being strategy, city information, resident education history, vacation policy, call requirements, and domestic considerations. Less than one-half of programs provided information on program director contact information, interview date, faculty contact information, sample contracts, resident profiles, alumni listings, debt management, educational fund, parking, nonnational visa information, meal allowance, and work hours. Fewer than one-quarter of the programs provided information on program director message, alumni career placement, interview details, and resident contact information. Almost no programs provided information on selection process, alumni education history, program chair message, board examination performance, and alumni contact information (Table 2).

For program education, the majority of programs provided information on rotation schedule, departmental research interests, didactic instruction, and research requirements. Less than one-half of programs provided information on journal club, meetings attended, elective rotation, curriculum, national organization links, and simulation training. Fewer than one-quarter of the programs provided information on operative experiences, conference schedule, faculty mentorship, and evaluation criteria (Table 2). Almost no programs provided information on vascular lab training (RPVI), which should be expected as it is not a core component of thoracic surgery.

Interventional Radiology

For program recruitment, almost all programs provided information on program description, faculty listing, faculty education, administrator or coordinator contact information, and descriptive faculty profiles. The majority of programs provided information on application requirements, a functional link to ERAS, facility description, vacation policy, benefits, the number of incoming positions available, salary, current residents, program director contact information, and faculty publications. Less than one-half of programs provided information on well-being strategies, city information, educational fund, parking, nonnational visa information, meal allowance, call requirement, resident education history, domestic considerations, interview dates, faculty contact information, and sample contracts. -quarter of the programs provided information on a program director message, resident profiles, interview details, alumni listing, alumni career placement, selection process, and debt management. Almost no programs provided information on resident contact information, work hours, alumni education history, program chair message, board examination performance, and alumni contact information (Table 2).

For program education, the majority of programs provided information on rotation schedule, departmental research interests (49/74, 66%), didactic instruction, and research requirements. Less than one-half of programs provided information on operative experiences, journal club, meetings attended, and elective rotations. Fewer than one-quarter of the programs provided information on conference schedule, curriculum, faculty mentorship, vascular lab training (RPVI), and national organization links. Almost no programs provided information on evaluation criteria, simulation training, and cardiovascular product company links (Table 2).

Comparison of Content Availability

Vascular surgery webpages provided the most information on rotation schedule, journal club, and vascular lab (as compared to thoracic surgery and interventional radiology webpages (*P*=.02, *P*=.008, and *P*<.001, respectively). Vascular surgery webpages provided less information on departmental research interests as compared to thoracic surgery and interventional radiology webpages (*P*=.008; Table 2).

Thoracic surgery webpages provided the most information on call requirement, national organization link, and work hours as compared to vascular surgery and interventional radiology webpages (*P*=.02, *P*=.04, and *P*<.001, respectively). Thoracic surgery webpages provided less information on descriptive faculty profile as compared to vascular surgery and interventional radiology webpages (*P*=.02; Table 2).

Interventional radiology webpages provided the most information on operative experience and program director contact information as compared to vascular and thoracic surgery webpages (*P*=.008 and *P*=.03, respectively). Interventional radiology webpages provided less information on facility description, debt management (*P*=.007), research requirements (*P*<.001), curriculum (*P*=.04), and simulation training (*P*<.001) as compared to vascular surgery and thoracic surgery webpages (Table 2).

Quality of Websites

On an overall assessment, integrated vascular surgery, thoracic surgery, and interventional radiology websites were found to be comparable. The average vascular surgery website score was 2.66 (SD 0.95), the average thoracic surgery website score was 2.18 (SD 0.92), and the average interventional radiology website score was 2.25 (SD 0.88). The vascular surgery websites had the highest scores in content, design, organization, and user-friendliness. The thoracic surgery websites had the lowest scores in content, organization, and user-friendliness, while the interventional radiology websites had the lowest score in design. Additional details regarding website quality, broken down by category, are visible in Table 3.

Table 3. Quality of US integrated vascular surgery, integrated thoracic surgery, and integrated interventional radiology training program websites[a].

	Content, mean (SD)	Design, mean (SD)	Organization, mean (SD)	User friendliness, mean (SD)	Average quality, mean (SD)
Vascular surgery	2.57 (0.95)	2.59 (0.99)	2.75 (0.96)	2.73 (0.90)	2.66 (0.95)
Thoracic surgery	2.04 (0.85)	2.22 (0.89)	2.19 (1.08)	2.26 (0.86)	2.18 (0.92)
Interventional radiology	2.27 (0.90)	2.05 (0.77)	2.32 (0.97)	2.34 (0.88)	2.25 (0.88)

[a]Scale: 1=poor, 2=acceptable, 3=good, 4=great.

Discussion

Principal Findings

As resident recruitment moves to a virtual platform, the internet is an increasingly important resource for residency applicants as they research programs. Thoracic surgery program webpages had the most information, followed by vascular surgery program webpages, with interventional radiology program webpages reporting the least content. This trend in availability of content items mirrors the percent of positions filled by each specialty, with 100% of PGY-1 thoracic surgery positions filled, 97% of vascular surgery PGY-1 positions filled, and 97% of PGY-1 interventional radiology positions filled (with 94% of PGY-2 interventional radiology positions filled), as reported by the NRMP 2020 Main Residency Match Results and Data report [28].

Other factors, beyond program websites, that have been identified to influence applicant interest in a program include geography, advice from a mentor, advice from a peer, and other online information. The integrated vascular track was first accredited by the ACGME in 2006 [27], the first integrated thoracic surgery program accepted residents in 2007 [29], and the first integrated interventional radiology programs participated in the NRMP in 2016 [30]. The majority of these integrated programs have been established for less than 10 years. This increase in the number of integrated programs, though necessary to meet the high demand for integrated residency positions, means that many programs do not have an established national presence. Applicants cannot receive the same quality of advice from mentors and peers on newer programs, as compared to programs that have been established for longer periods of time. Furthermore, many programs are geographically clustered, specifically in the northeast and along the west coast (see Figure 1). These factors combine to place additional weight on program websites, perhaps serving as the initial source of information for potential applicants and allowing for comparison.

Of ACGME-accredited programs, 56 of 61 (92%) vascular surgery programs, 27 of 27 (100%) thoracic surgery programs, and 74 of 85 (87%) interventional radiology programs had functional websites. Thoracic surgery program webpages had the most information (content item median 27, IQR 21-32), then vascular surgery program webpages (content item median 26, IQR 20-32), with interventional radiology program webpages reporting the least content (content item median 23, IQR 18-27). The greater amount of content on vascular surgery and thoracic surgery program webpages could be expected, given the young age of many interventional radiology programs. Previous studies have acknowledged integrated interventional radiology program webpages to be a work in progress [15]. This study confirms that finding in relation to longer-established vascular surgery and thoracic surgery program webpages.

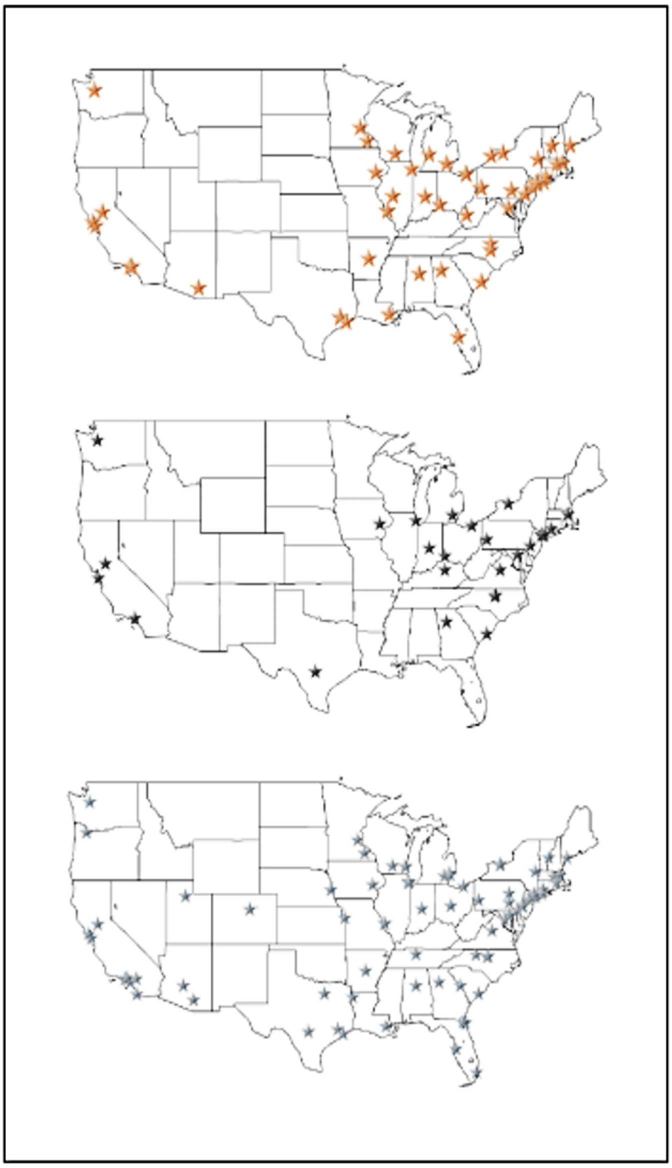
Two content items that have been identified to be highly influential to the applicant program decision are operative experience and faculty mentorship [31-33]. This analysis found those items to be reported at 32% (18/56) and 11% (6/56) for vascular surgery, 19% (5/27) and 11% (3/27) for thoracic surgery, and 50% (37/74) and 15% (11/74) for interventional radiology programs (*P*=.008 and *P*=.75), respectively. Additional notable deficits for vascular surgery websites were resident profiles, sample contracts, and departmental research interests. Thoracic surgery websites lacked program director contact information and message as well as information on operative experience. Interventional radiology websites had deficits in work hours, debt management, and curriculum. All specialty websites had deficits in evaluation criteria and faculty mentorship. In addition to addressing the deficits in program recruitment and education content items, the deficits in lifestyle management cannot be disregarded; medical students increasingly report controllable lifestyle as a major factor in specialty choice [34,35].

The deficits identified by this analysis are comparable to deficits identified for other specialties. Other studies have found considerable deficits in newsletter, resident listings and photographs, faculty contact information, and away elective rotation information for dermatology websites [3]; resident call schedule, alumni career placement, and salary for orthopedic surgery websites [7]; academic conference schedule, call schedule, operative case listing, graduate fellowship information, and board exam performance for plastic surgery websites [19]; evaluation criteria, call schedule, operative exposure, national meetings attended, debt management, alumni contact, and work hours for neurosurgery websites [20]; call schedule, away elective rotation information, resident profiles, and faculty research for general surgery websites [21]; and call schedule, active/past research projects, area information, message from the program director or chair, selection criteria, salary, and surgical statistics for otolaryngology websites [22].

Overall, we recommend that programs address the deficits in specific content items identified by this analysis. Given the increasingly important role of online information in the residency application process and the anticipated transition to a virtual application process for the 2021 cycle, it would behoove programs to increase their online presence. In addition to the content items included in this analysis, it might be fitting for programs to include more personal information (ie, more detailed resident and attending profiles) to give applicants a better idea of the personality of different programs, replacing the role previously served by in-person away rotations and interviews.

This study had several limitations. First, this data is representative of what information was available online at the time of data collection. It is possible that websites could have been edited or new program websites could have been published since that time. Additionally, though an extensive list of content items were evaluated by reviewers regarding program recruitment and education, it is possible that other unmentioned content items could hold bearing on an applicant's decision. Finally, reviewers were not blinded to what program they were evaluating. Thus, any inherent bias reviewers might have had for particular programs was not controlled for. The nature of this study did not lend itself to evaluating the association between website content, to what specialty and to what programs applicants apply, and ultimate applicant program placement. Future studies could seek to characterize this trajectory.

Figure 1. Map of integrated vascular surgery (orange), thoracic surgery (black), and interventional radiology (silver) training programs.



Conclusion

This study has uncovered key differences in the availability of online content for residency programs recruiting from the same pool of applicants. Thoracic surgery program webpages have the most information, then vascular surgery program webpages, with interventional radiology program webpages reporting the least content. Recruitment is highly competitive between vascular surgery, thoracic surgery, and interventional radiology with the ratio of applicants to positions being highest for thoracic surgery, then interventional radiology, and lastly vascular surgery, as reported by ERAS. To attract valuable candidates, programs should aim to increase the availability of online content for potential applicants.

Acknowledgments

This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

The research was presented at the Society for Vascular Surgery Virtual in Toronto, ON in June 2020.

Authors' Contributions

KJ, QY, and MGD were responsible for study planning. KJ, QY, and MGD were responsible for data collection. KJ, QY, and MGD were responsible for statistical analysis. All authors were responsible for data interpretation. KJ, QY, and MGD were responsible for manuscript drafting, and all authors were responsible for critical revision and overall content.

Conflicts of Interest

None declared.

References

1. Powell R, Brown K, Davies M, Hart J, Hsu J, Johnson B, SVS Valuation Work Group. The value of the modern vascular surgeon to the health care system: a report from the Society for Vascular Surgery Valuation Work Group. J Vasc Surg 2021 Feb;73(2):359-371.e3. [doi: 10.1016/j.jvs.2020.05.056] [Medline: 32585182]
2. Veith FJ. A look at the future of vascular surgery. J Vasc Surg 2016 Oct;64(4):885-890 [FREE Full text] [doi: 10.1016/j.jvs.2016.07.096] [Medline: 27666442]
3. Ashack KA, Burton KA, Soh JM, Lanoue J, Boyd AH, Milford EE, et al. Evaluating dermatology residency program websites. Dermatol Online J 2016 Mar 16;22(3):13030/qt7rx3j2dn [FREE Full text] [Medline: 27136628]
4. Oladeji LO, Yu JC, Oladeji AK, Ponce BA. How useful are orthopedic surgery residency web pages? J Surg Educ 2015;72(6):1185-1189. [doi: 10.1016/j.jsurg.2015.05.012] [Medline: 26089163]
5. Rozental TD, Lonner JH, Parekh SG. The Internet as a communication tool for academic orthopaedic surgery departments in the United States. J Bone Joint Surg Am 2001 Jul;83(7):987-991. [doi: 10.2106/00004623-200107000-00002] [Medline: 11451966]
6. Winters RC, Hendey GW. Do web sites catch residency applicants? Acad Emerg Med 1999 Sep;6(9):968-972 [FREE Full text] [doi: 10.1111/j.1553-2712.1999.tb01252.x] [Medline: 10490265]
7. Davidson AR, Loftis CM, Throckmorton TW, Kelly DM. Accessibility and availability of online information for orthopedic surgery residency programs. Iowa Orthop J 2016;36:31-36 [FREE Full text] [Medline: 27528832]
8. Davidson AR, Murphy RF, Spence DD, Kelly DM, Warner WC, Sawyer JR. Accessibility and quality of online information for pediatric orthopaedic surgery fellowships. J Pediatr Orthop 2014 Dec;34(8):831-834. [doi: 10.1097/BPO.0000000000000217] [Medline: 24840658]
9. Hashmi A, Policherla R, Campbell H, Khan FA, Schumaier A, Al-Mufarrej F. How informative are the plastic surgery residency websites to prospective applicants? J Surg Educ 2017;74(1):74-78. [doi: 10.1016/j.jsurg.2016.08.002] [Medline: 27717704]
10. Hu J, Zhen M, Olteanu C, Avram R. An evaluation of accessibility and content of microsurgery fellowship websites. Plast Surg (Oakv) 2016;24(3):187-190 [FREE Full text] [doi: 10.4172/plastic-surgery.1000974] [Medline: 28439508]
11. Homer N, Yoon MK. Evaluation of the American Society of Ophthalmic Plastic and Reconstructive Surgery (ASOPRS) fellowship program website content and quality. Ophthalmic Plast Reconstr Surg 2017;33(6):471-473. [doi: 10.1097/IOP.0000000000000830] [Medline: 27997464]
12. Miller VM, Padilla LA, Schuh A, Mauchley D, Cleveland D, Aburjania Z, et al. Evaluation of cardiothoracic surgery residency and fellowship program websites. J Surg Res 2020 Feb;246:200-206. [doi: 10.1016/j.jss.2019.09.009] [Medline: 31604181]
13. Mahler SA, Wagner M, Church A, Sokolosky M, Cline DM. Importance of residency program web sites to emergency medicine applicants. J Emerg Med 2009 Jan;36(1):83-88. [doi: 10.1016/j.jemermed.2007.10.055] [Medline: 18439790]
14. Mulcahey MK, Gosselin MM, Fadale PD. Evaluation of the content and accessibility of web sites for accredited orthopaedic sports medicine fellowships. J Bone Joint Surg Am 2013 Jun 19;95(12):e85. [doi: 10.2106/JBJS.L.00785] [Medline: 23783216]
15. Novin S, Yi PH, Vanderplas T, Yim D, Hong K. Integrated interventional radiology residency program websites: a development in progress. AJR Am J Roentgenol 2018 Jul;211(1):211-216. [doi: 10.2214/AJR.17.19008] [Medline: 29792738]
16. Reilly EF, Leibrandt TJ, Zonno AJ, Simpson MC, Morris JB. General surgery residency program websites: usefulness and usability for resident applicants. Curr Surg 2004;61(2):236-240. [doi: 10.1016/j.cursur.2003.10.006] [Medline: 15051272]
17. Shaath MK, Yeranosian MG, Ippolito JA, Adams MR, Sirkin MS, Reilly MC. Evaluation of the content and accessibility of web sites for accredited orthopaedic trauma surgery fellowships. J Bone Joint Surg Am 2018 May 02;100(9):e60. [doi: 10.2106/JBJS.17.01112] [Medline: 29715233]
18. Silvestre J, Guzman JZ, Abbatematteo JM, Chang B, Levin LS. Evaluation of content and accessibility of hand fellowship websites. Hand (N Y) 2015 Sep;10(3):516-521 [FREE Full text] [doi: 10.1007/s11552-014-9732-9] [Medline: 26330788]
19. Silvestre J, Tomlinson-Hansen S, Fosnot J, Taylor JA. Plastic surgery residency websites: a critical analysis of accessibility and content. Ann Plast Surg 2014 Mar;72(3):265-269. [doi: 10.1097/SAP.0000000000000125] [Medline: 24509136]
20. Skovrlj B, Silvestre J, Ibeh C, Abbatematteo JM, Mocco J. Neurosurgery residency websites: a critical evaluation. World Neurosurg 2015 Sep;84(3):727-733. [doi: 10.1016/j.wneu.2015.04.051] [Medline: 25940210]
21. Stoeger SM, Freeman H, Bitter B, Helmer SD, Reyes J, Vincent KB. Evaluation of general surgery residency program websites. Am J Surg 2019 Apr;217(4):794-799. [doi: 10.1016/j.amjsurg.2018.12.060] [Medline: 30638725]
22. Svider PF, Gupta A, Johnson AP, Zuliani G, Shkoukani MA, Eloy JA, et al. Evaluation of otolaryngology residency program websites. JAMA Otolaryngol Head Neck Surg 2014 Oct;140(10):956-960. [doi: 10.1001/jamaoto.2014.1714] [Medline: 25188904]

23. Young BL, Oladeji LO, Cichos K, Ponce B. Content and accessibility of shoulder and elbow fellowship web sites in the United States. Iowa Orthop J 2016;36:36-40 [FREE Full text] [Medline: 27528833]
24. Cantrell CK, Dahlgren NJ, Young BL, Hendershot KM. Evaluation of accredited trauma, surgical critical care, and acute care surgery fellowship websites. J Surg Res 2018 Dec;232:160-163. [doi: 10.1016/j.jss.2018.05.072] [Medline: 30463713]
25. Ahmed SA, Hyman C, Eltorai AEM, Ahn SH. Evaluation of integrated interventional radiology residency websites. R I Med J (2013) 2019 Aug 01;102(6):19-23. [Medline: 31398963]
26. Charalel RA, Pua BB, Galla N, Trehan SK, Madoff DC. Interventional radiology fellowship website content: what is the relevance to potential applicants? Clin Imaging 2016;40(6):1070-1074. [doi: 10.1016/j.clinimag.2016.06.011] [Medline: 27421081]
27. Huang BY, Hicks TD, Haidar GM, Pounds LL, Davies MG. An evaluation of the availability, accessibility, and quality of online content of vascular surgery training program websites for residency and fellowship applicants. J Vasc Surg 2017 Dec;66(6):1892-1901 [FREE Full text] [doi: 10.1016/j.jvs.2017.08.064] [Medline: 29169545]
28. Results and data: 2020 Main Residency Match®. National Resident Matching Program. 2020. URL: https://www.nrmp.org/wp-content/uploads/2020/06/MM_Results_and-Data_2020-1.pdf [accessed 2022-06-08]
29. Lebastchi AH, Tackett JJ, Argenziano M, Calhoon JH, Gasparri MG, Halkos ME, et al. First nationwide survey of US integrated 6-year cardiothoracic surgical residency program directors. J Thorac Cardiovasc Surg 2014 Aug;148(2):408-15.e1 [FREE Full text] [doi: 10.1016/j.jtcvs.2014.04.004] [Medline: 24820188]
30. Marx MV. Evolution of interventional radiology and ABR certification: birth of the ABR certificate in interventional radiology and diagnostic radiology. American Board of Radiology. 2020. URL: https://www.theabr.org/wp-content/uploads/2017/09/IRDR-history.pdf [accessed 2020-12-31]
31. Lee JT, Teshome M, de Virgilio C, Ishaque B, Qiu M, Dalman RL. A survey of demographics, motivations, and backgrounds among applicants to the integrated 0 + 5 vascular surgery residency. J Vasc Surg 2010 Feb;51(2):496-502; discussion 502 [FREE Full text] [doi: 10.1016/j.jvs.2009.08.076] [Medline: 20022205]
32. Calligaro KD, Dougherty MJ. Choice of vascular surgery as a specialty: what is important? Semin Vasc Surg 2006 Dec;19(4):180-186. [doi: 10.1053/j.semvascsurg.2006.08.004] [Medline: 17178319]
33. Calligaro KD, Dougherty MJ, Sidawy AN, Cronenwett JL. Choice of vascular surgery as a specialty: survey of vascular surgery residents, general surgery chief residents, and medical students at hospitals with vascular surgery training programs. J Vasc Surg 2004 Nov;40(5):978-984. [doi: 10.1016/j.jvs.2004.08.036] [Medline: 15557914]
34. Resneck JS. The influence of controllable lifestyle on medical student specialty choice. Virtual Mentor 2006 Aug 01;8(8):529-532 [FREE Full text] [doi: 10.1001/virtualmentor.2006.8.8.msoc1-0608] [Medline: 23234693]
35. Dorsey ER, Jarjoura D, Rutecki GW. Influence of controllable lifestyle on recent trends in specialty choice by US medical students. JAMA 2003 Sep 03;290(9):1173-1178. [doi: 10.1001/jama.290.9.1173] [Medline: 12952999]

Abbreviations

ACGME: Accreditation Council for Graduate Medical Education
ERAS: Electronic Residency Application Service
NRMP: National Resident Matching Program
RPVI: Registered Physician in Vascular Interpretation

Edited by N Zary; submitted 19.11.21; peer-reviewed by K Ashack, E Mohammadi; comments to author 16.03.22; revised version received 08.05.22; accepted 27.05.22; published 29.06.22.

Please cite as:
Jensen K, Yan Q, Davies MG
Critical Comparison of the Quality and Content of Integrated Vascular Surgery, Thoracic Surgery, and Interventional Radiology Residency Training Program Websites: Qualitative Study
JMIR Med Educ 2022;8(2):e35074
URL: https://mededu.jmir.org/2022/2/e35074
doi:10.2196/35074
PMID:35767342

Original Paper

Design and First Impressions of a Small Private Online Course in Clinical Workplace Learning: Questionnaire and Interview Study

Esther C Hamoen[1], MD; Peter G M De Jong[2], PhD; Floris M Van Blankenstein[2], PhD; Marlies E J Reinders[1,3], Prof Dr

[1]Department of Internal Medicine, Leiden University Medical Center, Leiden, Netherlands

[2]Center for Innovation in Medical Education, Leiden University Medical Center, Leiden, Netherlands

[3]Nephrology and Transplantation, Internal Medicine, Erasmus Medical Center Transplantation Institute, Erasmus Medical Center, Rotterdam, Netherlands

Corresponding Author:
Esther C Hamoen, MD
Department of Internal Medicine
Leiden University Medical Center
Albinusdreef 2
Leiden, 2333 ZA
Netherlands
Phone: 31 715298777
Email: e.c.hamoen@lumc.nl

Abstract

Background: Clinical workplace learning takes place in a dynamic and complex learning environment that is designated as a site for patient care and education. Challenges in clinical training can be overcome by implementing blended learning, as it offers flexible learning programs suitable for student-centered learning, web-based collaboration, and peer learning.

Objective: The aim of this study is to evaluate the Small Private Online Course (SPOC) by interns' first impressions and satisfaction measures (N=20) on using the SPOC. This study describes the design process of a SPOC from a theoretical and practical perspective and how it has been integrated into a clinical internship in internal medicine.

Methods: The design of the SPOC was based on general theoretical principles that learning should be constructive, contextual, collaborative, and self-regulated, and the self-determination theory to stimulate intrinsic motivation. Interns' impressions and level of satisfaction were evaluated with a web-based questionnaire and group interview.

Results: Interns thought the web-based learning environment to be a useful and accessible alternative to improve knowledge and skills. Peer learning and web-based collaboration through peer interaction was perceived as less effective, as student feedback was felt inferior to teacher feedback. The interns would prefer more flexibility within the course, which could improve self-regulated learning and autonomy.

Conclusions: The evaluation shows that the SPOC is a useful and accessible addition to the clinical learning environment, providing an alternative opportunity to improve knowledge and skills. Further research is needed to improve web-based collaboration and interaction in our course.

(JMIR Med Educ 2022;8(2):e29624) doi:10.2196/29624

KEYWORDS

blended learning; design-based research; web-based learning; workplace learning; medical education; clinical internship

Introduction

Blended Clinical Workplace Learning

Clinical workplace learning (WPL) mostly takes place during normal daily collaboration and patient care activities, or organized formal learning activities [1,2]. It happens in a complex learning environment that is known to face many challenges. Patient cases tend to increase in complexity, whereas educational exposure is often insufficient, and time pressure leads to insufficient observation and assessment of the learner and suboptimal support within the diagnostic process [3-6]. Another challenge is the lack of sustained relationships among students, teachers, and patients [7,8]. These challenges, among others, lead to suboptimal clinical training.

Blended learning can be used to remedy several of these problems. Blended learning refers to a deliberate blending of face-to-face and web-based learning, with the goal of stimulating and supporting learning [9]. When thoughtfully designed, blended learning can improve education [10]. It may shift education to a more active and learner-centered approach, where the learner is *in control*, and may better fit the needs of different learning styles that students might have [11-13]. Overall, blended learning is more effective than traditional learning and, when carefully designed, has been demonstrated to have better effects on knowledge outcomes, learner motivation, and satisfaction than traditional face-to-face learning [10,14-17].

The web-based component of blended learning permits flexible education at a time, place, and pace convenient for the learner [18]. It can also help learners to share knowledge and experiences through web-based discussion forums and collaborative assignments with others although geographically dispersed. As web-based learning is complementary to instructor-led training, it can best be integrated in a blended learning curriculum [18,19]. The web-based component of blended learning in medical education can help students develop clinical reasoning skills by adding web-based patient experiences to real-world patient exposure.

Small Private Online Courses

A Small Private Online Course (SPOC) is one possible instrument to blend web-based learning with clinical WPL. The SPOC concept was first introduced in 2013, and it has been progressively implemented in higher education thereafter. This type of course is often used locally with on-campus students and has a limited number of students that can enroll in the course [20]. Previous reports have shown that SPOCs can be feasible and suitable environments for student learning and fulfill students' need for social interaction [19,21-23]. In medical education, SPOCs can positively impact professional practice and are thought to improve the management of patients [21,24]. It was shown that SPOCs need a flexible program and supportive environment to make them work [18]. SPOCs are relatively new in clinical WPL, and much is still unknown about how SPOCs can be optimally developed and integrated in clinical WPL. This information is required to improve the deployment of such blended programs in clinical training and in the end to improve the training of our future physicians.

The development of a dedicated SPOC instead of using publicly available course materials has the advantage that several secondary conditions such as contents definition, availability to interns, the alignment of goals, desired teaching modes, and assessments can be addressed by design. This avoids many of the current challenges with using open web-based education from others as described by de Jong et al [25] and Hendriks et al [26,27], such as limited constructive learning and a lack of certain desired teaching modes.

Background and Objective

In 2017, we developed a SPOC for the Internal Medicine internship at Leiden University Medical Center (LUMC) in the Netherlands. The theoretical framework of the course is based on the self-determination theory (SDT) [28] and the general learning principles that learning should be constructive, contextual, collaborative, and self-regulated [29]. SDT offers a framework for driving intrinsic student motivation by stimulating autonomy, competence, and relatedness [28,30]. In the SPOC, groups of interns work on authentic clinical scenarios, use resources, and discuss with peers and teachers on a forum. The SPOC has been fully integrated into WPL; this means that knowledge and skills that are trained on the web can be directly transferred to the clinical environment where the interns have practical training and vice versa. In this paper, we evaluate the final design of the SPOC from a theoretical perspective. We report on the perceptions of interns on using this web-based resource in clinical WPL and their level of satisfaction. With the results, we hope to gain insight in the added value of introducing the SPOC in the WPL environment of our clinical internship.

Methods

Context

In the Netherlands, medicine students enter medical school at the bachelor level, which is followed by internships at the master level. At LUMC, each month approximately 20 interns start their clinical internships. In the first 4 weeks, the interns attend a joint program at the university. The so-called introductory internship (2 weeks) prepares them for the internships in general and is being followed by a specific 2-week theoretical course as preparation for the Internal Medicine internship. Thereafter, they start their clinical internship in Internal Medicine (12 weeks), in which they work in different affiliated hospitals in the region and cannot meet each other physically. A major aim during the internship is to obtain clinical reasoning skills for a broad range of clinical scenarios. However, the interns only have limited exposure to new patients who have not yet been diagnosed by other physicians, and the clinical scenarios that interns face in practice do not cover all the clinical scenarios that they need to know and understand. Collaboration between peers is limited because the group of interns is split up to have their internship in different hospitals.

Design of the SPOC

Overview

To overcome several of the limitations mentioned, a SPOC has been developed. The course was designed using a design-based research approach in which practical and theoretical aspects are integrated in the educational design [29,31]. Attending the SPOC is facultative but highly recommended. However, once an intern decides to participate, several of the learning activities are obligatory to complete the lessons.

Practical Aspects

For the development of the course, a group of stakeholders has been identified including clinical interns, clinical teachers, educational experts, technical experts, and a graphic designer. During a dedicated learning experience design (LED) session the stakeholders set a framework for the SPOC, including the team and course's goals (eg, improve patient expose), students' needs (eg, track progression within the course), learning goals (eg, improve clinical reasoning skills), and the aimed look and feel (eg, authenticity of the cases).

Description of the SPOC

The outcomes of the LED session were used to define the exact content and learning activities centered around authentic clinical problems that are typically encountered in internal medicine. NEO Learning Management System was used as a platform for the SPOC. Authentic cases in Dutch were developed to train clinical reasoning skills and medical knowledge. The SPOC has a modular design that involves preparing for the internships (2 weeks), internal medicine (2 weeks), and several inpatient, outpatient, and emergency room cases (12 weeks). Every week, the interns can study 1 new activity. A course overview of the 12 clinical weeks is shown in Figure 1. The lessons consist of various obligatory and optional assignments and learning activities including simulated patient cases, virtual reality applications (a virtual reality ward experience and professional or unprofessional behavior experience), group assignments, videos, e-learnings, e-readings, assignments, web-based exams, discussion forums, and peer feedback sessions. Exemplary screenshots of the intervention can be found in Figure 2. New knowledge can be directly applied to the clinical workplace where the interns work. Interns' progression is tracked in the course, and it offers access to resources they can use in the workplace.

Figure 1. Small Private Online Course (SPOC) layout. Assessments during the 12 weeks of the clinical internship are displayed only (lessons presented in preparting for the internships and internal medicine are not displayed). The lessons include several inpatient, outpatient, and emergency room cases and assignments related to those participants.

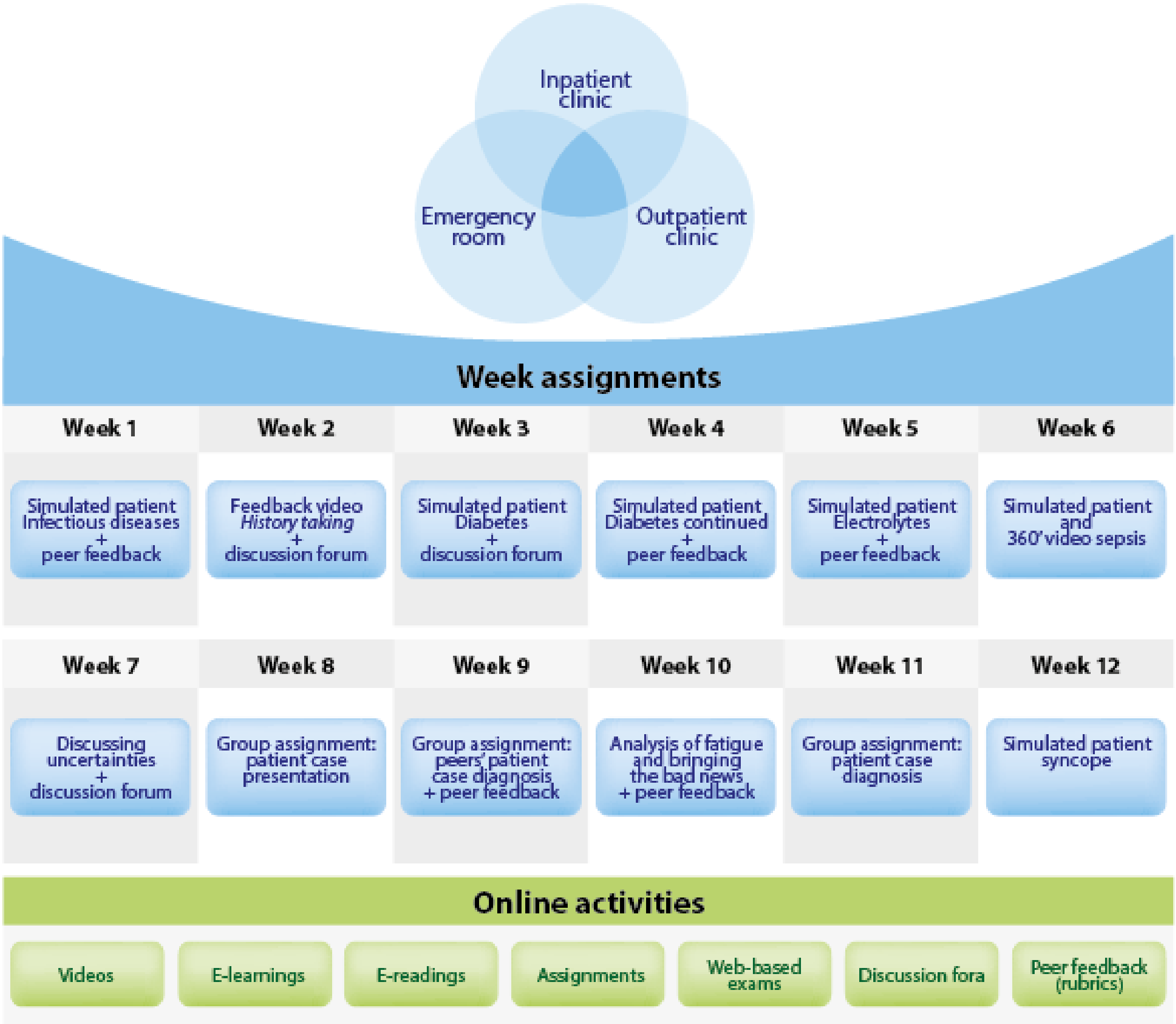
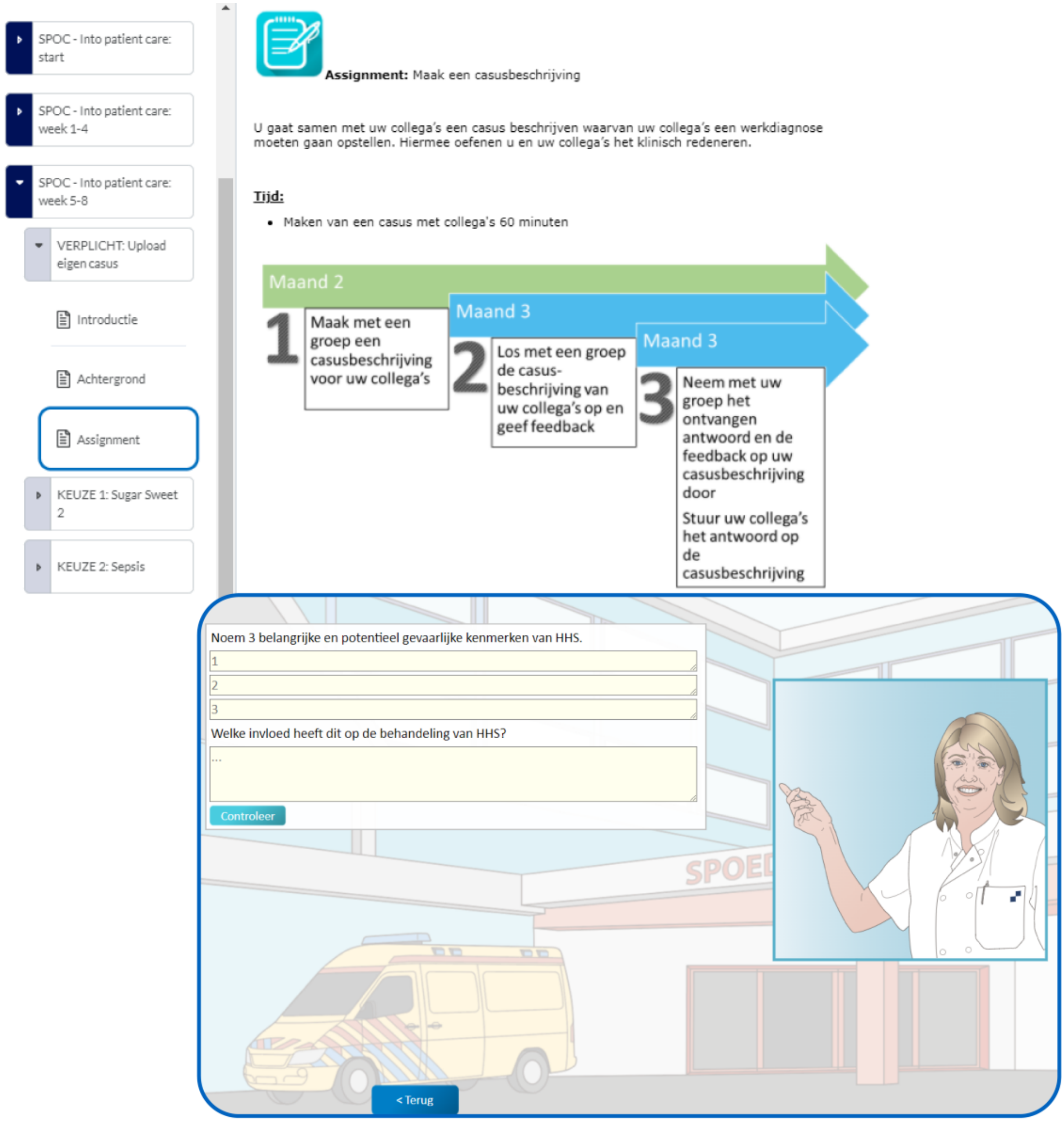


Figure 2. Exemplary screenshots of exercises in the Small Private Online Courses (SPOCs) as they are offered to the students.



Theoretical Aspects

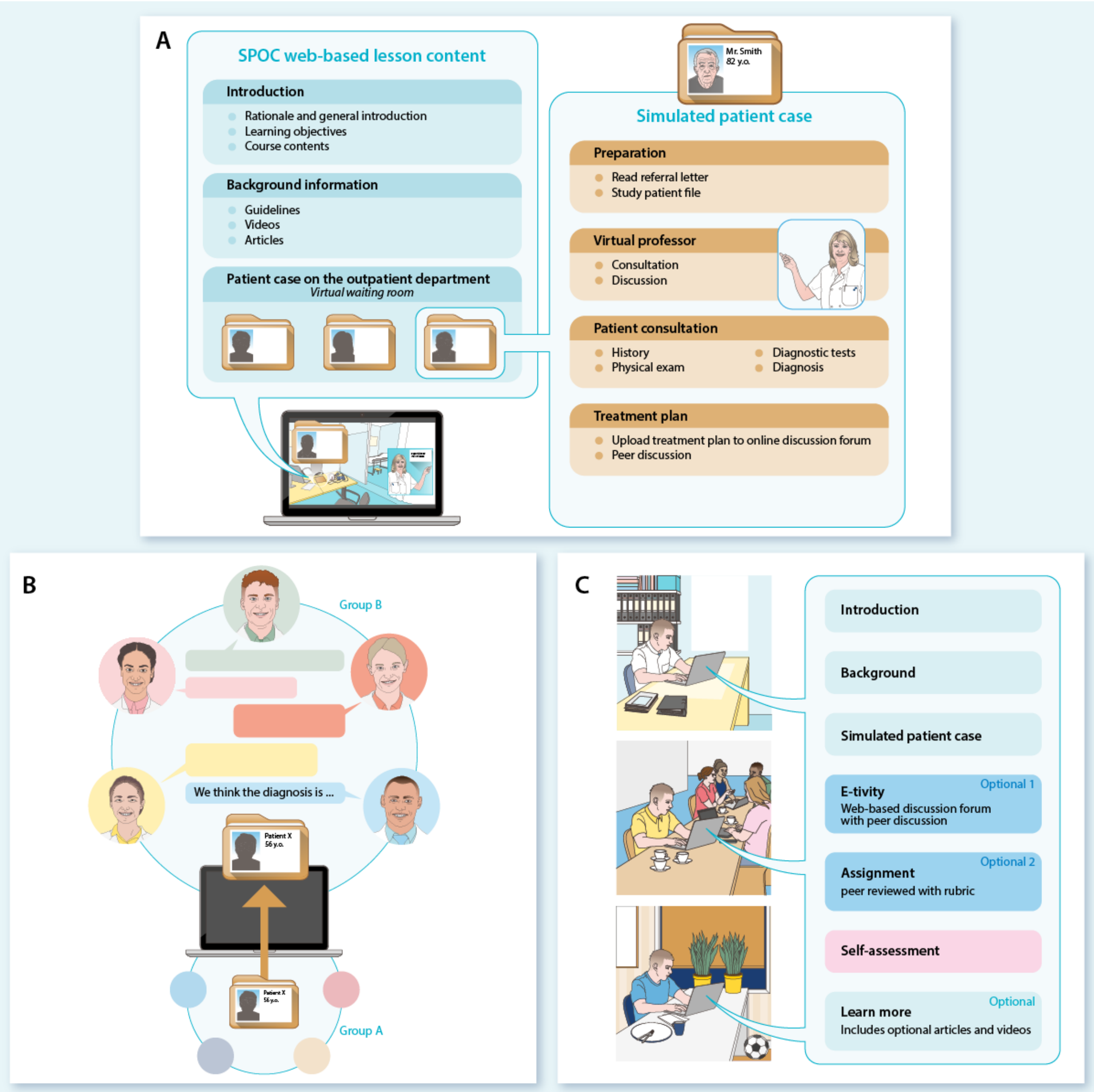
On the basis of the SDT and the general learning principles that learning should be constructive, contextual, collaborative, and self-regulated, authentic patient cases that are centered around virtual reality patients with complaints of diabetes, electrolyte disorders, infectious diseases, oncological diseases, and tiredness in a setting of the inpatient clinic, outpatient clinic, or emergency room have been developed. A 3D virtual reality patient ward has been developed, giving the interns a realistic impression of the inpatient setting. It aims to introduce interns to the ward and ward rounds and train professional standards such as hygiene regulations. Constructive learning activities such as doing rounds on 3D virtual reality patients with increasing complexity are used to gather knowledge and to improve learning effectiveness in the workplace.

Relatedness is promoted in the SPOC, as the interns (who work in different hospitals and do not have contact) meet up on the web on discussion forums, during peer feedback sessions and group assignments. This stimulates collaboration and relatedness among interns who would normally not meet each other during the internship. Collaborative forum assignments are included for peer learning and direct feedback by both peers and dedicated clinical teachers. Peer feedback is either given through open peer discussion on a forum or using a rubric format. Peer discussion, in this case on the web, might help interns to develop their critical thinking and clinical reasoning skills. Interns are prepared on their role as assessors by a web-based peer feedback training.

Besides obligatory content, the SPOC contains several optional assignments and resources that interns can choose from. The roster during the internship is different for each intern. An intern can spend 2 weeks in the cardiology department, whereas another intern works in the emergency room. Therefore, the interns can choose several sequential assignments for more extensive learning, depending on the training they need at that specific moment. By stimulating autonomy and attention to precourse goal setting and tracking and ranking course activity, it is aimed that interns can control their own learning process.

Training and assessment of competencies occur through peer feedback and self-assessment. Interns that have self-assessed their competencies and know that they are on the right track feel more confident while carrying out their new skills on real patients. In this blended program, skills that are trained on the web can be directly transferred to clinical practice. Practical examples of the integration of theory in the SPOC are shown in Figure 3.

Figure 3. Examples of theoretical integration in the Small Private Online Course (SPOC). (A) Constructive learning in the clinical context. The figure illustrates how interns go through an authentic simulated patient case in the SPOC. The learning objectives are defined for each learning activity separately. For example, in one of the diabetes cases, the learning objectives are defined as follows. By the end of this lesson, the intern (1) is able to distinguish between type 1 and type 2 diabetes based on epidemiology, history, physical examination, and additional tests; (2) knows the complications of diabetes mellitus and knows the screening protocols; (3) knows the general treatment modalities for diabetes mellitus and cardiovascular risk management; and (4) knows the chain of care for patients with diabetes and knows the physician’s role within the chain. (B) Collaborative learning and relatedness. This figure displays a group assignment in the SPOC. A clinical scenario is described by one group of interns (group A), and another group of interns (group B) elaborates their clinical reasoning process that is completed by a diagnosis. Thereafter, feedback is provided by both groups on the quality of the case and the diagnosis, respectively. (C) Self-regulated learning or autonomy. The courses’ lessons contain required and facultative issues for further learning.



Participants

Participants were a first group of 20 interns who worked with the SPOC during a pilot period. They enrolled during their introductory internship and remained in the SPOC until completion of their Internal Medicine internship. The participants were in the first year of their clinical phase. In this phase, interns are aged approximately 21 to 25 years, and, on average, 60% to 70% are women. As the aim is to evaluate the impressions of the first group, no sample size calculation has been performed.

Instruments

The evaluation existed of a web-based questionnaire (Multimedia Appendix 1) and an interview. The questionnaire consisted of 15 Likert-scale questions and 5 open-ended questions. The questionnaire contained questions that were mostly based investigation of the SPOC's design; for example, what are the interns' impressions of the aspects of the LED session, the SDT, and learning principles that were used in the SPOC's design? It contained questions about the amount of time invested in the SPOC and whether the interns thought the SPOC to be useful, informative, and motivating. Interns were also asked whether they experienced more patient exposure while working in the SPOC and if they thought the cases were authentic. The questionnaire also contained questions about perceived competence, relatedness, and autonomy in the course. The educational context was quite unique, and therefore, there was no existing validated instrument available to measure the intended outcomes.

Ethics Approval

As the study did not involve patients and no health intervention has been administered to participants, the study was not subjected to the Dutch WMO (Medical Research involving Human Subjects Act). The study has been conducted in compliance with the European Union General Data Protection Regulation 2016/679, and data have been anonymized and stored according to the *Nederlandse Gedragscode Wetenschapsbeoefening* of the Universiteiten van Nederland (Association of Universities The Netherlands). Institutional educational review board approval was obtained under reference OEC/ERRB/20220208/1. All interns provided written informed consent to participate in the study. In the information letter, the full study procedure was explained, as well as the option for the interns to opt out of the study at any moment without any reason.

Procedure

The students completed the weekly assignments and remained in the SPOC until they had completed their Internal Medicine internship. Subsequently, they were asked to fill out the web-based questionnaire. A group interview was led by an independent interviewer who had no intern-teaching relationship with the interns. The interview used the *snowball method*: the participants first individually recalled their own experiences with the course, and then 2 participants paired up to discuss their experiences and wrote down their most important findings, which was repeated in groups of 4 participants. All the experiences were shared with the group by oral presentation. The interviews were audio recorded and converted into a transcript by the first author (ECH).

Data Analysis

Questionnaire

Means and SDs were calculated for the Likert-scale questions in the web-based questionnaire. The answers to the 5 open-ended questions were summarized.

Interview

The first and third authors discussed the transcript coding template, which was based on the different items in the theoretical framework, until consensus was obtained. The outcome was a template consisting of six predefined, overarching themes: (1) contextual learning, (2) collaboration and relatedness, (3) constructive learning, (4) self-regulated learning or autonomy, (5) competence, and (6) other. For each category, the same authors agreed on a definition for each theme. The principal investigator (ECH) analyzed the interview data and clustered the answers in the template using a Microsoft Word.

Results

Collected Data

The aim of this study is to investigate the perceptions of interns concerning the use of the SPOC using a questionnaire and interview. Of the 20 interns eligible to enroll in the SPOC, 19 (95%) actually enrolled in the course, 10 (50%) finished the whole course, and 1 (5%) never started. Ten interns filled out the web-based questionnaire. Only questionnaires that were fully completed were included in the analysis. All 20 interns attended the group interview.

Questionnaire: Likert-Scale Questions

The results are shown in Table 1. The SPOC was valued as being informative and useful to most of the interns, and they felt that the patient cases were authentic. However, the interaction with peers was found inadequate and not useful. Interns' perceptions on motivation to learn in the SPOC was not optimal.

Table 1. Outcomes of the web-based questionnaire (n=10).

Question	Value, mean (SD)[a]
1. Time per week spent in the course (hours)	1.8 (0.92)[b]
2. Following the course is useful during the internship	4.5 (1.96)
3. The course was informative	5 (1.33)
4. The SPOC[c] was a motivation for learning	3 (1.49)
5. I would recommend the course to peer students	3.9 (1.91)
6. The knowledge obtained from the SPOC was fairly applicable to clinical practice	4 (1.89)
7. Through working in the SPOC, my patient exposure has been increased	3.3 (1.64)
8. The cases in the SPOC were authentic	4.8 (1.23)
9. Making the cases, I really felt like a physician making decisions	3.6 (2.01)
10. I felt more competent in clinical reasoning after finalization of the SPOC	4.1 (1.79)
11. The SPOC had a good construction of increasing difficulty	4.7 (1.06)
12. I could organize my own time well within the SPOC	4.1 (1.79)
13. In the course, I had good interaction with peers	2.4 (1.71)
14. The interaction with peers was useful	1.8 (1.03)
15. The assignments and tests were challenging	4.3 (1.70)

[a]The 7-point scale ranges as follows: (1) totally disagree to (4) neither disagree nor agree to (7) totally agree.

[b]Note that the number of of hours spent per week is displayed in this row.

[c]SPOC: Small Private Online Course.

Questionnaire: Open-ended Questions

The interns highlighted the patient cases as a positive aspect of the course. They particularly valued their connection with clinical practice and the elaboration and variety of relevant cases. Furthermore, the interns appreciated the group assignment in which they *solve* a patient case by finding the diagnosis. Interns liked the graphical layout of the course. The SPOC supported training clinical reasoning skills, although some interns felt that those skills are better trained in practice or when the specific SPOC cases were also experienced in clinical practice. Concerning peer feedback, the interaction during the patient cases was thought to be the most useful, as were the feedback training and having insight in the answers of peers. Some interns indicated that the technology to give feedback did not always work well or that interaction in clinical practice was more useful. The information about deadlines, the instruction for how to give peer feedback, the quality of the feedback received from peers, and the fixed order of the assignments were mentioned as limitations of the course. Owing to time constraints, some interns indicated they had difficulties finalizing the mandatory contents within the lessons. They would have preferred optional content only.

Interview

The interns were interviewed in their last week of the internship. Table 2 shows a more detailed overview of the results of the group interview. Constructive and collaborative learning were clustered under 1 category because the interview data overlapped in both categories.

The interns experienced the SPOC to be accessible and adequately designed for them as a target group. The participants felt the patient cases and SPOC content were useful and informative. However, although authenticity was integrated in the design of the SPOC, the participants felt that the contents did not match the real world. It seemed unclear to them how the SPOC should complement clinical WPL. They would prefer assignments that matched the clinical problems they encountered at that time instead of fixed weekly assignments.

Development of critical thinking skills by peer discussion was also integrated in the SPOC's design; however, the interns indicated that they preferred model answers over peer discussion. The participants also felt it was not really useful to receive feedback from peers instead of a teacher, because in their opinion peers know as much as they do themselves. In general, feedback was perceived to be very short.

The interns indicated they had enough time to finish the assignments during their normal day shifts, although not everyone agreed. They also appreciated the facultative character of the SPOC. They felt that more facultative assignments would be helpful, although they addressed the possibility that nobody would mind finishing them. They also indicated that they needed more flexibility and choice in the course, and the fixed order of the assignments did not work for everyone. Some participants felt they had "finally reached the clinical phase of their internship" and therefore did not appreciate completing web-based assignments in this stage of the curriculum. They also preferred a complete overview of a certain clinical presentation instead of looking up information themselves through links in the SPOC.

Table 2. Group interview concerning SPOC[a] perceptions[b].

Theme	Positive	Negative
Contextual learning	• SPOC was adequately targeting interns • Useful patient cases • Content was handy and informative	• Assignments did not match clinical problems encountered in WPL[c] • Learning objectives were unclear • Unclear how SPOC complements WPL • Contents did not match the real world
Collaboration, relatedness, and constructive learning	—[d]	• Peer feedback is less useful than feedback from teacher • Interaction • Interns wanted answer sheets instead of peer discussion
Self-regulated learning or autonomy	• Sufficient time to finish assignments during daily shift • Nice that SPOC is not obligatory	• Tight deadlines • Time-consuming • Insufficient time to finish assignments during daily shift • Need more choice instead of fixed assignments • Do not want theoretical assignments during the practical phase • Less obligatory and more optional assignments • Want a complete overview of a clinical presentation, instead of looking up information • Wished the SPOC to be more motivating
Competence	• Good patient cases, e-learnings, and quizzes • Useful lessons, mainly virtual reality patients • Useful for interns with less patient contacts and less moments for clinical reasoning • Enjoyed videos observing others taking history • History taking videos led to discussion among peers • Those videos must be part of the training	• SPOC did not fill gaps encountered in clinical practice • Watching videos seeing others taking history is not active learning • Learned more from observation in workplace • Need more specific physical exam tools instead of observation general physical examination • Need more assignments that specifically enhance knowledge
Other	• SPOC was accessible • Good learning environment • Videos were enjoyable	• Unclear deadlines • Problems planning patient-related assignments (suitable patients dismissed) • Technical shortcomings • Insufficient support

[a]SPOC: Small Private Online Course.

[b]The positive and negative perceptions of the interns are displayed in the middle and right columns. Those were assembled in 6 predefined clusters (left column). Collaboration and constructive learning have been clustered because of overlapping interview data.

[c]WPL: workplace learning.

[d]Not available.

The interns particularly valued the competence training received through completion of the patient cases, e-learning, and quizzes. The virtual reality patient cases and videos were rated as useful and likable. The videos concerning observation of patient interviews were enjoyable and led to useful discussion between peers. However, others felt that observation in clinical practice is more informative than watching web-based videos. Although the virtual reality patient cases can be an alternative for practicing clinical scenarios that have not been encountered in live patient contact, the interns indicated that the SPOC did not completely fill the gaps owing to the limited number of virtual reality patient cases addressed in the SPOC.

In general, the interns appreciated the accessibility of the SPOC and its learning environment, and the videos were enjoyable. They were less content about the deadlines that sometimes appeared to be unclear and the planning of SPOC activities that were directly patient related, which was challenging because of fast patient turnover rates. Completion of some of the assignments was difficult because of technical shortcomings or insufficient support.

Discussion

Principal Findings

In this study, a SPOC has been developed and implemented into an internship curriculum of Internal Medicine at LUMC. Course development was based on general learning principles and SDT, aiming to promote learner motivation and optimize the course's quality and integration in the clinical curriculum. The SPOC was evaluated by measuring the satisfaction of the interns with the course. Interns thought the SPOC to be a useful and accessible addition to the clinical learning environment. Peer learning and web-based collaboration through peer interaction were perceived as less effective, as student feedback was felt inferior to teacher feedback. The interns preferred more flexibility within the course, which could improve self-regulated learning and autonomy. Overall, the interns felt that the web-based learning environment provided an alternative opportunity to improve knowledge and skills.

Our SPOC was integrated into clinical practice by making it part of the internship program, instead of using it as a stand-alone web-based training. The advantage of this is that learners can directly apply their new knowledge in clinical practice and go back to the web-based resources for further exploration of a topic. In our opinion, using design sessions and a theoretical framework can facilitate integration of such a course into an existing curriculum and can be an effective solution for some of the complexities faced in clinical training.

From the results, we learned that the SPOC meets the expectations for some of the categories fairly well, although the overall student satisfaction seems to be modest. In general, completing the SPOC assignments during the internships seemed feasible but not for all interns. Concerning learning in the clinical context, interns appreciated the authenticity and usefulness of patient cases as an additional opportunity to improve their clinical knowledge and skills. Regarding the clinical skills development, the SPOC therefore meets our hypothesis; however, this is not the case for stimulation of the social cohesion of the interns through working in the SPOC. The quantitative data show a modest to very low outcome on several topics (score ≤3.3: motivation, interaction, and usefulness of interaction with peers and feeling like a physician making decisions).

It appeared that the interns were not satisfied with the collaboration and relatedness aspects of the SPOC, in particular the web-based interaction and peer feedback. Other studies confirm that students prefer teacher feedback over peer feedback [32-34]. When peers lack knowledge or are not critical, the peer feedback is often considered inadequate [35]. In our SPOC, peer feedback was a prespecified learning objective and was supposed to be a formative process to promote learning. It is possible that the interns perceived peer feedback merely as a mandatory assignment for completing the SPOC and that they did not understand its importance in their own learning process and that of their peers. Other studies describe that guidance on assessment and the requirements hereof, training in giving peer feedback, and clarification of the role the student takes in the feedback process are key principles of effective feedback [36,37]. Furthermore, it is important to explain the purpose of peer feedback to the students [38]. Despite the fact that our SPOC contains a peer feedback training, we might need to strengthen the guidance of the interns in their role as peer assessors. In addition, the learning objectives of the peer-assessed activities could be clearer on the purpose providing and receiving feedback.

Learning activities in a SPOC are on the web and asynchronous, with only written interaction and no visual interaction and body language [39]. This may have affected the learning experience of the interns, as interaction on cognitive, social, and teaching levels is required to promote deep learning [40]. Possible interventions that might improve interaction in these three levels are more visual teacher presence in our SPOC, synchronized learning activities to stimulate just-in-time learning, more insight into interns' learning needs, adaptivity of the teaching strategy, social cohesion, and promoting deep learning by considering feedback as a dialogical process while considering asynchronous learning in the SPOC [39,41].

Although flexibility was integrated in the design of the course based on previous literature [18], interns indicated they needed more flexibility and that deadlines were too tight. They also requested more autonomy in choosing when to complete which lessons. Self-regulated learning skills are essential in web-based learning owing to the freedom provided in web-based education. It gives learners autonomy in how they organize their learning, and they need to deal with that to be successful [42]. Our evaluation was conducted among undergraduate medical students. Previous studies have shown that self-regulation, such as experienced in e-learning, might not fit novice learners that lack the maturity and experience to reach learning outcomes that are minimally guided [43-45]. Novice students may lack the cognitive, affective, and metacognitive self-regulated learning skills necessary to effectively navigate the abundance of information that is nonlinearly provided by the hypermedia [46]. This may affect the learners' acceptability and satisfaction of e-learning activities and consequently their emotional experience. Self-regulated learning within web-based learning environments is also influenced by the emotional experience of the students [47]. Studies found a negative relation between negative emotions and learning outcomes, and the emotional experience and subsequent learning may be improved by fostering students' emotion regulation [47-49]. The technical shortcomings of the SPOC may have attributed to a negative emotional experience of our interns.

Regarding the feelings of being a physician making decisions and contextual learning, it seemed that the SPOC somehow failed in this aspect. In the LED session it was predefined to make authentic simulated patient cases to enhance this feeling and the feeling of increased patient exposure; however, this was not experienced by the interns. Although studies have shown that virtual reality patients can improve clinical reasoning skills and knowledge, it should be recognized that patient simulations are not equivalent to real patients and cannot replace traditional clinical WPL [50-52].

Reflection and Improvements

Reflecting on our study, this may imply some improvements in our SPOC's design. First of all, we should critically review the technical shortcomings, deadlines, and obligatory components of the SPOC and improve the flexibility of the learning environment. For instance, replacing the weekly assignments by a more variable set of assignments on a monthly basis could be considered. We are currently investigating the feasibility of such adaptations and the expansion of the number of optional lessons. In addition, the SPOC will soon be transferred to another internet platform, which is technically supported by our University's Information Technology department.

Second, it may be helpful to incorporate training of self-regulated learning skills before using our hypermedia environment and more support and augmentation of learners' self-regulated learning within the learning system [46,48,53]. From the interview data, we also learn that the SPOC's general instructions were valued as suboptimal. Therefore, a stronger instructional guidance and guidance of learners' self-regulated learning in our SPOC may improve the effectiveness of the learning experience and learner satisfaction [46,48,53,54]. In addition, incorporation of emotion regulation in the learning activities of such a SPOC and in the design and implementation in clinical WPL may improve self-regulated learning and maximize positive effects on students' learning within the digital learning environment [47]. These topics might be subject to future research. We believe that, as discussed, more focus on self-regulated learning, the emotional experiences within the SPOC, guidance, and the purpose of feedback may also improve the students' motivation for learning within the SPOC.

Third, several other conditions may be met to increase student satisfaction. Student satisfaction in clinical internships is enhanced by supervisor support, perceived social value [55], dedicated faculty, working in teams, and continuity in intern-patient contacts [56]. Longitudinal relationships between supervisors and students also increase student satisfaction [57,58] and students' independence [59]. Therefore, for instance, dedicated supervisors could provide feedback in the SPOC in a longitudinal working relationship with their interns. In addition, students might be encouraged to use the learning resources in the SPOC in continued contacts with patients. This may promote the SPOC's social value.

Limitations

When critically looking at the quantitative data, it is apparent that the variance is modest and particularly low for the questions concerning the interaction with peers and increasing difficulty. An explanation might be that the interns know each other fairly well and their answers may not be independent or even biased by information, sharing leading to a group opinion. Limitations of the study were the small sample size and use of an unvalidated questionnaire. As the study was designed just to evaluate the first group of interns enrolled in the SPOC, no formal sample size calculation has been performed. However, we did include qualitative data to supplement these limited quantitative outcomes. We therefore think that the impressions still provide valuable insights in the SPOC's strengths and weaknesses that we can use for further adaptation.

Conclusions

From our study, we have learned that interns perceive several learning opportunities after adding the SPOC to their clinical learning environment, mainly in skills and knowledge acquisition. However, particularly web-based collaboration and perception of relatedness among the interns within the course need further improvement. In the future, interviews with the interns may be beneficial for a deeper investigation of this issue and others, their context, and which additional adaptations might be needed to our SPOC. Future research is also needed to further investigate how learning principles can be optimally integrated in web-based courses.

Acknowledgments

This research was funded by an educational grant of Leiden University (project reference 4020522012). The authors wish to thank Franka Luk and Manon Zuurmond for their contribution to this study and illustrations, and Mandy Segers for her assistance in finalizing the manuscript revisions.

Authors' Contributions

All authors were involved in the design of the study and draft of the manuscript. ECH and FMvB analyzed and interpreted the results, and PGMDJ had major contributions to the main conclusions. All authors read and approved the final manuscript. The authors alone are responsible for the content and writing of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire.
[DOCX File , 20 KB - mededu_v8i2e29624_app1.docx]

References

1. Mertens F, de Groot E, Meijer L, Wens J, Gemma Cherry M, Deveugele M, et al. Workplace learning through collaboration in primary healthcare: a BEME realist review of what works, for whom and in what circumstances: BEME guide No. 46. Med Teach 2018;40(2):117-134. [doi: 10.1080/0142159X.2017.1390216] [Medline: 29094626]
2. Hamoen EC, van Blankenstein FM, de Jong PG, Ray A, Reinders ME. Development of a clinical teaching unit in internal medicine to promote interprofessional and multidisciplinary learning: a practical intervention. Teach Learn Med 2021;33(1):78-88. [doi: 10.1080/10401334.2020.1792309] [Medline: 32917107]

3. Daelmans HE, Hoogenboom RJ, Donker AJ, Scherpbier AJ, Stehouwer CD, van der Vleuten CP. Effectiveness of clinical rotations as a learning environment for achieving competences. Med Teach 2004;26(4):305-312. [doi: 10.1080/01421590410001683195] [Medline: 15203842]
4. Reilly BM. Inconvenient truths about effective clinical teaching. Lancet 2007;370(9588):705-711. [doi: 10.1016/S0140-6736(07)61347-6] [Medline: 17720022]
5. Dyrbye LN, Thomas MR, Shanafelt TD. Medical student distress: causes, consequences, and proposed solutions. Mayo Clin Proc 2005;80(12):1613-1622. [doi: 10.4065/80.12.1613] [Medline: 16342655]
6. Van Der Hem-Stokroos HH, Scherpbier AJ, Van Der Vleuten CP, De Vries H, Haarman HJ. How effective is a clerkship as a learning environment? Med Teach 2001;23(6):599-604. [doi: 10.1080/01421590120091000] [Medline: 12098483]
7. Irby DM. Educational continuity in clinical clerkships. N Engl J Med 2007;356(8):856-857. [doi: 10.1056/NEJMe068275] [Medline: 17314347]
8. Dorsey JK, Beason AM, Verhulst SJ. Relationships matter: enhancing trainee development with a (simple) clerkship curriculum reform. Teach Learn Med 2019;31(1):76-86. [doi: 10.1080/10401334.2018.1479264] [Medline: 30321063]
9. Boelens R, De Wever B, Voet M. Four key challenges to the design of blended learning: a systematic literature review. Educ Res Rev 2017;22:1-18. [doi: 10.1016/j.edurev.2017.06.001]
10. Spanjers IA, Könings KD, Leppink J, Verstegen DM, de Jong N, Czabanowska K, et al. The promised land of blended learning: quizzes as a moderator. Educ Res Rev 2015;15:59-74. [doi: 10.1016/j.edurev.2015.05.001]
11. Sarkar S, Sharma S, Raheja S. Implementation of blended learning approach for improving anatomy lectures of phase I MBBS students - learner satisfaction survey. Adv Med Educ Pract 2021;12:413-420 [FREE Full text] [doi: 10.2147/AMEP.S301634] [Medline: 33935527]
12. Sajid MR, Laheji AF, Abothenain F, Salam Y, AlJayar D, Obeidat A. Can blended learning and the flipped classroom improve student learning and satisfaction in Saudi Arabia? Int J Med Educ 2016;7:281-285 [FREE Full text] [doi: 10.5116/ijme.57a7.83d4] [Medline: 27591930]
13. Coyne E, Rands H, Frommolt V, Kain V, Plugge M, Mitchell M. Investigation of blended learning video resources to teach health students clinical skills: an integrative review. Nurse Educ Today 2018;63:101-107. [doi: 10.1016/j.nedt.2018.01.021] [Medline: 29425738]
14. 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;22(8):e16504 [FREE Full text] [doi: 10.2196/16504] [Medline: 32773378]
15. Lozano-Lozano M, Fernández-Lao C, Cantarero-Villanueva I, Noguerol I, Álvarez-Salvago F, Cruz-Fernández M, et al. A blended learning system to improve motivation, mood state, and satisfaction in undergraduate students: randomized controlled trial. J Med Internet Res 2020;22(5):e17101 [FREE Full text] [doi: 10.2196/17101] [Medline: 32441655]
16. Means B, Toyama Y, Murphy R, Baki M. The effectiveness of online and blended learning: a meta-analysis of the empirical literature. Teach Coll Rec 2013;115(3):1-47. [doi: 10.1177/016146811311500307]
17. Evans KH, Thompson AC, O'Brien C, Bryant M, Basaviah P, Prober C, et al. An innovative blended preclinical curriculum in clinical epidemiology and biostatistics: impact on student satisfaction and performance. Acad Med 2016;91(5):696-700. [doi: 10.1097/ACM.0000000000001085] [Medline: 26796089]
18. Pickering JD, Henningsohn L, DeRuiter MC, de Jong PG, Reinders ME. Twelve tips for developing and delivering a massive open online course in medical education. Med Teach 2017;39(7):691-696. [doi: 10.1080/0142159X.2017.1322189] [Medline: 28485629]
19. Erbe D, Eichert HC, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
20. Kaplan AM, Haenlein M. Higher education and the digital revolution: about MOOCs, SPOCs, social media, and the cookie monster. Bus Horiz 2016;59(4):441-450. [doi: 10.1016/j.bushor.2016.03.008]
21. Vaysse C, Fize AL, Despas F, Chantalat E, Serrano E, Beyne-Rauzy O, et al. Small private online course in teaching oncology-feedback after 1 year: what lessons? J Cancer Educ 2021;36(1):65-71. [doi: 10.1007/s13187-019-01598-z] [Medline: 31410821]
22. Uijl S, Filius R, Ten Cate O. Student interaction in small private online courses. Med Sci Educ 2017;27(2):237-242. [doi: 10.1007/s40670-017-0380-x]
23. Ruiz-Palmero J, Fernández-Lacorte JM, Sánchez-Rivas E, Colomo-Magaña E. The implementation of small private online courses (SPOC) as a new approach to education. Int J Educ Technol High Educ 2020;17(1):27. [doi: 10.1186/s41239-020-00206-1]
24. Vaysse C, Chantalat E, Beyne-Rauzy O, Morineau L, Despas F, Bachaud JM, et al. The impact of a small private online course as a new approach to teaching oncology: development and evaluation. JMIR Med Educ 2018;4(1):e6 [FREE Full text] [doi: 10.2196/mededu.9185] [Medline: 29506968]
25. de Jong PG, Pickering JD, Hendriks RA, Swinnerton BJ, Goshtasbpour F, Reinders ME. Twelve tips for integrating massive open online course content into classroom teaching. Med Teach 2020;42(4):393-397. [doi: 10.1080/0142159X.2019.1571569] [Medline: 30860930]

26. Hendriks RA, de Jong PG, Admiraal WF, Reinders ME. Teaching modes and social-epistemological dimensions in medical massive open online courses: lessons for integration in campus education. Med Teach 2019;41(8):917-926. [doi: 10.1080/0142159X.2019.1592140] [Medline: 31008667]
27. Hendriks RA, de Jong PG, Admiraal WF, Reinders ME. Instructional design quality in medical massive open online courses for integration into campus education. Med Teach 2020;42(2):156-163. [doi: 10.1080/0142159X.2019.1665634] [Medline: 31566999]
28. 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] [Medline: 11392867]
29. Dolmans DH, De Grave W, Wolfhagen IH, van der Vleuten CP. Problem-based learning: future challenges for educational practice and research. Med Educ 2005;39(7):732-741. [doi: 10.1111/j.1365-2929.2005.02205.x] [Medline: 15960794]
30. Ten Cate TJ, Kusurkar RA, Williams GC. How self-determination theory can assist our understanding of the teaching and learning processes in medical education. AMEE guide No. 59. Med Teach 2011;33(12):961-973. [doi: 10.3109/0142159X.2011.595435] [Medline: 22225433]
31. Dolmans DH, Tigelaar D. Building bridges between theory and practice in medical education using a design-based research approach: AMEE Guide No. 60. Med Teach 2012;34(1):1-10. [doi: 10.3109/0142159X.2011.595437] [Medline: 22250671]
32. Strijbos JW, Sluijsmans D. Unravelling peer assessment: methodological, functional, and conceptual developments. Learn Instr 2010;20(4):265-269. [doi: 10.1016/j.learninstruc.2009.08.002]
33. Schut S, van Tartwijk J, Driessen E, van der Vleuten C, Heeneman S. Understanding the influence of teacher-learner relationships on learners' assessment perception. Adv Health Sci Educ Theory Pract 2020;25(2):441-456 [FREE Full text] [doi: 10.1007/s10459-019-09935-z] [Medline: 31664546]
34. Heeneman S, Oudkerk Pool A, Schuwirth LW, van der Vleuten CP, Driessen EW. The impact of programmatic assessment on student learning: theory versus practice. Med Educ 2015;49(5):487-498. [doi: 10.1111/medu.12645] [Medline: 25924124]
35. Bryant DA, Carless DR. Peer assessment in a test-dominated setting: empowering, boring or facilitating examination preparation? Educ Res Policy Prac 2010;9(1):3-15. [doi: 10.1007/s10671-009-9077-2]
36. Evans C. Making sense of assessment feedback in higher education. Rev Educ Res 2013;83(1):70-120. [doi: 10.3102/0034654312474350]
37. Simonsmeier BA, Peiffer H, Flaig M, Schneider M. Peer feedback improves students' academic self-concept in higher education. Res High Educ 2020;61(6):706-724. [doi: 10.1007/s11162-020-09591-y]
38. Lerchenfeldt S, Mi M, Eng M. The utilization of peer feedback during collaborative learning in undergraduate medical education: a systematic review. BMC Med Educ 2019;19(1):321 [FREE Full text] [doi: 10.1186/s12909-019-1755-z] [Medline: 31443705]
39. Filius RM, De Kleijn RA, Uijl SG, Prins FJ, Van Rijen HV, Grobbee DE. Challenges concerning deep learning in SPOCs. Int J Technol Enhanc Learn 2018;10(1/2):111. [doi: 10.1504/ijtel.2018.088341]
40. Garrison DR, Kanuka H. Blended learning: uncovering its transformative potential in higher education. Internet High Educ 2004;7(2):95-105. [doi: 10.1016/j.iheduc.2004.02.001]
41. Filius RM, de Kleijn RA, Uijl SG, Prins FJ, Rijen HV, Grobbee DE. Promoting deep learning through online feedback in SPOCs. Frontline Learn Res 2018;6(2):92-113. [doi: 10.14786/flr.v6i2.350]
42. Jansen R. Dealing with autonomy: self-regulated learning in open online education. Utrecht University Repository. 2019. URL: https://dspace.library.uu.nl/bitstream/handle/1874/383982/proefschrift_reneejansen.pdf?sequence=1&isAllowed=y [accessed 2022-03-31]
43. Robinson JD, Persky AM. Developing self-directed learners. Am J Pharm Educ 2020;84(3):847512 [FREE Full text] [doi: 10.5688/ajpe847512] [Medline: 32313284]
44. Zamberg I, Schiffer E, Stoermann-Chopard C. Novice and advanced learners' satisfaction and perceptions of an e-learning renal semiology module during the COVID-19 pandemic: mixed methods study. JMIR Med Educ 2021;7(2):e29216 [FREE Full text] [doi: 10.2196/29216] [Medline: 34048357]
45. George PP, Papachristou N, Belisario JM, Wang W, Wark PA, Cotic Z, et al. Online eLearning for undergraduates in health professions: a systematic review of the impact on knowledge, skills, attitudes and satisfaction. J Glob Health 2014;4(1):010406 [FREE Full text] [doi: 10.7189/jogh.04.010406] [Medline: 24976965]
46. Azevedo R, Taub M, Mudrick NV. Understanding and reasoning about real-time cognitive, affective, and metacognitive processes to foster self-regulation with advanced learning technologies. In: Alexander PA, Schunk DH, Greene JA, editors. Handbook of self-regulation of learning and performance. New York, NY: Routledge; 2017.
47. Wortha F, Azevedo R, Taub M, Narciss S. Multiple negative emotions during learning with digital learning environments - evidence on their detrimental effect on learning from two methodological approaches. Front Psychol 2019;10:2678 [FREE Full text] [doi: 10.3389/fpsyg.2019.02678] [Medline: 31849780]
48. Taub M, Sawyer R, Lester J, Azevedo R. The impact of contextualized emotions on self-regulated learning and scientific reasoning during learning with a game-based learning environment. Int J Artif Intell Educ 2020;30(1):97-120. [doi: 10.1007/s40593-019-00191-1]
49. Harley JM, Pekrun R, Taxer JL, Gross JJ. Emotion regulation in achievement situations: an integrated model. Educ Psychol 2019;54(2):106-126. [doi: 10.1080/00461520.2019.1587297]

50. Kononowicz AA, Woodham LA, Edelbring S, Stathakarou N, Davies D, Saxena N, et al. Virtual patient simulations in health professions education: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019;21(7):e14676 [FREE Full text] [doi: 10.2196/14676] [Medline: 31267981]
51. Combs C, Combs PF. Emerging roles of virtual patients in the age of AI. AMA J Ethics 2019;21(2):E153-E159 [FREE Full text] [doi: 10.1001/amajethics.2019.153] [Medline: 30794125]
52. Murphy S, Imam B, Whitehouse LM. Twelve tips for utilizing virtual patients to teach professionalism. MedEdPublish 2016;5:107. [doi: 10.15694/mep.2016.000107]
53. Azevedo R, Cromley JG. Does training on self-regulated learning facilitate students' learning with hypermedia? J Educ Psychol 2004;96(3):523-535. [doi: 10.1037/0022-0663.96.3.523]
54. Kirschner PA, Sweller J, Clark RE. Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educ Psychol 2006;41(2):75-86. [doi: 10.1207/s15326985ep4102_1]
55. To WM, Lung JW. Factors influencing internship satisfaction among Chinese students. Educ Train 2020;62(5):543-558. [doi: 10.1108/et-01-2020-0023]
56. Stepczynski J, Holt SR, Ellman MS, Tobin D, Doolittle BR. Factors affecting resident satisfaction in continuity clinic-a systematic review. J Gen Intern Med 2018;33(8):1386-1393 [FREE Full text] [doi: 10.1007/s11606-018-4469-8] [Medline: 29736753]
57. Hense H, Harst L, Küster D, Walther F, Schmitt J. Implementing longitudinal integrated curricula: systematic review of barriers and facilitators. Med Educ 2021;55(5):558-573. [doi: 10.1111/medu.14401] [Medline: 33099784]
58. Cant R, Ryan C, Cooper S. Nursing students' evaluation of clinical practice placements using the clinical learning environment, supervision and nurse teacher scale - a systematic review. Nurse Educ Today 2021;104:104983. [doi: 10.1016/j.nedt.2021.104983] [Medline: 34229233]
59. Bonnie LH, Cremers GR, Nasori M, Kramer AW, van Dijk N. Longitudinal training models for entrusting students with independent patient care?: a systematic review. Med Educ 2022;56(2):159-169. [doi: 10.1111/medu.14607] [Medline: 34383965]

Abbreviations

LED: learning experience design
LUMC: Leiden University Medical Center
SDT: self-determination theory
SPOC: Small Private Online Course
WPL: workplace learning

Edited by T Leung; submitted 14.04.21; peer-reviewed by LA Lee, I Zamberg, M Manzanares; comments to author 28.06.21; revised version received 19.08.21; accepted 22.02.22; published 07.04.22.

Please cite as:
Hamoen EC, De Jong PGM, Van Blankenstein FM, Reinders MEJ
Design and First Impressions of a Small Private Online Course in Clinical Workplace Learning: Questionnaire and Interview Study
JMIR Med Educ 2022;8(2):e29624
URL: https://mededu.jmir.org/2022/2/e29624
doi:10.2196/29624
PMID:35389362

Original Paper

Perspectives of 360-Degree Cinematic Virtual Reality: Interview Study Among Health Care Professionals

Elizabeth Beverly[1*], PhD; Brooke Rigot[1*], BS; Carrie Love[2*], MFA; Matt Love[2*], MFA

[1]Ohio University Heritage College of Osteopathic Medicine, Athens, OH, United States

[2]J Warren McClure School of Emerging Communication Technologies, Ohio University, Athens, OH, United States

[*]all authors contributed equally

Corresponding Author:
Elizabeth Beverly, PhD
Ohio University Heritage College of Osteopathic Medicine
1 Ohio University
Irvine Hall 307
Athens, OH, 45701
United States
Phone: 1 740 593 4616
Fax: 1 740 593 2205
Email: beverle1@ohio.edu

Abstract

Background: The global market for medical education is projected to increase exponentially over the next 5 years. A mode of delivery expected to drive the growth of this market is virtual reality (VR). VR simulates real-world objects, events, locations, and interactions in 3D multimedia sensory environments. It has been used successfully in medical education for surgical training, learning anatomy, and advancing drug discovery. New VR research has been used to simulate role-playing and clinical encounters; however, most of this research has been conducted with health professions students and not current health care professionals. Thus, more research is needed to explore how health care professionals experience VR with role-playing and clinical encounters.

Objective: The aim of this study was to explore health care professionals' experiences with a cinematic VR (cine-VR) training program focused on role-playing and clinical encounters addressing social determinants of health, Appalachian culture, and diabetes. Cine-VR leverages 360-degree video with the narrative storytelling of cinema to create an engaging educational experience.

Methods: We conducted in-depth telephone interviews with health care professionals who participated in the cine-VR training. The interviews were audio recorded and transcribed verbatim. A multidisciplinary team coded and analyzed the data using content and thematic analyses with NVivo software.

Results: We conducted 24 in-depth interviews with health care professionals (age=45.3, SD 11.3, years; n=16, 67%, women; n=22, 92%, White; and n=4, 17%, physicians) to explore their experiences with the cine-VR training. Qualitative analysis revealed five themes: immersed in the virtual world: seeing a 360-degree sphere allowed participants to immerse themselves in the virtual world; facilitated knowledge acquisition: all the participants accurately recalled the culture of Appalachia and listed the social determinants of health presented in the training; empathized with multiple perspectives: the cine-VR provided a glimpse into the real life of the main character, and participants described thinking about, feeling, and empathizing with the character's frustrations and disappointments; perceived ease of use of cine-VR: 96% (23/24) of the participants described the cine-VR as easy to use, and they liked the 360-degree movement, image resolution, and sound quality but noted limitations with the buttons on the headsets and risk for motion sickness; and perceived utility of cine-VR as a teaching tool: participants described cine-VR as an effective teaching tool because it activated visual and affective learning for them.

Conclusions: Participants emphasized the realism of the cine-VR training program. They attributed the utility of the cine-VR to visual learning in conjunction with the emotional connection to the VR characters. Furthermore, participants reported that the cine-VR increased their empathy for people. More research is needed to confirm an association between the level of immersion and empathy in cine-VR training for health care professionals.

(JMIR Med Educ 2022;8(2):e32657) doi:10.2196/32657

KEYWORDS

virtual reality; qualitative; medical education; health care; digital learning; learning platform; health care providers

Introduction

Background

The global market for medical education is expected to grow by US $143.3 billion during 2021-2025 [1]. Although numerous factors will contribute to this rise, new modes of delivery are expected to be a major driving factor [1]. Virtual reality (VR) is one mode of delivery propelling the growth of the medical education market [2]. In 2019, the global VR health care market was valued at US $2.1 billion [3]. By 2026, it is expected to reach US $30.4 billion, with a compound annual growth rate of 42.4%. Advances in VR technology and improved accessibility and affordability of wearable devices are driving the demand for VR in medical education [2]. Moreover, the COVID-19 pandemic emphasized the need for virtual learning opportunities in medical education [4].

VR is technology that simulates real-world objects, events, locations, and interactions in 3D multimedia sensory environments [5,6]. Users explore sensory environments in real time (ie, first-person active learning) through different levels of immersion [6]. Levels of immersion range from low to moderate to high, depending on the degree to which the VR is inclusive, extensive, surrounding, vivid, and matching [7]. Inclusive refers to whether the VR removes signals (eg, external noise) from the physical world. Extensive refers to the number of senses engaged in the VR. Surrounding refers to the presentation of the VR and the degree to which the physical world is blocked out. Vivid refers to the fidelity and resolution of the VR. Matching refers to whether the VR is adapted to fit proprioceptive feedback from, or spatial awareness of, the user. Fully immersive VR reaches high levels of inclusive, extensive, surrounding, vivid, and matching in the virtual world such that the user experiences the virtual world as if it were the real world [2]. An example of fully immersive VR is viewing 360-degree video using a headset with controllers so that the user can see, hear, and move in, as well as interact with, the virtual world. In contrast, semi-immersive VR provides a user with a virtual environment that is not fully inclusive, extensive, surrounding, vivid, or matching [2]. It provides a user with a complete picture of the virtual world; however, the user is limited in their ability to move in, or interact with, this virtual world [2]. In other words, the user has a strong connection to the virtual world, but they are not completely isolated from the real world. An example of semi-immersive VR is viewing a 360-degree video on a web-based or smart device–based platform.

Objectives

Traditionally, VR in medical education has been used to supplement surgical training [8,9], teach human anatomy [10], and visualize molecular complexes to advance drug discovery [11]. More recently, VR has been used to simulate clinical skills training and role-playing [12,13]. VR offers the user repeated attempts to practice difficult conversations or scenarios, with fewer time constraints and less perceived pressure from observation by superiors [12,14]. Research has shown that VR with role-playing and clinically based scenarios enhances empathy toward patients in medically and culturally diverse populations [14-16]. Most of this research has focused on health professions students and not professionals, although a new study demonstrated the effectiveness of improving cultural self-efficacy and attitudes toward diabetes among health care professionals [17]. Thus, more research is needed to understand how VR with role-playing and clinical encounters can be used effectively for practicing health care professionals. The aim of this study was to explore health care professionals' experiences with a cinematic VR (cine-VR) training program. Specifically, this qualitative study assessed the educational impact and knowledge learned among health care professionals after the cine-VR training.

Methods

Research Design

Cine-VR is close to a fully immersive experience. Users have the ability to look around the virtual world in 360 degrees and hear sounds with spatialized audio, thereby creating a highly inclusive, surrounding, and vivid experience. This experience reinforces the belief that the user is present in the virtual world. Cine-VR differs from traditional VR, which uses computer-generated characters and worlds; in contrast, cine-VR uses live images filmed through a camera such as in cinema. In cine-VR, filmmakers leverage 360-degree video and apply the techniques of cinema to VR. Techniques include narrative storytelling, scripts, actors, lighting, framing, lens choices, focus, color, and so on. These techniques create an engaging educational experience for users. Studies have shown that immersion technologies such as cine-VR enhance education by allowing multiple perspectives, situated learning, and improved knowledge transfer to other situations [17,18].

We used narrative inquiry to understand health care professionals' experiences with the cine-VR technology and training program [19]. For the purposes of this study, health care professionals included physicians, nurse practitioners, nurses, psychologists, exercise physiologists, physical therapists, dietitians, pharmacists, certified diabetes educators, community health workers, and certified health education specialists. Narrative inquiry captures the relationship between the individual experience and the greater cultural context through the communication of knowledge and the experience of time [19]. In this study, narrative inquiry described the lived experience of participating in an innovative cine-VR training program for continuing medical education and continuing education credits in the evolving landscape of medical education.

Cine-VR Training

We conducted in-depth telephone interviews with a subset of health care professionals who participated in the 3-hour cine-VR training program entitled *Using Virtual Reality to Visualize Diabetes in Appalachia*. The purpose of this cine-VR training program was to educate health care professionals in Ohio about implicit bias and social determinants of health per the funding announcement. We selected education to address bias toward Appalachian culture, social determinants of health, and type 2 diabetes, considering that more than one-third of the counties in Ohio are Appalachian and their diabetes rates are nearly double the national average (20% vs 11%) [20]. The participating professionals watched 10 cine-VR simulations and 2 traditional films that displayed interactions between a woman aged 72 years with type 2 diabetes and her primary care physician, pharmacist, family members, and community. The cine-VR featured 4 *guided simulations*, which were face-to-face role-playing conversations with the main character and her providers. All cine-VR simulations were screened in an Oculus Go (Reality Labs) head-mounted display to allow participants to turn their head and body in any direction to gather information as though they were present in the actual location. In addition to the cine-VR simulations, we delivered 12 brief didactics (eg, 3 to 5 minutes for each didactic) that addressed the following content: (1) diabetes burnout, (2) food insecurity, (3) Appalachian cultural strengths, (4) transportation barriers in rural areas, (5) person-centered care, (6) psychosocial issues in diabetes, (7) financial insecurity and the cost of diabetes medications, (8) lack of social support, (9) Appalachian cultural values, (10) diabetes complications, (11) diabetes comorbidities, and (12) effective communication. The participating health care professionals received 3.0 continuing medical education or continuing education credits at no cost. The primary aim of the original study was to improve participants' cultural self-efficacy and diabetes attitudes, wherein we observed statistically significant improvements in all subscales of cultural self-efficacy and diabetes attitudes after training; a detailed description of these findings has been published elsewhere [17].

Cine-VR Technology

A variety of cameras were used to create the cine-VR images. Specifically, we used Insta360 OneX and Insta360 Pro 2 with a Sennheiser Ambeo soundbar and Zoom F6 microphones. For most of the filming, we used the Insta360 Pro 2, which allowed us to capture the entire 360-degree space at the same time. This was achieved with multiple sensor-lens combinations that capture different portions of the image simultaneously. These disparate images were combined into a single 360-degree panorama, either in real time before the image was saved to the memory card or later using Final Cut Pro (Apple Inc) and DaVinci Resolve (Blackmagic Design).

Ethics Approval

Ethics approval for the study was obtained from the Ohio University Office of Research Compliance Institutional Review Board (20-X-111). In complying with federal, state, and local laws and regulations for human subjects research, we ensured that our research met the requirements set forth in the regulations on public welfare in part 46 of title 45 of the Code of Federal Regulations, the principles set forth in the *Belmont Report*, and the Helsinki Declaration of 1975.

Sample

We used maximum variation sampling, a form of purposive sampling [21], to recruit a wide range of participants from different disciplines. The inclusion criteria for participation included adults aged ≥18 years who could read and speak in English and were currently employed as health care professionals (ie, physicians, nurse practitioners, nurses, psychologists, exercise physiologists, physical therapists, dietitians, pharmacists, certified diabetes educators, community health workers, and certified health education specialists) in Ohio. Participants were recruited through email and word of mouth. Specifically, we emailed participants who shared their email addresses with us after participating in the training or we spoke with individuals who completed the training.

Data Collection

All interviews were conducted by two trained qualitative researchers (EB and CL), who asked participants broad, open-ended questions about the cine-VR content and educational experience. Specifically, participants were asked what they learned about diabetes, social determinants of health, and Appalachian culture during the cine-VR training program. The researchers used directive probes to clarify questions and elicit additional information from the participants (Textbox 1). All interviews were conducted through telephone because of the COVID-19 pandemic and restrictions with in-person human subjects research because of state-mandated lockdowns. The interviews ranged from 20 to 45 minutes in length. We collected data until we achieved data saturation; that is, when variation in the data leveled off or no new perspectives emerged from the interviews [21]. All interviews were digitally audio recorded and transcribed verbatim. The researchers performed quality checks of the transcribed files by listening to half of the interview recordings to validate the data. The transcripts were cleaned to remove participants' names and identifiers to protect confidentiality.

Textbox 1. Interview questions.

Questions

- What is your clinical, teaching, or academic role?
- Overall, how would you describe your experience with the virtual reality?
- In your own words, how would you have described Appalachian culture before you watched the virtual reality?
- Did your view of Appalachia change after seeing the virtual reality? Please explain.
- After watching the virtual reality, how do you believe diabetes is affecting Appalachian Ohio?
- In your opinion, how is diabetes viewed in rural Appalachia?
- Prior to the virtual reality, what did you think were the biggest barriers to diabetes care in Appalachian Ohio?
- After watching the virtual reality, what are the biggest barriers to diabetes care in Appalachian Ohio?
- What was the most important factual information you gained as part of this experience?
- What aspect of this experience had the most impact on your attitudes and beliefs about this people with diabetes in Appalachian Ohio?
- How will this experience most impact your future behavior when working with people in Appalachian Ohio?
- How does virtual reality compare to the different learning styles you have experienced in your career?
- In your opinion, what makes the virtual reality effective or ineffective?
- What suggestions or recommendations do you have for improving the virtual reality experience?
- Is there anything else about this virtual reality experience that you want to share?

Qualitative Analysis

The multidisciplinary research team, which consisted of a behavioral diabetes researcher, a medical student, and 2 VR experts, used standard qualitative techniques to analyze the data. First, two members of the research team (EB and BR) performed content analysis by independently coding common words, phrases, and ideas in the qualitative data [22-24]. They met to review coded data to establish intercoder reliability; all discrepancies were reviewed, discussed, and resolved through consensus [24]. The Cohen κ coefficient between the 2 coders was 0.951, indicating almost perfect agreement [25,26]. No negative cases were excluded from the analysis [27]. After all the transcripts were coded and reviewed, one member of the research team (BR) entered the coded transcripts in NVivo software (QSR International) to organize the coded data. The two remaining members of the team (CL and ML) reviewed the codes to achieve researcher corroboration. The research team selected themes that characterized the participants' experiences with the cine-VR technology and training program that occurred multiple times, both within and across transcripts.

Rigor

We used investigator triangulation with team members representing different disciplines to support the credibility (ie, validity) of the data. Investigator triangulation also provided a means to identify cognitive biases in the analysis [28]. We reviewed findings with 21% (5/24) of the participants to achieve participant corroboration and establish validity of the accounts [29]. Next, we supported transferability (ie, external validity) through rich descriptions of the participants' experiences with the cine-VR training program and the inclusion of verbatim quotations [27]. To support dependability (ie, reliability), we invited a researcher, Marilyn D Ritholz, not involved with the study to conduct an inquiry audit to examine the process and the product to determine whether the findings and conclusions were supported by the data [27]. Finally, to establish confirmability (ie, objectivity), we created an audit trail to document the research steps we took from the start of the study to the reporting of the findings [27].

Results

Overview

A total of 24 health care professionals participated in in-depth telephone interviews; the mean age of the participants was 45.3 (SD 11.3) years; 18 (75%) identified as women, 4 (17%) as men, 1 (4%) as gender queer, and 1 (4%) as transgender or nonbinary. Of the 24 participants, 22 (92%) self-identified as White and 2 (8%) as Asian (Table 1); 5 (21%) were community health workers or certified health education specialists, 4 (17%) were physicians, 4 (17%) were nurses, 3 (13%) were dietitians (13%), 2 (8%) were pharmacists, 2 (8%) were exercise physiologists, 2 (8%) were medical directors, 1 (4%) was a physical therapist, and 1 (4%) was a psychologist. Years of clinical or academic experience ranged from 1 to 5 years (7/24, 29%), 6 to 10 years (4/24, 17%), 16 to 20 years (7/24, 29%), 21 to 25 years (4/24, 17%), and >25 years (2/24, 8%).

Table 1. Participants' demographic characteristics (N=24).

Characteristic	Values
Age (years), mean (SD)	45.3 (11.3)
Gender, n (%)	
Woman	18 (75)
Man	4 (17)
Gender queer	1 (4)
Transgender or nonbinary	1 (4)
Race, n (%)	
American Indian or Alaska Native	0 (0)
Asian	2 (8)
Black or African American	0 (0)
Hispanic or Latinx	0 (0)
Native Hawaiian or Pacific Islander	0 (0)
White	22 (92)
Two or more races	0 (0)
Another race not listed	0 (0)
Occupation, n (%)	
Community health worker or certified health education specialist	5 (21)
Physician	4 (17)
Nurse	4 (17)
Dietitian	3 (13)
Pharmacist	2 (8)
Exercise physiologist	2 (8)
Medical directors	2 (8)
Physical therapist	1 (4)
Psychologist	1 (4)
Years in health care, n (%)	
1 to 5	7 (29)
6 to 10	4 (17)
11 to 15	0 (0)
16 to 20	7 (29)
21 to 25	4 (17)
>25	2 (8)

Transcript identifiers are included with quotations indicating participant number and discipline. The following themes emerged from the data analysis:

Immersed in the Virtual World

Of the 24 participants, 22 (92%) described feeling immersed or connected to the virtual world. They commented on the realism they felt in the cine-VR simulations. Participants identified with the dress, the cars, the houses, the people, and the barriers they faced:

> *I identified with the video. I identified with the dress. The vehicles. The scenarios and problems that those families face. The coexisting disorders. The housing. How things are undone and unfinished. Raising grandchildren. You know, taking care of everyone else but yourself, and being viewed and stereotyped as someone who is heavy and lazy, and maybe just doesn't care when really that is not the case in impoverished areas such as southeast Ohio.* [ID 103, Nurse]

> *Just for me, [it was] the realism of it. It's our neighbors. It's what we see here in Appalachia. It's the patients that I've dealt with since 1994 when I became a nurse. So I think it was well done. I think*

> *the realism was really important and impactful for a lot of folks who aren't from here.* [ID 101, Nurse]

The cine-VR gave participants privileged, unfiltered access to the main character's environment, social interactions, and daily challenges. In particular, the cine-VR slice-of-life vignettes depicted the main character's struggles with social determinants of health, including poverty, food insecurity, housing instability, transportation barriers, lack of access to health care, and limited social support. Seeing a 360-degree sphere allowed participants to immerse themselves in the simulations, feeling the disarray and chaos of the main character's home or the substantial burden of the family car breaking down (Figure 1):

Figure 1. Photograph of the main character, her daughter, and a tow truck driver. The main character's car breaks down on the side of a rural road. Photograph by ML.

> *The first time I watched it I was just in awe. I felt like I was a fly on the wall. I can look all around. I can look at the unfinished walls. I can look at the clutter. I can look at the unwashed dishes that might been two or three days, maybe an overflowed trashcan. Kids in and out. The chaos, the child with the traumatic brain injury that's just kind of part of the wallpaper. I mean, they're existing in there and we're all working our lives around them. I think that happens all the time in rural Appalachia. But to be inside someone's home and inside their life, it's almost like being invisible and seeing all the things that people want to hide either purposely or indirectly that affect their life. It's like a glimpse inside someone's personal life and why things are the way they are.* [ID 103, Nurse]

> *I think that the whole scene with the flat tire with the car I think was huge because it wasn't just the flat tire, right? There was the flat tire, there was the fact that then she didn't have money to pay the tow truck driver and she was going to thank him with a cake because that's how she shows her caring for people is through baking sweets. I think all of that happens all the time...There's just so much...I felt like that scene stacked everything up really nicely. This could 100% happen to anyone on any given day because it felt very realistic.* [ID 112, Medical director]

These participants viewed the main character, her environment, and her interactions with professionals as true to life and clinical practice in Appalachian Ohio. In fact, some participants felt that the cine-VR was so realistic and immersive that the main character must be a real person or one of their patients:

> *So I knew it was simulated, right? It's stuff I see every day. But I think what I liked about it is it solidified all the things I thought. I'm like, "Well, I don't think people can get in their car and go, I think people smoke, and I think they have a hard time taking care of their families." It was like taking all the research that everybody's done and putting it into an actuality that actually is actual. I still think about it almost like it was my patient, right? I remember thinking of her in the car and twisting her [ribbon] around her finger, and her husband dying. I remember all of it like she was one of my patients.* [ID 109, Pharmacist]

Furthermore, the realism of the cine-VR left a lasting impression on the participants. Several of the participants stated that they look back on the simulations to inform their current clinical practice:

> *It was engaging. It's very realistic. It resonated well with what I can envision my patients dealing with and brought to light a lot of the issues that I don't think a lot of physicians can envision well without seeing something like this...To see it come to life in that format really made it real, and helps it stick so I remember that when I'm dealing with patients every day. And it's probably not exaggerating to say that I think about what I saw in the Lula Mae videos when I'm taking care of patients in the emergency*

> *department. And I think it's probably been a year or so since I've looked at the videos and I'm still flashing little scenes of what I saw in those videos when I'm interacting with patients, so it's powerful. Realistic. Really hit the mark to bring a lot of things to light.* [ID 123, Physician]

Facilitated Knowledge Acquisition

All the participants accurately recalled the culture of rural Appalachian Ohio and listed the social determinants of health presented in the cine-VR training program. As all interviews were conducted 12 to 18 months after the training program, this suggested that the participants successfully learned and retained the information. They described the cultural strengths of Appalachia, including caregiving, loyalty, and generosity, as follows:

> *I would describe the people as close-knit communities...Everyone knows each other. Everybody helps each other out for the most part. If someone needs something, people dig deep, and give 'til it hurts when they don't have much to give at all.* [ID 103, Nurse]

> *I would describe Appalachian culture as a passionate group of people, people who love their families and those that they claim as family, which may or may not be blood-related. I would describe them as very self-reliant.* [ID 104, Nurse]

> *I think you really hit it with that, especially because she's [Lula Mae] clearly the matriarch in that situation. How many family units do we see in Athens County and surrounding counties that are based on that matriarch, and that's huge. I think that's a really positive thing.* [ID 112, Medical director]

Participants also remembered the numerous social determinants of health affecting people in this rural and underserved area. Participants identified financial insecurity, food insecurity, transportation barriers, lack of access to health care, low educational achievement, and social isolation as social determinants of health negatively affecting people living in rural Appalachian Ohio:

> *There are many social factors that are barriers. It's not just the health care infrastructure that's the problem. It's all of the other life stressors and challenges that make it difficult to prioritize a person's individual health. Poverty, food insecurities, transportation barriers.* [ID 117, Certified health education specialist]

> *Money. Not only access to care, but money to purchase the things needed. Education to understand how to take care of yourself. How to monitor your own glucose, et cetera.* [ID 102, Psychologist]

> *I would say transportation for sure. Maybe not that the person doesn't have a car; they might have a car, but maybe they don't have money to put the gas in the car. Or, maybe it's not reliable enough to get them where it is that they need to go.* [ID 104, Nurse]

> *I think one of the things that really are important to harp on are the isolation that a lot of people feel here.* [ID 111, Dietitian]

Multiple participants noted how low educational achievement in Appalachian Ohio was a barrier to health care, citing both a lack of education regarding one's personal health as well as lack of education regarding the severity of diabetes:

> *I feel like we are still very undereducated about diabetes in this area. I don't think people understand the impact that it makes. I don't think the people living here actually understand the impact that it makes on their health. I feel like—I guess I said it in my last statement when I said it's not a priority. I don't feel that people in this area feel like their health is a priority.* [ID 109, Pharmacist]

> *I think it really comes down to their lack of education and their lack of understanding not only to their predisposition for it, their risk of it, but how serious it is once you are, in fact, diagnosed with it and how you care for yourself.* [ID 102, Psychologist]

In addition, many participants discussed common beliefs about diabetes in Appalachian Ohio. They described the sense of fatalism surrounding a diabetes diagnosis because of the high prevalence of the disease in the area. Participants thought that this fatalist attitude toward diabetes prevented people from making behavior changes to prevent or delay the onset of diabetes:

> *I think that there is some, definitely some folks who think, "Well everyone in my family's had it so I'm going to get it too," because that's what they've seen and that's what they know.* [ID 104, Nurse]

> *I think most just are—I'm trying to think of the word—resigned to the fact that it's going to happen...If somebody else in their family has had it, they assume they're going to get it.* [ID 101, Nurse]

Empathized With Multiple Perspectives

Of the 24 participants, 21 (88%) commented on the empathy they experienced during the cine-VR simulations. The cine-VR provided a glimpse into the real life of the main character, and participants could sense what she was thinking and feeling. Perceiving these thoughts and feelings may have transferred the experiences to the users. Participants described feeling Lula Mae's frustrations, her disappointments, and "being torn between competing responsibilities" (Figure 2):

Figure 2. Throughout the cinematic virtual reality, the viewers watch the main character take care of everyone in her life and as a result put her health and well-being last. This photograph of the main character depicts the isolation and lack of support she receives in her diabetes management. Photograph by ML.

> *You really got to see life through her—I felt like I was walking in her shoes? I felt emotions as part of that learning experience. I felt the frustrations that she might've felt. The disappointments, and being torn between competing responsibilities. I felt like I understood why she wasn't taking her medication as prescribed, and it didn't seem like it had anything to do with her desire to address her own health, but more a function of her taking care of everyone else and all of the responsibilities that came with that. She kept putting herself last. I think it's a realistic reflection of what many people do when they're overburdened and overstressed and underresourced.* [ID 117, Certified health education specialist]

> *It's the emotions, yes. When it feels like you're right there observing Lula Mae in her environment, and her interactions with the family, with the doctor, with the community members, you can almost feel her or the other people's emotions. Like when Lula Mae was stuck on the road, I could feel her desperation because, hey, I was stuck on the road once, and I did not like it at all. You're empathizing with her. When she was at the doctor's office, I could feel her anxiety.* [ID 107, Medical director]

Participants also reflected on the thoughts and feelings of the other characters in the cine-VR. They recognized that identifying harmful emotions from the professionals and family members was as important as empathizing with the main character. They explained that understanding the perspectives of each of the characters helped them to identify how they could support people with diabetes in the future:

> *The way that it makes you internally reflect on how that situation could apply to you or how you could possibly be as a provider or as an educator in the shoes to help someone like Lula Mae. Because I would say she just felt beat down and almost helpless—one thing after another. It made me reflect and it made me think like if I were in even the nurse's shoes at the doctor's office or the medical assistant at the doctor's office, if I were in that person's shoes, what could I do? Or her family support was not there. She was the support for the whole family. So, as a family member, how can you be there for people, or as a friend, how can you be there for people so that you don't have a friend that feels like they're in Lula Mae's situation? Because I think that's what it made me look at is I never want one of my friends who has diabetes to feel like that and like they're alone and try to figure it out all on their own. So, I think one of the best things it gives you is it truly gives you a feeling for what it's like, and it gives you that time to reflect and discuss what's going on.* [ID 113, Exercise physiologist]

> *Looking at it from different perspectives—it was more impactful that way. Because you got to see and kind of empathize, if you will, just how it was coming from the perspective of that patient who maybe didn't have the money for medication, or from the doctor's perspective, being so overwhelmed and so tired in the health care system he's working in right now. Or the feeling of you've got a patient who has so many roles to fulfill because of the plate that she has been handed. Like taking care of grandchildren or taking care of the family member that needed her. Then also the spiritual side of it where she was actually trying to help other people by doing what she felt was right at the food pantry. It was just wholehearted.* [ID 119, Nurse]

Some of the participants expressed an intention to change how they practiced medicine after viewing the simulations. They explained that it increased their empathy for patients and enhanced their understanding of the challenges faced by people in rural Appalachia:

> *It changed my practice in the sense it made me a heck of a lot more empathetic towards what they may or may not be saying. I always look at body language and can read that pretty well, but now I understand a little bit of what might be going on. It also helped me to understand the culture of my patient population and what things may be important, not important to them, and what may be playing a role in why they're making the decisions that they're making or making the decisions where they're not able to take care of their health because they are so burdened with so many issues in their lives.* [ID 122, Physician]

Perceived Ease of Use of the Technology

Of the 24 participants, 5 (21%) had prior experience with VR technology; however, none of them had prior experience of cine-VR. When asked about its ease of use, 96% (23/24) of the participants described the cine-VR as easy or simpler than expected:

> *The technology was pretty simple actually. I mean, when we had somebody who was facilitating, who was very understanding of the technology, it was very easy...Clinicians are already using technology on a daily basis to do their charts and things so this is much less burdensome than [electronic health record system] so to speak. So as far as I'm concerned, no, this is really easy, very simple.* [ID 122, Physician]

> *It was a lot easier than I thought it would be, to be honest with you. Because aside from a couple of experiences using the headsets, I'm not a gamer and I don't use them for other things like people who do are used to that environment. But from picking it out of the box to turning it on, to putting it on my head, getting the right video, and getting it going was remarkably easy.* [ID 123, Physician]

Several participants reported feeling surprised by the live-action VR. They did not know it was possible to see live-action VR using a headset:

> *Live action virtual reality is what really grabbed me about this one. I've seen a lot of virtual reality avatar-based things and a lot of things where you're grabbing the virtual syringe and you're injecting the virtual syringe into the virtual type patient. It's okay. I think we're still probably like years off for where that really has the virtual tactile feel that I need to make a valuable trend. Where this one really hit was that it used live action where live action was the best way to do it. So, if I really want to feel what it's like to be in you know, Lula Mae's house, you know, then boy, this is the way to do it.* [ID 123, Physician]

Participants also commented on the ability to move 360 degrees in any direction during the simulation as well as the quality of the image and sound:

> *It was more than I expected. I was—I guess I was not expecting the 360-degree virtual reality and to have so much freedom with movement to be able to look all around me, around the room. That type of technology was pretty groundbreaking to me. I was more expecting to just look straight ahead and everybody would be seeing the same thing at the same time.* [ID 106, Dietitian]

> *I really liked the video and the sound quality, especially. My hearing isn't always the best. I was in marching band, and I played drum line, and I work at a really busy pharmacy. So being able to hear so clearly with that technology was really, really engaging. And then the video quality, I was thoroughly shocked by. I was expecting more like my other experience of virtual reality, where it's a little grainier.* [ID 108, Pharmacist]

The cine-VR training program had some technical and sensory issues. A few participants remembered accidentally hitting the buttons on the side of the headset, which turned off the simulations:

> *I think having a little presession, here's how you use the Oculus or don't press this button, because I accidentally turned it off twice, are important reminders. I am a technically savvy individual. However, even I struggled with it, despite the best efforts of all the support people there.* [ID 124, Physician]

The other main issue was worries about motion sickness (to address the potential dizziness and nausea from viewing the cine-VR, we disclosed that these risks were possible and offered alternative viewing devices). Of the 24 participants, 2 (8%) reported experiencing motion sickness and 3 (13%) thought that they would experience motion sickness but did not:

> *As far as the motion sickness that I personally experienced, a disclosure to let people know that there are bright lights flashing. The refresh rate of whatever the Oculus is may impair some folks and to be aware of that going in so that I didn't, I wouldn't be as sick, but that's me personally. Not everyone may experience that.* [ID 124, Physician]

> *So, it was easier to use than I had anticipated it to be. I thought it was going to be kind of complicated. I also was a little worried about it making me like motion sick, because sometimes movies, like 3D movies will do that, but it wasn't, none of that happened at all. It was super easy to use, easy to focus on things, and really neat to be able to look around the whole room and really get an entire picture of what's going on in different parts of the house for example during different things. So, I would say all in all, it was completely positive, a very awesome experience and a very neat and different way to learn*

things, but it kept me engaged as well. [ID 113, Exercise physiologist]

Perceived Utility of Cine-VR as a Teaching Tool

Of the 24 participants, 22 (92%) described cine-VR as an effective teaching tool and 23 (96%) expressed satisfaction with the learning experience. The cine-VR training was viewed favorably by participants both with and without prior experience with VR technology:

> *I thought it was super cool. I'm not a gamer. I mean, I'm not a major tech person, so it was a really cool experience for me. I'd never done anything like that, and I haven't done anything or seen anything like that since. So, if that's the way things are headed, I think it's pretty cool and it's certainly a great way for the next generation to learn things.* [ID 101, Nurse]

Participants believed that the realism of the cine-VR offered a way of learning that affected the user on a personal level:

> *I think, especially as a clinician, I think you can go through training and training, you could read books, you can watch videos, you can look at PowerPoints, but when you're forced to be in the room with someone like that and they're looking right at you, and you look around and you're the only one there, or—you know what I mean, you kind of have to be a little bit more involved so you have to be more attentive to what's going on so you respond better.* [ID 105, Exercise physiologist]

Participants recognized the applicability of cine-VR as an educational tool for multiple topic areas and levels of learning. They felt that the realism of the simulations and full immersion in the virtual environment made the training an effective tool. Moreover, they believed that the cine-VR tapped into affective learning. Participants believed that visual learning in conjunction with emotional connection to the VR characters made the training program a meaningful and memorable experience:

> *This format is our first experience doing this, and I think aside from being realistic in the stories and the content resonating, the format really immersed the learner in the environment. We strive for realism. We strive for full immersion. We strive for really activating all of the senses to create the best learning environment. When you get all the senses firing and you immerse somebody fully into this reality, it creates memories and you feed into that psychology in different parts. Because it activates your emotional centers and visual cortex and your auditory cortex and all of this works together to really form all these connections for you to remember or bring into our practice. So yeah, but the full immersion, the application in this area, that's what really has the power.* [ID 123, Physician]

> *Students nowadays are learning a completely different curriculum and a very case-based format. This is taking case-based formats to a whole new level because it's not on paper, you are visually seeing it. And many people do better. Yes, they can do paper, but to actually see it almost like a movie. To almost embed it into your visual memory, it goes without saying that this is superior. There's no question about it...And I think that visual piece in the moment in the life, it's just like a movie or a TV show where there are parts of it that you just take away and they embed, and you remember that and you can't forget it in a way that you can totally forget a piece of paper with writing on it.* [ID 122, Physician]

Discussion

Principal Findings

In this qualitative study, we explored health care professionals' experiences with a cine-VR training program in diabetes, social determinants of health, and Appalachian culture. Participants described feeling immersed in the cine-VR training. They commented on their ability to watch the main character interact with professionals, family members, and the environment by means of the cine-VR 360-degree video. They noted that the opportunity to see life through the main character's eyes elicited many emotions in them. Most of the participants reported feeling the frustrations, anxieties, and disappointments experienced by the main character and empathizing with her struggles. Participants agreed that these emotions and empathy were key to the learning process and knowledge they acquired. All participants were able to recall the social determinants of health addressed in the cine-VR as well as the cultural aspects of rural Appalachia. Finally, 96% (23/24) of the participants described the cine-VR as easy to use with surprising technical features, including live action, 360-degree movement, high resolution, and high sound quality. Drawbacks included the buttons on the headset and risk for motion sickness. In summary, these findings suggest that health care professionals perceived immersion and empathy to be key drivers in the success of the cine-VR training.

Comparison With Prior Work

The value of cine-VR training is that it gives health care professionals a glimpse inside the lives of their patients and why things are the way they are. The more professionals can understand their patients' personal lives, the more they can empathize with their challenges and struggles. Recent gaming research found that immersive VR from the first-person perspective of a person with chronic pain increased kindness (ie, a subscale in the Empathy Scale) and willingness to help after a simulation, suggesting an increase in implicit and explicit empathy [30]. Another study on perspective-taking tasks showed that VR-based tasks increased self-reported empathy more than narrative-based tasks in a simulation about the homeless population [31]. These studies support the use of VR to promote empathy in medical education. Empathy in diabetes care is critical, given prior work demonstrating a 40% to 50% lower risk of all-cause mortality at 10-year follow-up when patients newly diagnosed with type 2 diabetes experienced high levels of provider empathy in the first year after diagnosis [32]. Participants in our study described the empathy they felt for the characters in the cine-VR. Thus, cine-VR training has the potential to increase empathy among health care professionals and students; however, more mixed methods research is needed to measure empathy and other prosocial behaviors (ie, helping, sharing, and comforting) before and after completing this cine-VR training.

Participants highlighted the importance of feeling immersed in the virtual world. A recent systematic review and meta-analysis by Kyaw et al [6] found that VR improved knowledge and skills acquisition among health care professionals compared with traditional modes of education (eg, textbooks and lectures) and other digital education (eg, 2D images on a screen and web-based teaching). The findings from this systematic review and meta-analysis suggest that VR is an effective mode of delivery for medical education and VR is more effective than traditional or other digital education in knowledge and skills acquisition [6]. Similar findings have been reported among medical students: those participating in fully immersive VR reported significantly higher gains in knowledge than students in partially immersive VR [33]. Another recent study focused on social determinants of health and empathy in dental education found that 360-degree, immersive VR improved cognitive, affective, and skills-based learning in residents and faculty [34]. Overall, these findings combined with the qualitative responses from our participants underscore the importance of immersive VR to enhance the user experience and improve learning outcomes.

Limitations

The study limitations included homogeneity of the sample with regard to gender and race and ethnicity. Thus, the qualitative findings may not be transferable or generalizable to people not represented in the sample. Future research with a more diverse sample is necessary to explore experiences with the cine-VR training program. Additional limitations of the study included the small sample size and participant self-selection, which also limit the transferability of the findings. However, qualitative research differs from quantitative research in that it is not driven by sample size, randomness, and power calculations. Rather, qualitative research rests on the notion of data saturation or the point at which no new information is collected for data analysis. Therefore, sample size was not an indicator of rigor in our study. Historically, an adequate sample size for an in-depth individual interview study is 15 to 20 participants [21]. With regard to participant self-selection, individuals who volunteered to participate in the interviews may have had more positive experiences with the cine-VR training than participants who did not volunteer to participate in the study. In addition, we recruited participants through email and word of mouth, which increased our selection bias because participants were not randomly selected to participate in the interviews. Finally, self-reported data are vulnerable to social desirability bias. To minimize bias, the researchers informed participants that their responses were confidential and could not be linked back to their personal identity. Furthermore, the investigators emphasized the voluntary nature of participation and explicitly informed the participants that their responses had no bearing on their employment. Finally, our original study did not include a control group as a comparison. Research comparing cine-VR with a proper control condition is underway to examine the effectiveness of cine-VR in changing health care professionals' knowledge, beliefs, and empathy.

Conclusions

The participating health care professionals perceived the cine-VR training to be a valuable educational experience that generated empathy toward the VR characters. They attributed the value of the cine-VR to the immersive and realistic nature of the 360-degree virtual environment. Future research is needed to examine the impact of cine-VR training on quantitative measures of immersion, empathy, and prosocial behaviors among current professionals and health professions students. Cine-VR has the potential to play an integral role in clinical training as medical education expands to meet the growing need for virtual platforms.

Acknowledgments

This study was part of the Medicaid Simulation Project funded by the Ohio Department of Medicaid and administered by the Ohio Colleges of Medicine Government Resource Center. The views expressed in the cine-VR simulations and this manuscript are solely those of the creators and do not represent the views of the state of Ohio or federal Medicaid programs. Participant compensation was funded by the Osteopathic Heritage Foundation Ralph S Licklider, DO, Endowed Professorship in Behavioral Diabetes awarded to EB.

Conflicts of Interest

None declared.

References

1. TechNavio. Global medical education market 2021-2025. London, UK: TechNavio; 2021.
2. Pottle J. Virtual reality and the transformation of medical education. Future Healthc J 2019 Oct;6(3):181-185 [FREE Full text] [doi: 10.7861/fhj.2019-0036] [Medline: 31660522]
3. Global virtual reality in healthcare market size by end-use(research and diagnostics, laboratories, hospitals and clinics), by application (patient care management, education and training, fitness management,), by geographic scope and forecast. Jersey City, NJ, USA: Verified Market Research; 2021.
4. Tabatabai S. COVID-19 impact and virtual medical education. J Adv Med Educ Prof 2020 Jul;8(3):140-143 [FREE Full text] [doi: 10.30476/jamp.2020.86070.1213] [Medline: 32802908]

5. Nesenbergs K, Abolins V, Ormanis J, Mednis A. Use of augmented and virtual reality in remote higher education: a systematic umbrella review. Educ Sci 2021;11(1):8. [doi: 10.3390/educsci11010008]
6. 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] [Medline: 30668519]
7. Slater M, Wilbur S. A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. Presence (Camb) 1997 Dec;6(6):603-616. [doi: 10.1162/pres.1997.6.6.603]
8. Gurusamy KS, Aggarwal R, Palanivelu L, Davidson BR. Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane Database Syst Rev 2009 Jan 21(1):CD006575. [doi: 10.1002/14651858.CD006575.pub2] [Medline: 19160288]
9. Nemirovsky DR, Garcia AJ, Gupta P, Shoen E, Walia N. Evaluation of surgical improvement of clinical knowledge ops (SICKO), an interactive training platform. J Digit Imaging 2021 Aug;34(4):1067-1071. [doi: 10.1007/s10278-021-00482-x] [Medline: 34244878]
10. Triepels CP, Smeets CF, Notten KJ, Kruitwagen RF, Futterer JJ, Vergeldt TF, et al. Does three-dimensional anatomy improve student understanding? Clin Anat 2020 Jan;33(1):25-33 [FREE Full text] [doi: 10.1002/ca.23405] [Medline: 31087400]
11. Ratamero EM, Bellini D, Dowson CG, Römer RA. Touching proteins with virtual bare hands : visualizing protein-drug complexes and their dynamics in self-made virtual reality using gaming hardware. J Comput Aided Mol Des 2018 Jun;32(6):703-709 [FREE Full text] [doi: 10.1007/s10822-018-0123-0] [Medline: 29882064]
12. Sapkaroski D, Mundy M, Dimmock MR. Virtual reality versus conventional clinical role-play for radiographic positioning training: a students' perception study. Radiography (Lond) 2020 Feb;26(1):57-62. [doi: 10.1016/j.radi.2019.08.001] [Medline: 31902456]
13. Ma Z, Huang KT, Yao L. Feasibility of a computer role-playing game to promote empathy in nursing students: the role of immersiveness and perspective. Cyberpsychol Behav Soc Netw 2021 Nov;24(11):750-755. [doi: 10.1089/cyber.2020.0371] [Medline: 33989057]
14. Bricker DJ, Pardee CJ. Nurse experts jump-start clinical simulation in rehabilitation nursing: supporting new graduate transition to competence. Nurs Educ Perspect 2011;32(1):34-36. [doi: 10.5480/1536-5026-32.1.34] [Medline: 21473480]
15. Everson N, Levett-Jones T, Lapkin S, Pitt V, van der Riet P, Rossiter R, et al. Measuring the impact of a 3D simulation experience on nursing students' cultural empathy using a modified version of the Kiersma-Chen Empathy Scale. J Clin Nurs 2015 Oct;24(19-20):2849-2858. [doi: 10.1111/jocn.12893] [Medline: 26178187]
16. Dyer E, Swartzlander BJ, Gugliucci MR. Using virtual reality in medical education to teach empathy. J Med Libr Assoc 2018 Oct;106(4):498-500 [FREE Full text] [doi: 10.5195/jmla.2018.518] [Medline: 30271295]
17. Beverly EA, Love C, Love M, Williams E, Bowditch J. Using virtual reality to improve health care providers' cultural self-efficacy and diabetes attitudes: pilot questionnaire study. JMIR Diabetes 2021 Jan 27;6(1):e23708 [FREE Full text] [doi: 10.2196/23708] [Medline: 33502335]
18. Dede C. Immersive interfaces for engagement and learning. Science 2009 Jan 02;323(5910):66-69. [doi: 10.1126/science.1167311] [Medline: 19119219]
19. Connelly FM, Clandinin DJ. Stories of experience and narrative inquiry. Educ Res 1990 Jun 1;19(5):2-14. [doi: 10.3102/0013189x019005002]
20. Ruhil A, Johnson L, Cook K, Trainer M, Beverly E, Olson M, et al. What does diabetes look like in our region: a summary of the regional diabetes needs assessment study. The Diabetes Institute, Ohio University Heritage College of Osteopathic Medicine. 2017. URL: https://www.ohio.edu/voinovich-school/sites/ohio.edu.voinovich-school/files/sites/A-Summary-of-the-Regional-Diabetes-Needs-Assessment-Study-2017_January-5-2018.pdf [accessed 2022-04-03]
21. Morse JM, Field PA. Qualitative research methods for health professionals. 2nd edition. Thousand Oaks, CA, USA: Sage Publications; 1995.
22. Pope C, Mays N. Qualitative research in health care. 2nd edition. London, UK: BMJ Books; 2000.
23. Krippendorf K. Content analysis: an introduction to its methodology. 2nd edition. Thousand Oaks, CA, USA: Sage Publications; 2004.
24. Neuendorf KA. The content analysis guidebook. Thousands Oaks, CA, USA: Sage Publications; 2002.
25. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968 Oct;70(4):213-220. [doi: 10.1037/h0026256] [Medline: 19673146]
26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
27. Lincoln YS, Guba EG. Naturalistic inquiry. New York, NY, USA: Sage Publications; 1985.
28. Miles MB, Huberman AM. Qualitative data analysis: an expanded sourcebook. 2nd edition. Thousand Oaks, CA, USA: Sage Publications; 1994.
29. Denzin NK. The research act: a theoretical introduction to sociological methods. 2nd edition. Milton Park, UK: Routledge; 2009.

30. Tong X, Gromala D, Kiaei Ziabari SP, Shaw CD. Designing a virtual reality game for promoting empathy toward patients with chronic pain: feasibility and usability study. JMIR Serious Games 2020 Aug 07;8(3):e17354 [FREE Full text] [doi: 10.2196/17354] [Medline: 32763883]
31. 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 Oct 17;13(10):e0204494 [FREE Full text] [doi: 10.1371/journal.pone.0204494] [Medline: 30332407]
32. Dambha-Miller H, Feldman AL, Kinmonth AL, Griffin SJ. Association between primary care practitioner empathy and risk of cardiovascular events and all-cause mortality among patients with type 2 diabetes: a population-based prospective cohort study. Ann Fam Med 2019 Jul;17(4):311-318 [FREE Full text] [doi: 10.1370/afm.2421] [Medline: 31285208]
33. Gutiérrez F, Pierce J, Vergara VM, Coulter R, Saland L, Caudell TP, et al. The effect of degree of immersion upon learning performance in virtual reality simulations for medical education. Stud Health Technol Inform 2007;125:155-160. [Medline: 17377256]
34. Amini H, Gregory ME, Abrams MA, Luna J, Roland M, Sova LN, et al. Feasibility and usability study of a pilot immersive virtual reality-based empathy training for dental providers. J Dent Educ 2021 Jun;85(6):856-865. [doi: 10.1002/jdd.12566] [Medline: 33638168]

Abbreviations

cine-VR: cinematic virtual reality
VR: virtual reality

Edited by T Leung; submitted 05.08.21; peer-reviewed by G Kurillo, M Linnett; comments to author 11.10.21; revised version received 20.12.21; accepted 22.03.22; published 29.04.22.

Please cite as:
Beverly E, Rigot B, Love C, Love M
Perspectives of 360-Degree Cinematic Virtual Reality: Interview Study Among Health Care Professionals
JMIR Med Educ 2022;8(2):e32657
URL: https://mededu.jmir.org/2022/2/e32657
doi:10.2196/32657
PMID:35486427

Original Paper

Assessing Physician's Motivational Communication Skills: 5-Step Mixed Methods Development Study of the Motivational Communication Competency Assessment Test

Vincent Gosselin Boucher[1,2], PhD; Simon Bacon[1,3], PhD; Brigitte Voisard[1,2], BA; Anda I Dragomir[1,2], PhD; Claudia Gemme[1,2], BA; Florent Larue[4], MD, PhD; Sara Labbé[1,2], MSc; Geneviève Szczepanik[1], PhD; Kimberly Corace[5,6], PhD; Tavis Campbell[7,8], PhD; Michael Vallis[9], PhD; Gary Garber[10,11], MD; Codie Rouleau[5], PhD; Jean G Diodati[1], MD; Doreen Rabi[12], MD, PhD; Serge Sultan[13], PhD; Kim Lavoie[1,2], PhD; Network For Health Behavior Change And Promotion (CAN-Change)[14]

[1]Montreal Behavioural Medicine Centre, Centre Intégré Universitaire de santé et services sociaux du Nord-de-l'Ile-de-Montréal (CIUSSSNIM), Montreal, QC, Canada

[2]Department of Psychology, Université du Québec à Montréal, Montréal, QC, Canada

[3]Department of Health, Kinesiology and Applied Physiology, Concordia University, Montreal, QC, Canada

[4]Faculty of Medicine of Montpellier, Montpellier, France

[5]Department of Psychiatry, University of Ottawa, Ottawa, ON, Canada

[6]The Royal's Institute of Mental Health Research, Ottawa, ON, Canada

[7]Total Cardiology Cardiac Rehabilitation, Calgary, AB, Canada

[8]Department of Psychology, University of Calgary, Calgary, AB, Canada

[9]Department of Family Medicine, Dalhousie University, Halifax, NS, Canada

[10]Department of Medicine, University of Ottawa, Ottawa, ON, Canada

[11]Department of Medicine, University of Toronto, Toronto, ON, Canada

[12]Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada

[13]Department of Pediatrics, Université de Montréal, Montreal, QC, Canada

[14]Canadian Network for Health Behavior Change and Promotion (CAN-Change), Montreal, QC, Canada

Corresponding Author:
Vincent Gosselin Boucher, PhD
Montreal Behavioural Medicine Centre
Centre Intégré Universitaire de santé et services sociaux du Nord-de-l'Ile-de-Montréal (CIUSSSNIM)
5400, boul Gouin O,
Montreal, QC, H2G 3G2
Canada
Phone: 1 4388808310
Email: vincent.gosselin.boucher@gmail.com

Abstract

Background: Training physicians to provide effective behavior change counseling using approaches such as motivational communication (MC) is an important aspect of noncommunicable chronic disease prevention and management. However, existing evaluation tools for MC skills are complex, invasive, time consuming, and impractical for use within the medical context.

Objective: The objective of this study is to develop and validate a short web-based tool for evaluating health care provider (HCP) skills in MC—the Motivational Communication Competency Assessment Test (MC-CAT).

Methods: Between 2016 and 2021, starting with a set of 11 previously identified core MC competencies and using a 5-step, mixed methods, integrated knowledge translation approach, the MC-CAT was created by developing a series of 4 base cases and a scoring scheme, validating the base cases and scoring scheme with international experts, creating 3 alternative versions of the 4 base cases (to create a bank of 16 cases, 4 of each type of *base* case) and translating the cases into French, integrating the cases into the web-based MC-CAT platform, and conducting initial internal validity assessments with university health students.

Results: The MC-CAT assesses MC competency in 20 minutes by presenting HCPs with 4 out of a possible 16 cases (randomly selected and ordered) addressing various behavioral targets (eg, smoking, physical activity, diet, and medication adherence). Individual and global competency scores were calculated automatically for the 11 competency items across the 4 cases, providing automatic scores out of 100. From the factorial analysis of variance for the difference in competency and ranking scores, no significant differences were identified between the different case versions across individual and global competency (*P*=.26 to *P*=.97) and ranking scores (*P*=.24 to *P*=.89). The initial tests of internal consistency for rank order among the 24 student participants were in the *acceptable* range (α=.78).

Conclusions: The results suggest that MC-CAT is an internally valid tool to facilitate the evaluation of MC competencies among HCPs and is ready to undergo comprehensive psychometric property analyses with a national sample of health care providers. Once psychometric property assessments have been completed, this tool is expected to facilitate the assessment of MC skills among HCPs, skills that will better support patients in adopting healthier lifestyles, which will significantly reduce the personal, social, and economic burdens of noncommunicable chronic diseases.

(JMIR Med Educ 2022;8(2):e31489) doi:10.2196/31489

KEYWORDS

assessment; motivational communication; tool development; physicians; health promotion

Introduction

The World Health Organization estimates that >71% of deaths worldwide result from noncommunicable diseases (NCDs), including cardiovascular disease, cancer, chronic lung disease, diabetes, and obesity [1]. Despite advances in genetic, pharmacological, and surgical medicine, the prevalence and associated social and economic burden of NCDs are increasing rather than decreasing [2]. This is unsurprising, given that the underlying cause of most NCDs is not biological factors but harmful human behaviors (eg, smoking, physical inactivity, and poor diet) that are poorly addressed by current biomedical approaches [3,4].

As part of offering comprehensive care, health care providers (HCPs) are often responsible for providing some form of behavior change counseling (BCC) to patients who exhibit health risk behaviors. At present, this typically takes the form of offering *persuasive information and advice* [5-7], which has been shown to be either ineffective or counterproductive because patients feel as if they are being *told what to do* [8,9]. When evidence-based BCC approaches are offered by HCPs, they tend to have positive impacts on patient engagement in and the adoption of healthy lifestyle choices [10]. However, one of the most popular of these approaches (ie, motivational interviewing) has generally demonstrated poor uptake by physicians. This has been attributed to perceptions of it being too rigid, taking too much time to implement in practice, and lying outside the physician's scope of practice [11,12].

To address the limitations of motivational interviewing, we codeveloped (with behavior change experts and HCPs) a new BCC approach called motivational communication (MC), which is based on motivational interviewing and theoretical models of behavior change (eg, self-determination theory [13], social-cognitive theory [14], and transtheoretical model [15]) and incorporates more cognitive behavioral therapy–based components and practical considerations regarding real-world clinical encounters in an NCD management context. Designed as a behavior change communication style specifically developed for HCPs, it is evidence–based and time–efficient and can be used to promote patient engagement, adoption of healthy behaviors, and sustained self-management of chronic conditions [16]. MC was defined as reflecting 11 core communication competencies that have a solid evidence base for behavior change in the context of NCD management [16]. These 11 competencies were summarized under the mnemonic "LEARN tHE BASICs." These competencies are reflective listening, expressing empathy, demonstrating acceptance, tolerance, and respect, responding to resistance, (not) negatively judging or blaming, (not) expressing hostility or impatience, eliciting *change-talk or* evocation, (not) being argumentative or confrontational, setting goals, providing information neutrally, and being collaborative [16].

After defining MC as well as developing the content of the MC training program to be delivered to HCPs (the *MOTIVATOR* program), we also developed an accompanying MC competency assessment tool to evaluate skill acquisition among HCPs receiving training in this approach. A recent review of the literature on the quality of existing communication assessment tools among HCPs revealed a great deal of heterogeneity over the 45 different assessment tools that were identified. This review also indicated that few tools were developed using appropriate theoretical models (49%), and many failed to clearly define or describe the communication competencies they were designed to evaluate (19%) [17]. In addition, 65% used scoring methods that required extensive training on the part of external assessors, and 93% of the tools required the use of standardized (ie, a person playing the role of a patient; 61%) or real patients (32%) to complete their evaluations [17], potentially undermining the feasibility of implementing this type of evaluation in *real life*. Existing competency assessment tools are hence complex, invasive, time-consuming, and impractical for use in many medical contexts.

Effective, feasible, and user-friendly competency evaluation tools are important not only for assessing the quality and efficacy of training programs but also for ensuring that patients benefit from the BCC methods used by HCPs. Using an integrated knowledge translation (iKT) approach, which is a collaborative model of knowledge production between stakeholders and researchers [18,19], the objective of this study was to develop a new web-based MC competency assessment tool called the Motivational Communication Competency Assessment Test (MC-CAT), to conduct initial internal validity assessments, and to evaluate the ranking and competency score consistency between the base cases and the modified cases as part of a larger iterative development process for this new tool.

Methods

Concept

The concept of the web-based MC-CAT assessment tool is to present HCPs with a series of patient cases with a specific behavioral target (eg, engaging in more physical activity). Each case comprises a simulated interaction between a virtual patient and the provider, which focuses on engaging the patient in a discussion about changing their health behavior. Patient information (ie, patient's picture, age, sex, health condition, health behavior status, and medications) is accessible by clicking on the icon in the top right-hand corner (Figure 1). Each MC-CAT assessment requires completing 4 cases selected at random from a 16-case bank for a total assessment time ranging from 15 to 20 minutes (approximately 5 minutes per case). The cases were designed to be relatively short to maximize the tool's acceptability and uptake by the busy HCPs.

Figure 1. Example of patient chart and interaction between the physician and the virtual patient case. (A) and (B) patient information; always accessible by clicking on the icon in the top right-hand corner; (C) patient's initial statement (with audio); (D) list of answers; each answer was associated with a score on different motivational communication competencies.

Ethics Approval

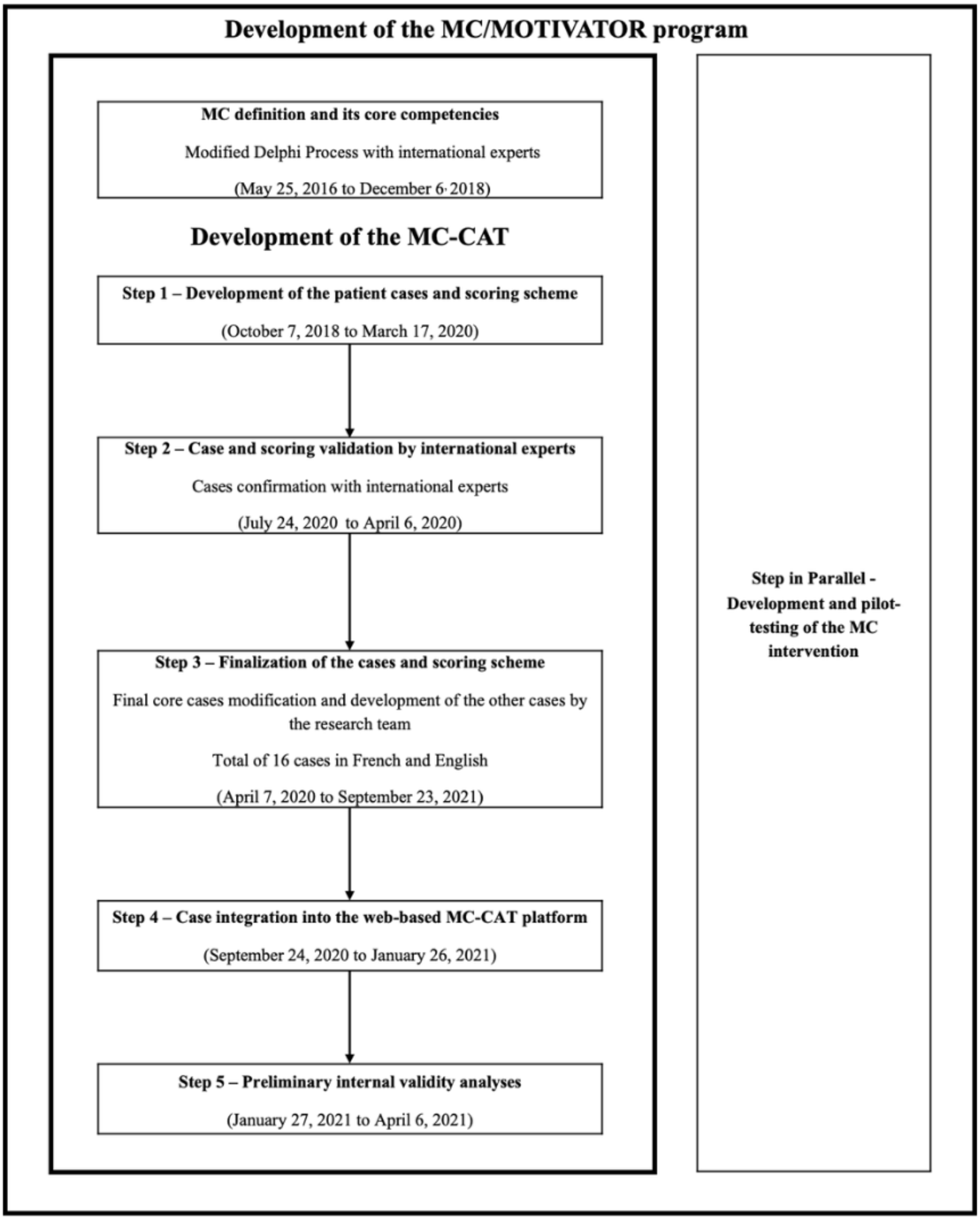
Ethics approval was provided by the *Centre Intégré Universitaires de Santé et de Service Sociaux du Nord-de-l'île-de-Montréal* (number 2016-1206), and all participants provided informed consent electronically.

Development

Overview

We have previously defined the 11 core communication competencies of MC [16], which represent the individual competencies evaluated by the MC-CAT. To develop the MC-CAT, we followed a 5-step, mixed methods, iKT approach based on established methods for instrument development and validation, which engaged relevant knowledge users (physicians, HCPs, researchers, and health care administrators) [20-22]. The steps are shown in Figure 2, and are as follows: (1) developing a series of 4 base cases and a scoring scheme to assess the 11 communication competencies of MC, (2) validating the content of the base cases and scoring scheme with international experts, (3) creating 3 alternative versions of the 4 base cases (resulting in a bank of 16 cases, 4 of each type of base case) and translating the cases into French (necessary for a Canadian audience), (4) integrating the cases into the web-based MC-CAT platform, and (5) conducting initial internal validity assessments with a sample of 31 university-allied health students.

Figure 2. Overview of the MC-CAT development steps. MC: motivational communication; MC-CAT: Motivational Communication Competency Assessment Test.



Step 1: Development of the Patient Cases and Scoring Scheme

The goal of this step was to create 4 base cases (A1, B1, C1, and D1) to assess all 11 MC competencies through simulated medical consultations. One of the primary aims was to develop an assessment tool that would accurately reflect real-life medical consultations targeting health behavior changes in the context of NCD management. To ensure that the behavioral targets of our cases were relevant to clinical practice, we sent an open invitation to Canadian physicians from 4 target specialties known to treat a high volume of patients with NCDs (ie, cardiology, respirology, internal medicine, and general practitioners) to complete a brief (10-minute) web-based survey (LimeSurvey GmbH), which was available in English and French. Invitations were emailed to physicians throughout

January 2018 through relevant medical associations (eg, Hypertension Canada and Diabetes Canada) and networks (eg, the Canadian Respiratory Research Network). In this survey, physicians were asked to identify what they believed to be the most important health risk behaviors that they encountered with patients in their efforts to prevent or treat NCDs (checklist with the possibility of adding behaviors) and rank them in order of their relative importance (from most important to least important). This resulted in the identification of the 4 most common health risk behaviors encountered in the context of NCD prevention and management, which would form the behavioral targets of the 4 base cases of MC-CAT.

The research team then proceeded to develop 4 base cases around the 4 most important health risk behaviors as identified by physicians, which acted as templates to develop alternate case versions to expand the tool to 16 cases (4 different health behaviors—represented by the letters A, B, C, and D—targeted across 4 different patient cases, numbered from 1 to 4). Base cases were designed to simulate a consultation between an HCP and a patient, which focused on engaging the patient in a conversation about changing their health behavior. Between October 2018 and March 2020, the research team worked on developing 4 base case scripts. The goal was to create a conversation flow beginning with the patient providing an opening statement of concern, after which the HCP was prompted to reply by selecting 1 of 5 multiple-choice responses, reflecting the most MC-consistent responses to the least MC-inconsistent responses (scored from 1 to 5). Each base case included 6 to 7 levels of exchange, each providing the HCP the opportunity to demonstrate ≥1 of the 11 core MC competencies. This resulted in a scoring scheme that indicated the extent to which responses were MC consistent or inconsistent on a 5-point scale (2=very MC consistent, 1=somewhat MC consistent, 0=neither MC consistent nor inconsistent, −1=MC inconsistent, and −2=very MC inconsistent). The scoring format was constructed with a range from +2 to −2 to reflect the nonneutrality of responding in an MC-inconsistent manner (which may be counterproductive for behavior change). Care was taken to provide HCPs with multiple opportunities to demonstrate each of the 11 MC competencies across the 4 cases to ensure that >1 interaction could be used to calculate an individual competency score. MC competency scores were calculated automatically by averaging the individual competency scores across the 4 cases, which were then summed and converted to a global score out of 100, reflecting overall MC competency (theoretical range −93.6 to 100).

Step 2: Case and Scoring Validation by International Experts

After creating the 4 base cases (A1, B1, C1, and D1), we proceeded to validate the presence of the different MC competencies reflected in each case, as well as their rank order from most to least MC consistent, using a multi-round survey among a sample of international experts (7/14, 50% women with an average of 21 years of experience in BCC, SD 9 years; Table 1 provides a summary of expert characteristics). The first survey was launched in July 2019 and ended in September 2019. The results were used to calculate the consensus score, reflecting the level of agreement between the experts and the original classifications attributed by the research team. The original classifications were considered *good* if the agreement with the experts was perfect, *acceptable* if the expert rank was a +1-point or –1-point deviation from the rank the research team had indicated (eg, ranked 5 instead of 4), and *poor* if the expert order was a +2-point –2-point deviation from the research team rank (eg, ranked 1 instead of 3). These results were used to make minor modifications to some response items based on the criteria summarized in Table 2. This resulted in the construction of a second confirmatory survey, in which the same experts were asked to confirm their agreement with the new competency classification and ranking. The survey was launched in October 2019 and ended in April 2020.

Table 1. Demographic information of health care providers and international experts.

Variable	Health care providers (N=80), n (%)	International experts (N=14), n (%)
Gender		
Women	36 (45)	7 (50)
Men	44 (55)	7 (50)
Language spoken		
English speaking	30 (38)	12 (86)
French speaking	50 (62)	2 (14)
Age (years)		
<30	2 (3)	—[a]
30-39	27 (34)	3 (21)
40-49	9 (11)	6 (43)
50-59	19 (24)	2 (14)
60-69	23 (29)	2 (14)
≥70	—	1 (7)
Duration of practice (years)		
0-5	20 (25)	—
6-10	9 (11)	1 (7)
11-15	7 (9)	3 (21)
16-20	3 (4)	6 (43)
21-25	13 (16)	—
26-30	10 (13)	1 (7)
>30	18 (23)	3 (21)
Province of practice or country		
Canada		
Alberta	2 (3)	2 (14)
British Columbia	3 (4)	—
New Brunswick	1 (1)	—
Nova Scotia	2 (3)	2 (14)
Ontario	17 (21)	3 (21)
Quebec	55 (69)	4 (29)
Sweden	—	1 (7)
United Kingdom	—	1 (7)
United States	—	1 (7)

[a]Data not available.

Table 2. Criteria for case modification after evaluation by behavior change experts.

Criteria	Modifications
Ranking the choices of the different base cases physician's options	
≥70% of good agreement[a] and ≤10% of poor agreement[b] between external experts and research team rankings[c]	No modification
≥90% of good and acceptable[d] agreement and ≤10% of poor agreement between external experts and research team rankings	No modification
Competency identification	
≥70% agreement between external experts and research team identification	Competencies kept or added if there was agreement
Between 40% and 69% of agreement between external experts and research team identification	Competencies may be kept or added depending on the research team's consensus
≤30% agreement between external experts and research team identification	Competencies deleted

[a]If agreement with the experts and the research team was perfect.

[b]The expert order was +2 or –2 deviations in rank from the research team (eg, ranked 1 instead of 3).

[c]If this criterion is not met, the modifications must result in a minimum of 70% perfect agreement or 90% perfect and partial agreement and <10% of complete disagreement.

[d]If the expert rank was +1 or –1 point deviation in rank from the research team (eg, ranked 5 instead of 4).

Step 3: Finalization of the Cases and Scoring Scheme

Overview

One of the aims was to design the MC-CAT to allow for the variation (or flexible *programming*) in the demographic variables of the virtual patients (ie, age, sex, race, culture, and language), NCDs, and contextual variables (eg, personal information) across cases. To achieve this, we adapted the 4 base cases to expand the test bank (ie, create alternate versions of the 4 base cases) without altering the core cases' original structure (allowing us to maintain the integrity of the scoring algorithm and conversation branching across established at step 2). As such, every MC-CAT assessment of 4 randomly selected cases should maintain the integrity and psychometric properties by including a variant of each of the 4 original cases. This resulted in the creation of 16 unique patient cases (cases A1 to A4, B1 to B4, C1 to C4, and D1 to D4) reflecting a variety of cases that would be appropriate for multiple types of NCD management situations and relevant across different medical specialties. These 16 cases were then formally translated into French (using back translation to ensure equivalence) [23] to permit use with French-speaking physicians, which is relevant in the Canadian context.

Scoring Algorithm

The MC-CAT provides a subscale score for each of the 11 MC competencies, as well as a global score summarizing overall MC competency. Individual competency scores were calculated by considering the number of times the person had chosen responses that included a certain competency (eg, reflective listening) across the 4 cases, divided by the total number of times the competency could have been chosen across the evaluation, and multiplied by the relative proportion of opportunities to demonstrate that competency over all the competencies evaluated.

To obtain scores for global competency, the scores for the positive competencies (reflective listening; expressing empathy; demonstrating acceptance, tolerance, and respect; responding to resistance; eliciting *change-talk* or evocation; setting goals; providing information neutrally; and being collaborative) are aggregated together, and the negative competencies (negatively judging or blaming, expressing hostility or impatience, and being argumentative or confrontational) are subtracted from the sum, as reflected in the following equation:

> Competencies score (%) = ([reflective listening / number of reflective listening occasions × percentage of reflective listening cases] + [expressing empathy / number of expressing empathy occasions × percentage of expressing empathy cases] + [evocation / number of evocation occasions × percentage of evocation cases] + [responding to resistance / number of responding to resistance occasions × percentage of responding to resistance cases] + [setting goals / number of setting goals occasions × percentage of setting goals cases] + [acceptance, tolerance, and respect / number of acceptance, tolerance, and respect occasions × percentage of acceptance, tolerance, and respect cases] + [being collaborative / number of being collaborative occasions × percentage of being collaborative cases] + [providing information neutrally / number of providing information neutrally occasions × percentage of providing information neutrally cases]) – ([hostility + negatively judging + argumentative] × [1 / number of hostility, negatively judging, argumentative occasion cases × 100]) **(1)**

The score for the ranking (ie, whether the physician selected the most consistent response with MC [1] or the least consistent response with MC [5]) is calculated by adding the number of times a participant selected the ranks of response choices multiplied by a constant ranging from 2 to −2 (eg, 10 times the second choice is multiplied by 1). The following is the equation for ranking scores:

Ranking score (%) = ([number of first choices × 2] + [number of second choices × 1] + [number of third choices × 0] + [number of fourth choices × −1] + [number of fifth choices × −2]) / (50×100) **(2)**

Step 4: Case Integration Into the Web-Based MC-CAT Platform

Following the completion of steps 2 to 3, the first web-based computerized version of the MC-CAT (version 1.0), including 32 virtual patient cases (16 in French and 16 in English, which are identical and translated) and a demographic questionnaire (including sex, age, location of practice, primary medicine specialty, clinical setting, years of practice, number of patients, and physicians' attitudes toward addressing health risk behaviors) was designed and created in collaboration with 42 Comets Inc, a software developer with expertise in the creation of electronic education and training programs. To test the user interface of the MC-CAT, we conducted user experience research with 27 volunteer graduate students, HCPs, and behavior change experts also involved in step 2. The goal was to determine (1) the clarity of the instructions and tasks, (2) the navigability of the platform, (3) the synchronicity between audio and video information, and (4) the acceptability of the duration of the assessment. The responses were used to refine the aspects of the web-based interface to optimize the functionality of the program.

Step 5: Preliminary Internal Validity Analyses

Overview

We refined the aspects of the web-based interface based on user experience testing. The final step in the development process was to collect preliminary psychometric properties of the MC-CAT from a sample of MC-naïve undergraduate allied HCP students who were recruited via email and invited to complete 2 MC-CAT assessments approximately 12 weeks apart. Each assessment involved completing a basic sociodemographic questionnaire followed by the MC-CAT (4 randomly selected cases from the 16-case bank, 1 from each series A to D).

Case Consistency Analyses

To evaluate the consistency of the ranking and competency scores between the base cases (A1, B1, C1, and D1) and the modified cases (cases A2 to A4, B2 to B4, C2 to C4, and D2 to D4), a factorial ANOVA for the difference in competency and ranking scores was used to determine differences between case variations (ie, case A1 vs A2 vs A3 vs A4) across individual competency scores, global competency scores, and ranking scores.

Internal Consistency of the Tool

The internal consistency of the MC-CAT was obtained by calculating the Cronbach α coefficient [24]. The rank of the response selected by the participant (ranging from 1=most consistent with MC to 5=least consistent with MC) for each of the response choices per case (A, B, C, and D) and for the entire MC-CAT assessment (all responses over 4 cases) were used to calculate the coefficient. Thus, this analysis aimed to determine whether the responses chosen by the participants were consistent across the 4 cases. An acceptable score for the Cronbach α coefficient is between .70 and .95 [24,25], which was adopted as our target criterion for moving forward with the tool.

Results

Step 1: Development of the Patient Cases and Scoring Scheme

We received 154 surveys, of which 80 (52%) physicians had complete data (n=22, 28% cardiologists; n=22, 28% respirologists; n=15, 19% internists; and n=21, 26% general practitioners) and were included in the analyses. The mean age was 49 (SD 12.9) years. Of the 80 physicians, 44 (55%) were male, and 50 (63%) identified French as the first language. The mean duration of practice of the physicians was 18 (SD 11.9) years, and 69% (55/80) of physicians were working in a university hospital setting and had a mean of 38 (SD 24.2) weekly NCD consultations (Table 1 presents the participants' information).

The health risk behaviors most frequently identified by physicians were physical inactivity (75/80, 94%), smoking (73/80, 91%), medication nonadherence (71/80, 89%), and unhealthy diet (69/80, 86%). Physicians ranked smoking first, medication nonadherence second, physical inactivity third, and unhealthy diet fourth in the list of most important health risk behaviors to address in the context of NCD management. The following health risk behaviors were perceived as the most prevalent among their patients: (1) physical inactivity (mean 58%, SD 19.7%), (2) unhealthy diet (mean 47.2%, SD 18.2%), and (3) difficulty in managing stress (mean 44.8%, SD 18.9%). On the basis of these results, we designed 4 core cases of the MC-CAT to feature cases with health risk behaviors—smoking, physical inactivity, nonadherence to medication, and poor diet—representing a range of NCDs (eg, obesity, asthma, diabetes, and hypertension) [26]. The average number of opportunities was 11.4 (SD 6.4, range 2-24).

Step 2: Case and Scoring Validation by International Experts—Content Validity

The initial percentage of agreement between our classification and the experts' for the rank order of responses across all 4 base cases was 60.9% (SD 14.0%; range 37.1%-84.3%). The competency identification agreement across all 4 base cases was 44.9% (SD 8.4%; range 30.5%-60.2%). In response to these results and considering the specific feedback provided by our international experts (74 comments over the 4 cases), we made 8 modifications to the rank ordering of statements and 23 modifications to aspects of the dialog (eg, making a statement more or less consistent with MC; Table 3). The experts were then asked to assess whether they agreed with the new rankings of the modified cases and the competencies identified. After this evaluation, an increase was noted in agreement for both the rank order (mean 87.6%, SD 16.8%; range 16.7%-100%) and the competency identification (mean 78.1%, SD 14.3%; range 48.6%-100%), which is considered as an acceptable level of agreement [27].

Table 3. Percentage agreement of the rank order of responses across all 4 base cases.

Choice of response for each case	Agreement			Agreement after modification		
	Good	Acceptable	Poor	Good	Acceptable	Poor
Case A, (%)						
1	57.1	25.7	17.1	78.3	12.9	10
2	58.6	37.1	4.3	70	25.7	4.3
3	37.1	38.6	24.3	73.3	26.7	0
4	70	27.1	2.9	80	20	0
5	60	30	10	86.7	13.3	0
6	45.7	40	14.3	70	28.3	1.7
Mean (SD)	54.8 (11.6)	33.1 (6.2)	12.2 (8.1)	76.4 (6.5)	21.1 (6.8)	2.7 (4.0)
Case B, (%)						
1	47.1	27.1	25.7	80	15	3.3
2	57.1	27.1	15.7	91.7	6.7	1.7
3	78.6	18.6	2.9	78.3	20	0
4	65.7	30	4.3	85	11.7	1.7
5	45.7	48.6	5.7	86.7	13.3	0
6	68.6	28.6	2.9	81.7	16.7	1.7
Mean (SD)	60.5 (12.9)	30 (9.9)	9.5 (9.3)	83.9 (4.9)	13.9 (4.5)	1.4 (1.2)
Case C, (%)						
1	78.6	15.7	4.3	81.7	13.3	3.3
2	45.7	44.3	10	86.7	6.7	6.7
3	47.1	40	12.9	80	16.7	3.3
4	74.3	22.9	2.9	70	26.7	3.3
5	78.6	17.1	4.3	81.7	13.3	5
6	84.3	14.3	1.4	86.7	13.3	0
7	57.1	31.4	11.4	80	20	0
Mean (SD)	66.5 (16.2)	26.5 (12.2)	6.7 (4.6)	81.0 (5.6)	15.7 (6.3)	3.1 (2.4)
Case D, (%)						
1	73.3	18.3	8.3	73.3	18.3	8.3
2	43.3	53.3	3.3	90	10	0
3	76.7	20	3.3	76.7	20	3.3
4	73.3	23.3	3.3	93.3	6.7	0
5	60	26.7	13.3	93.3	6.7	0
6	48.3	40	8.3	81.6	16.7	1.7
Mean (SD)	62.5 (14.2)	30.3 (13.7)	6.6 (4.1)	84.7 (8.7)	13.1 (6.0)	2.2 (3.3)
Overall agreement, mean (SD)	61.2 (13.8)	29.8 (10.5)	8.7 (6.8)	81.5 (6.9)	15.8 (6.4)	2.5 (2.8)

Steps 3 and 4: Finalization of the Cases, Scoring Scheme, and Integration Into the Web-Based MC-CAT Platform

As part of the log-in process, HCPs were asked to enter basic demographic information, including language preference, age, sex, and specialty, the latter of which was used to present HCPs with cases in their area of practice. Each case began by presenting respondents with relevant patient information (age, sex, diagnosis, and basic clinical information) in a *file* located at the top right-hand corner of the screen and accessible anytime during the assessment (Figure 2). The physician was then informed of the behavioral target (eg, increasing physical activity) and instructed to engage the patient in a conversation about changing their behavior. The case always started with a patient expressing ambivalence about health behavior changes.

The physician was then directed to select a response from 1 of 5 randomly ordered options (Figure 1 provides a visual example of a case). Each response corresponded to an opportunity to demonstrate ≥1 of the 11 core MC competencies, which are assessed multiple times per case and across cases and averaged to obtain a score for that competency (Table 4 presents the distribution of the competencies for each base case).

Of the 27 responses received during user experience testing, 10 (37%) comments reflected audio and visual elements (eg, synchronization of the voice with the appearance of the text and the mouth movement of the virtual patient), and 19 (70%) comments reflected instruction elements (eg, lack of clarity in sections of the consent form; typos). On the basis of this, the 29 comments were addressed by the research team and 42 Comets Inc when creating a new version of the MC-CAT platform (version 2.0).

Table 4. Distribution of competencies for each of the 4 core cases.

MC-CAT[a] competency	Possibility of expressing target behavior					Example
	Case A: physical activity, n (%)	Case B: smoking cessation, n (%)	Case C: healthy diet, n (%)	Case D: medication adherence, n (%)	Total, N	
Reflective listening	3 (19)	4 (25)	5 (31)	4 (25)	16	"So you recognize the potential benefits of a healthier diet, but it's challenging given your line of work."
Expressing empathy	2 (20)	3 (30)	3 (30)	2 (20)	10	"Changing your daily eating habits when there are barriers can be challenging. But exploring the benefits may help."
Eliciting "change-talk" or evocation	2 (22)	2 (22)	3 (33)	2 (22)	9	"You said you were fed up with feeling breathless, and recognize that smoking might be the cause. What would increase your confidence in your ability to quit?"
Responding to resistance	2 (33)	2 (33)	0 (0)	2 (33)	6	"It might help to know the benefits of exercise. Tell me what you think you would be able to do if you were in better shape?"
Goal setting	2 (29)	1 (14)	2 (29)	2 (29)	7	"Cooking would be a great place to start! And if it's something you enjoy, you are more likely to stick with it. What is your plan to get started?"
Demonstrating acceptance, tolerance, and respect	4 (31)	2 (15)	4 (31)	3 (23)	13	"It sounds like a great plan, and your willingness to getting more information this weekend demonstrates how important this is to you."
Being collaborative	2 (25)	2 (25)	1 (13)	3 (38)	8	"It sounds like we just need to find a routine that works for you. Could we explore some options together?"
(Not) expressing hostility or impatience	1 (9)	4 (36)	5 (45)	1 (9)	11	"If you want to avoid exacerbating your diabetes, you need to commit to a diet change, sooner rather than later."
(Not) negatively judging or blaming	5 (25)	6 (30)	4 (20)	5 (25)	20	"I think that's a good place to start, all you need to do is follow through."
(Not) being argumentative or confrontational	6 (25)	6 (25)	6 (25)	6 (25)	24	"Yes, but since you lack confidence you should also get behavioral counselling, you don't want to fail again!"
Providing information neutrally	0 (0)	1 (50)	0 (0)	1 (50)	2	"There are several options: nicotine replacement therapy, medications, and behavioral counselling have all been shown to be effective. What do you think would work best for you?"
Total	29 (23)	33 (26)	33 (26)	31 (25)	126	—[b]
Exchanges	6 (24)	6 (24)	7 (28)	6 (24)	25	—

[a]MC-CAT: Motivational Communication Competency Assessment Test.

[b]Data not available.

Step 5: Preliminary Internal Validity Analyses

We received 24 MC-CAT responses from undergraduate allied HCP students. All participants had completed both assessments (17/24, 71% female; living in the province of Quebec, Canada; English speaking; with 0-5 years of practice in BCC and no previous training in MC).

Case Consistency Analyses

To identify possible differences between the different versions of the base cases (cases A1 to A4, B1 to B4, C1 to C4, and D1 to D4), ANOVAs were performed by comparing the competency and ranking scores. No significant differences were identified between the different versions of the case across the 2 measurement times (Figures 3 and 4; Table 5).

Figure 3. Score differences between each case version for time 1 (N=24).

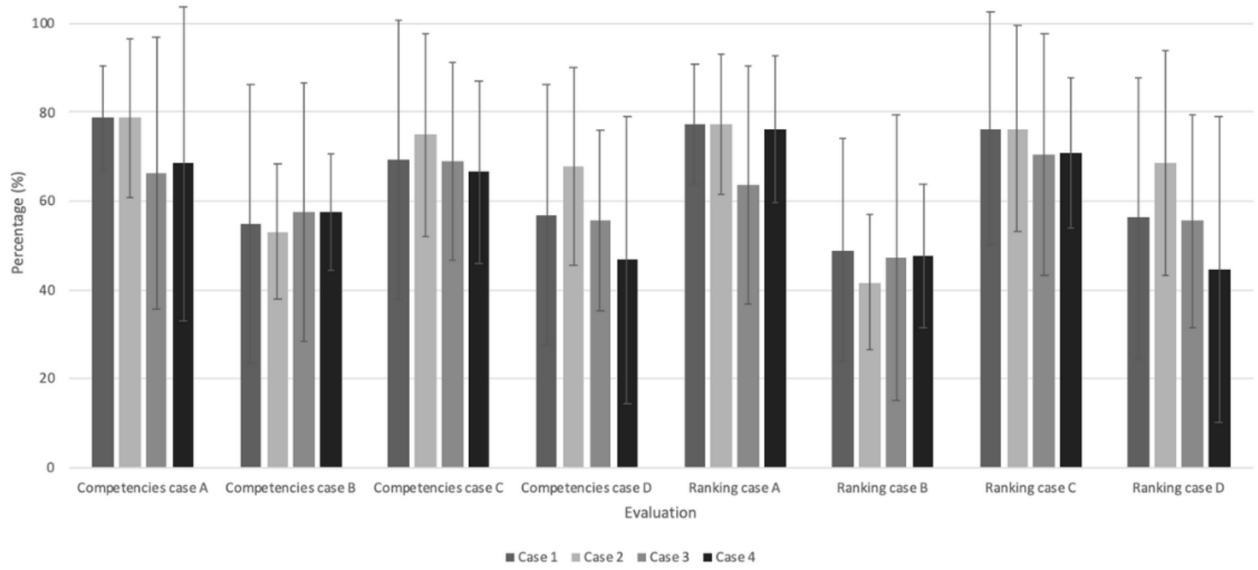


Figure 4. Score differences between each case version for time 2 (N=24).

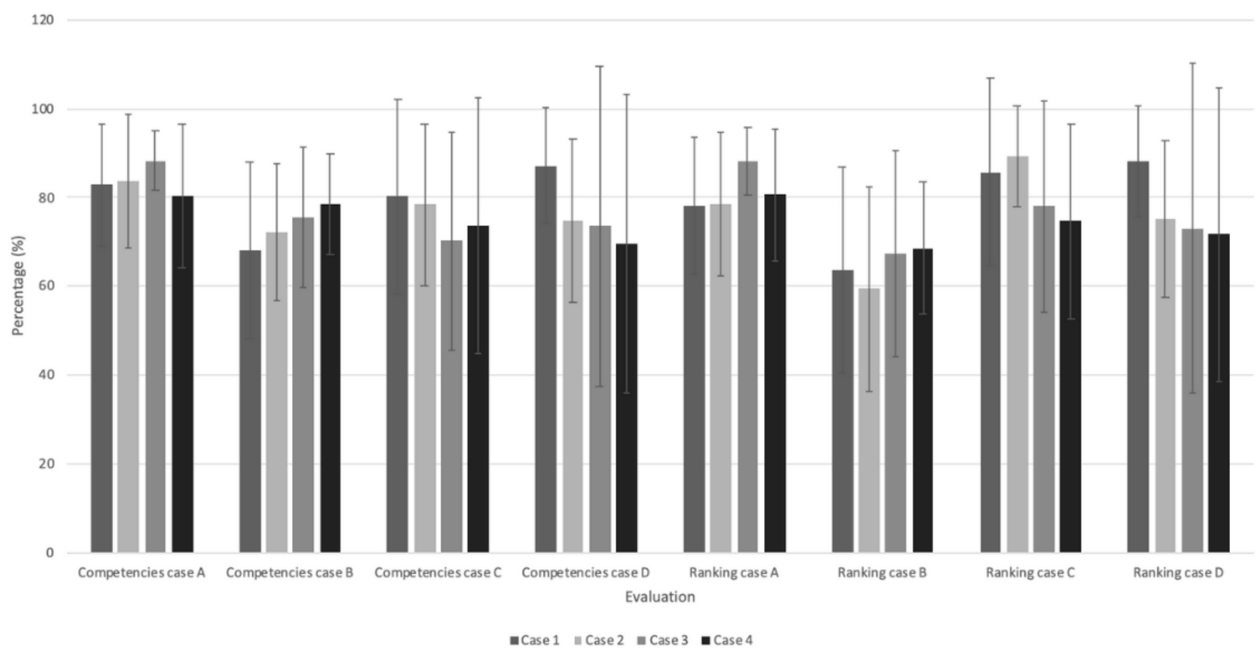


Table 5. Global competency and ranking weighted scores at assessments 1 and 2 (N=24).

Cases	All (%), mean (SD)	Case 1 (%), mean (SD)	Case 2 (%), mean (SD)	Case 3 (%), mean (SD)	Case 4 (%), mean (SD)	Difference between cases, *P* value
Precourse						
Competency (global competency score 64.9, SD 19.0)						
Case A	73.1 (25.6)	78.7 (11.6)	78.7 (18.0)	66.3 (30.6)	68.5 (35.4)	.51
Case B	55.8 (24.2)	54.9 (31.4)	53.1 (15.3)	57.5 (29.1)	57.5 (13.1)	.97
Case C	70.0 (32.3)	69.4 (31.5)	75.0 (22.8)	68.9 (22.1)	66.6 (20.6)	.82
Case D	56.8 (27.4)	56.9 (29.2)	67.9 (22.2)	55.5 (20.4)	46.8 (32.4)	.26
Ranking (global ranking score 65.8, SD 17.4)						
Case A	73.6 (18.9)	77.3 (13.5)	77.4 (15.8)	63.6 (26.9)	76.3 (16.6)	.24
Case B	46.4 (23.2)	49.0 (25.2)	41.7 (15.2)	47.2 (32.0)	47.7 (16.3)	.89
Case C	73.5 (22.6)	76.4 (26.1)	76.3 (23.3)	70.4 (27.2)	70.9 (16.9)	.87
Case D	56.3 (30.1)	56.2 (31.6)	68.6 (25.3)	55.6 (23.9)	44.6 (34.4)	.24
Postcourse						
Competency (global competency score 77.6, SD 16.5;*P*<.001)						
Case A	83.7 (13.3)	82.8 (13.7)	83.6 (15.0)	88.2 (6.6)	80.4 (16.1)	.54
Case B	73.3 (15.9)	68.2 (19.8)	72.0 (15.4)	75.5 (15.8)	78.4 (11.4)	.50
Case C	75.3 (23.5)	80.1 (21.8)	78.3 (18.2)	70.2 (24.5)	73.5 (28.7)	.68
Case D	77.7 (26.5)	87.2 (12.8)	74.6 (18.4)	73.5 (36.2)	69.6 (33.6)	.32
Ranking (global ranking score 77.8, SD 16.6;*P*<.001)						
Case A	81.3 (14.1)	78.2 (15.4)	78.5 (16.1)	88.2 (7.5)	80.6 (14.8)	.26
Case B	64.3 (21.6)	63.6 (23.4)	59.4 (23.1)	67.3 (23.2)	68.5 (14.9)	.72
Case C	81.5 (21.1)	85.7 (21.1)	89.3 (11.5)	78.1 (23.8)	74.7 (22.0)	.37
Case D	78.4 (26.5)	88.2 (12.5)	75.0 (17.7)	73.1 (37.0)	71.7 (32.9)	.31

Internal Consistency of the Tool

The MC-CAT tool showed acceptable values of internal consistency for global scores (25 items) at both time 1 (α=.78) and time 2 (α=.80).

Discussion

Principal Findings

The objectives of this study were to develop the MC-CAT, a new, web-based, *user-friendly* tool for assessing communication competencies among HCPs in the context of changing health behaviors among patients with NCDs, conduct an initial internal validity assessment, and evaluate score consistency between the base and modified cases. The MC-CAT was designed to simulate clinical interactions with *virtual* patients to provide both global and specific scores for 11 core communication skills [16]. It was co-designed in collaboration with key stakeholders (ie, physicians, HCPs, researchers, and health care administrators) using an iKT approach to ensure its clinical relevance and feasibility for use in practice. The web-based platform was also user-tested to ensure ease of navigability among target users.

The results of this 5-step mixed methods study indicate that the MC-CAT demonstrates acceptable levels of internal consistency for the global competency score (α=.78-.80), and little variance was found across different versions of the 4 base cases. This level of internal consistency is higher than the levels observed in many existing communication assessment tools such as the Pediatric Consultation Assessment Tool and the Four Habits Coding Scheme, which had Cronbach α values between .52 [28] and .66 [29]. However, it was slightly lower than the levels seen in other tools (eg, the Council of Emergency Medicine Residency Directors Standardized Direct Observation Assessment Tool [30] or the Doctors' Observable Use of Self-Efficacy Enhancing Interviewing Techniques measure [31], with Cronbach α=.93 and .94, respectively). Several factors can affect the results of a tool's internal consistency analysis (also considered a measure of scale reliability), such as the number of participants included in the analysis (varying between 19 and 82 participants for these 4 tools, in contrast to 32 participants for our analyses), the potential for evaluation biases associated with assessment methods (eg, self-report surveys and observational scales vs an objective scoring algorithm), and the number of items (varying between 10 and 26 items for these 4 tools). However, using a rigorous development process, we have developed an assessment tool that met our internal validity criterion (ie, α between .70 and .95), and we are satisfied that we can move forward with internal and external validation among HCPs.

Comparison With Prior Work

To our knowledge, this is the first study to describe the development of a web-based interactive communication skills assessment tool for HCPs [17]. The MC-CAT tool was also developed in parallel with a theory-driven and evidence-informed MC training framework for HCPs [16]. As such, the MC-CAT addresses the shortcomings of approximately 50% of the 45 existing communication competency assessment tools [17], which our recent review revealed were not informed by established theories of communication or behavior change. Furthermore, most tools did not explicitly define the communication competencies they were designed to assess [17], unlike MC-CAT, which was specifically designed to assess the 11 core communication competencies of MC, which HCPs and behavior change experts identified as being the most critical for changing health behaviors in the context of NCD management [16]. We developed a tool to assess MC competencies [16] as MC has become an increasingly popular communication style among HCPs [32,33]. The fact that MC-CAT assesses all MC competencies and not a subset of these skills also overcomes the limitations of previous tools that have not been developed to provide comprehensive assessments of specific communication frameworks [17]. The total scores on the MC-CAT also reflected the relative importance of each individual competency proportional to its use in practice. In other words, communication skills that are used more frequently during patient consultations (eg, asking open *evocative* questions, reflective listening, and expressing empathy) are given greater weight in the final scoring. This is an important strength of the tool, as we are aware of no existing tools that take into account the *real-world* frequency with which certain skills are used in their scoring algorithms. We also ensured that the different versions of our 4 base cases (cases A1 to A4, B1 to B4, C1 to C4, and D1 to D4) were comparable in terms of competency and ranking scores. This means that our adaptations were consistent with the original scoring algorithm and that we can use them to create further adaptations of the 4 base cases to further extend the case bank of the tool.

One of the most useful and attractive features of the MC-CAT is that it is scored automatically based on a preprogrammed algorithm, which eliminates the need for external raters to conduct assessments (ie, interrater reliabilities). The need for external, trained raters is a feature of all existing assessment tools [17]. Although more rigorous than self-reported assessments, a manual rating is associated with significant costs in terms of time and complexity. The fact that the MC-CAT is scored automatically also reduces the potential biases associated with the subjective nature of rater assessments and eliminates the need for multiple raters to assess agreement, which removes time and complexity. Indeed, in previous studies using external raters to assess physician communication competencies, training time averaged 14 hours and ranged from 1.5 to >90 hours [17], which may not be feasible to implement in practice. Most previous studies (61%) also failed to standardize the training of external raters [17], which can greatly affect the fidelity of the coding process and does not allow for the comparison of one evaluation with another. Finally, the MC-CAT was specifically designed to address the practical constraints of many NCD-focused physicians who may not have the time to undergo complex evaluations [11]. The MC-CAT tool is completed on the web using any electronic device and takes between 15 and 20 minutes to complete, which are features that our HCP collaborators have indicated as both acceptable and feasible.

Study Limitations and Strengths

First, this study may be limited by the fact that we did not specifically include patients with NCDs as part of our stakeholder groups. The rationale was that our target users were HCPs; hence, our focus was on engaging various physicians, HCPs, and health administrator stakeholders. There is already an evidence base demonstrating that patients whose physicians use MC-type approaches feel more understood, have more trust in their providers, are more adherent to treatment, are more satisfied with their care, and have better outcomes [10,34-38]. As such, our goal was not to validate this work but rather to focus on how to facilitate the implementation of these approaches into practice. Second, although the MC-CAT includes a range of cases that are intended to reflect real-world clinical encounters, it was not possible to create cases that reflected all behavioral issues involved in these diseases, which may limit the generalizability of the tool. Similarly, we attempted to include cases that reflected patient diversity in terms of sociodemographic characteristics (ie, age, sex, and race or culture); however, it was not possible to include all combinations and permutations of these characteristics, which may be seen as a limitation. However, now that we have validated the scoring integrity of our 4 base cases and their alternate versions, our next step is to create additional adaptations that will increase the heterogeneity of our case bank. Finally, the MC-CAT relies on computer and internet access, which may not be readily available to some providers.

Despite these limitations, this study also has several notable strengths. Critical to successfully developing a valid and reliable assessment tool, we integrated several key stakeholder groups in all steps of the development process (eg, content, testing, and recruitment) using the knowledge transfer cycle as the basis of our iKT strategy [18]. This is expected to optimize the uptake and impact of assessment tools in clinical practice and research. In addition, we did not neglect the design of the web-based platform and conducted careful user-testing using the User Experience Framework [39] to assess appeal, clarity, and navigability. This framework provides several dimensions to consider when designing and testing a web-based tool, presented on a continuum from abstract to concrete regarding visual design, interaction design, and functional specifications. Through this process, we further refined the tool to make it more *user-friendly* and intuitive. Another strength of our development process is that we created multiple versions of our base cases and tested whether they were comparable using a factorial analysis, which revealed no significant differences between cases on competency and ranking scores. This increases our confidence that the MC-CAT is now ready for comprehensive psychometric property analyses among target users (ie, HCPs: nurses and physicians), which is the next step in the development process.

Conclusions

The MC-CAT is a new web-based, interactive, user-friendly MC assessment tool that was codeveloped with a range of relevant stakeholders. The results demonstrated acceptable internal consistency for global competency scores, which indicates that it is ready for more comprehensive psychometric property analyses, including both internal and external validity tests (eg, positive and negative predictive values and convergent validity) in a national sample of HCPs across disciplines. We will continue to use an iterative approach during subsequent phases of development, and we are prepared to further refine the tool and its scoring algorithm as needed. Once developed, the MC-CAT will be the first web-based MC assessment tool that can be easily and widely accessed by a variety of HCPs and can be used not only as an evaluation tool but also as an adjunct to the MC training programs. Its accessibility, convenience, and *user-friendliness* are expected to increase the uptake and improve the quality of MC training programs designed to improve HCPs' ability to effectively motivate and support patients to adopt healthy behaviors in the context of NCD prevention and management.

Acknowledgments

The authors would like to thank the knowledge users (physicians, health care providers, researchers, and health care administrators) for their participation and time in the different steps of the development process. The authors would like to thank Dr Justin Presseau, Dr Robert D Reid, Dr Angela Pfammatter, Dr Anne H Berman, Dr Jill Chorney, Dr Marie Johnston, Dr John Kayser, Dr Catherine Laurin, and Dr David Ogez for their motivational communication and behavior change expertise.

Conflicts of Interest

None declared.

References

1. Fact sheets: noncommunicable diseases. World Health Organization. URL: https://www.who.int/topics/noncommunicable_diseases/factsheets/en/ [accessed 2021-03-01]
2. Noncommunicable diseases: key facts. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2021-03-01]
3. Overview - Preventing chronic diseases: a vital investment. World Health Organization. 2015. URL: https://www.who.int/chp/chronic_disease_report/part1/en/ [accessed 2021-03-01]
4. Marmot M, Bell R. Social determinants and non-communicable diseases: time for integrated action. BMJ 2019 Jan 28;364:l251 [FREE Full text] [doi: 10.1136/bmj.l251] [Medline: 30692093]
5. Elder JP, Ayala GX, Harris S. Theories and intervention approaches to health-behavior change in primary care. Am J Prev Med 1999 Nov;17(4):275-284. [doi: 10.1016/s0749-3797(99)00094-x] [Medline: 10606196]
6. Ubel PA, Scherr KA, Fagerlin A. Empowerment failure: how shortcomings in physician communication unwittingly undermine patient autonomy. Am J Bioeth 2017 Nov;17(11):31-39 [FREE Full text] [doi: 10.1080/15265161.2017.1378753] [Medline: 29111936]
7. Ledford CJ, Villagran MM, Kreps GL, Zhao X, McHorney C, Weathers M, et al. "Practicing medicine": patient perceptions of physician communication and the process of prescription. Patient Educ Couns 2010 Sep;80(3):384-392. [doi: 10.1016/j.pec.2010.06.033] [Medline: 20675096]
8. Stott NC, Pill RM. 'Advise yes, dictate no'. Patients' views on health promotion in the consultation. Fam Pract 1990 Jun;7(2):125-131. [doi: 10.1093/fampra/7.2.125] [Medline: 2369980]
9. Bock C, Diehl K, Schneider S, Diehm C, Litaker D. Behavioral counseling for cardiovascular disease prevention in primary care settings: a systematic review of practice and associated factors. Med Care Res Rev 2012 Oct;69(5):495-518. [doi: 10.1177/1077558712441084] [Medline: 22457269]
10. Dragomir AI, Julien CA, Bacon SL, Boucher VG, Lavoie KL, Canadian Network for Health Behavior ChangePromotion (CAN-Change). Training physicians in behavioural change counseling: a systematic review. Patient Educ Couns 2019 Jan;102(1):12-24. [doi: 10.1016/j.pec.2018.08.025] [Medline: 30172573]
11. Adler KG. The unexpected challenge of motivational interviewing. Fam Pract Manag 2016;23(5):9 [FREE Full text] [Medline: 27626112]
12. Miller W, Amrhein P, Yahne C, Tonigan J. A motivational interviewing failure: hazards of a therapy manual. University of New Mexico Center on Alcoholism, Substance Abuse and Addictions (CASAA). URL: https://motivationalinterviewing.org/sites/default/files/Hazards%20of%20a%20Therapy%20Manual.pdf [accessed 2021-03-01]
13. Deci E, Ryan R. Intrinsic Motivation and Self-Determination in Human Behavior. Cham: Springer; 1985.
14. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
15. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]

16. Dragomir AI, Boucher VG, Bacon SL, Gemme C, Szczepanik G, Corace K, et al. An international Delphi consensus study to define motivational communication in the context of developing a training program for physicians. Transl Behav Med 2021 Mar 16;11(2):642-652 [FREE Full text] [doi: 10.1093/tbm/ibaa015] [Medline: 32145022]
17. Boucher VG, Gemme C, Dragomir AI, Bacon SL, Larue F, Lavoie KL. Evaluation of communication skills among physicians: a systematic review of existing assessment tools. Psychosom Med 2020 Mar 11;82(4):440-451. [doi: 10.1097/psy.0000000000000794]
18. Gagliardi AR, Berta W, Kothari A, Boyko J, Urquhart R. Integrated knowledge translation (IKT) in health care: a scoping review. Implement Sci 2016 Mar 17;11(1):38 [FREE Full text] [doi: 10.1186/s13012-016-0399-1] [Medline: 26988000]
19. Kothari A, Wathen CN. Integrated knowledge translation: digging deeper, moving forward. J Epidemiol Community Health 2017 Jun 15;71(6):619-623. [doi: 10.1136/jech-2016-208490] [Medline: 28298415]
20. Méthodes Qualitatives, Quantitatives Et Mixtes, 2 edition: Dans la recherche en sciences humaines, sociales et de la santé. Quebec , Canada: PU QUEBEC; 2021.
21. Recker J, Rosemann M. A measurement instrument for process modeling research: development, test and procedural model. Scand J Info Syst 2010;22(2):1-28 [FREE Full text]
22. MacKenzie SB, Podsakoff PM, Podsakoff NP. Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques. MIS Q 2011;35(2):293. [doi: 10.2307/23044045]
23. Ozolins U, Hale S, Cheng X, Hyatt A, Schofield P. Translation and back-translation methodology in health research - a critique. Expert Rev Pharmacoecon Outcomes Res 2020 Feb;20(1):69-77. [doi: 10.1080/14737167.2020.1734453] [Medline: 32089017]
24. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55 [FREE Full text] [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]
25. Bujang MA, Omar ED, Baharum NA. A review on sample size determination for Cronbach's alpha test: a simple guide for researchers. Malays J Med Sci 2018 Nov;25(6):85-99 [FREE Full text] [doi: 10.21315/mjms2018.25.6.9] [Medline: 30914882]
26. Boucher V, Gemme C, Dragomir A, Szczepanik G, Bacon S, Lavoie K. The development of an evaluation tool for assessing the acquisition of motivational communication skills among physicians using an integrated KT approach. Ann Behav Med 2019;53(3):303.
27. Stemler S. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. Pract Assess Res Eval 2004;9(4). [doi: 10.7275/96jp-xz07]
28. Howells RJ, Davies HA, Silverman JD, Archer JC, Mellon AF. Assessment of doctors' consultation skills in the paediatric setting: the paediatric consultation assessment tool. Arch Dis Child 2010 May;95(5):323-329. [doi: 10.1136/adc.2008.146191] [Medline: 19019880]
29. Krupat E, Frankel R, Stein T, Irish J. The four habits coding scheme: validation of an instrument to assess clinicians' communication behavior. Patient Educ Couns 2006 Jul;62(1):38-45. [doi: 10.1016/j.pec.2005.04.015] [Medline: 15964736]
30. Shayne P, Gallahue F, Rinnert S, Anderson CL, Hern G, Katz E, CORD SDOT Study Group. Reliability of a core competency checklist assessment in the emergency department: the standardized direct observation assessment tool. Acad Emerg Med 2006 Jul;13(7):727-732 [FREE Full text] [doi: 10.1197/j.aem.2006.01.030] [Medline: 16636361]
31. Jerant A, Hanson B, Kravitz RL, Tancredi DJ, Hanes E, Grewal S, et al. Detecting the effects of physician training in self-care interviewing skills: coding of standardized patient (SP) visit recordings versus SP post-visit ratings. Patient Educ Couns 2017 Feb;100(2):367-371 [FREE Full text] [doi: 10.1016/j.pec.2016.08.021] [Medline: 27578271]
32. Rouleau C, Lavoie K, Bacon S, Vallis M, Corace K, Campbell T. Training healthcare providers in motivational communication for promoting physical activity and exercise in cardiometabolic health settings: do we know what we are doing? Curr Cardiovasc Risk Rep 2015 Apr 19;9(6):185-190. [doi: 10.1007/s12170-015-0457-2] [Medline: 11778978]
33. Dewez S, Laurin C, Ogez D, Bourque C, Curnier D, Laverdière C, et al. Elaboration and refinement of a motivational communication training program for healthcare professionals in pediatric oncology: a feasibility and acceptability study. Health Psychol Behav Med 2021 Mar 24;9(1):220-238 [FREE Full text] [doi: 10.1080/21642850.2021.1903326] [Medline: 34104558]
34. Boucher VG, Pelaez S, Gemme C, Labbe S, Lavoie KL. Understanding factors associated with vaccine uptake and vaccine hesitancy in patients with rheumatoid arthritis: a scoping literature review. Clin Rheumatol 2021 Feb;40(2):477-489. [doi: 10.1007/s10067-020-05059-7] [Medline: 32621081]
35. Abdul-Razzak A, Sherifali D, You J, Simon J, Brazil K. 'Talk to me': a mixed methods study on preferred physician behaviours during end-of-life communication from the patient perspective. Health Expect 2016 Aug;19(4):883-896 [FREE Full text] [doi: 10.1111/hex.12384] [Medline: 26176292]
36. Hanlon M, Hogan M, Durand H, Pilch M, Harney O, Molloy G, et al. Designing an e-learning tool to support health practitioners caring for patients taking multiple medications. HRB Open Res 2020;3:59 [FREE Full text] [doi: 10.12688/hrbopenres.13110.2] [Medline: 33954278]
37. Jacob A, Moullec G, Lavoie KL, Laurin C, Cowan T, Tisshaw C, et al. Impact of cognitive-behavioral interventions on weight loss and psychological outcomes: a meta-analysis. Health Psychol 2018 May;37(5):417-432. [doi: 10.1037/hea0000576] [Medline: 29698017]

38. Lavoie KL, Sedeno M, Hamilton A, Li P, De Sousa D, Troosters T, et al. Behavioural interventions targeting physical activity improve psychocognitive outcomes in COPD. ERJ Open Res 2019 Oct;5(4):00013-2019 [FREE Full text] [doi: 10.1183/23120541.00013-2019] [Medline: 31720294]
39. Garrett J. The Elements of User Experience: User-centered Design for the Web and Beyond - Voices That Matter. London, United Kingdom: Pearson Education; 2002.

Abbreviations

BCC: behavior change counseling
HCP: health care provider
iKT: integrated knowledge translation
MC: motivational communication
MC-CAT: Motivational Communication Competency Assessment Test
NCD: noncommunicable disease

Edited by T Leung; submitted 29.06.21; peer-reviewed by KS Gonçalves, Z Su; comments to author 22.01.22; revised version received 25.01.22; accepted 12.02.22; published 24.06.22.

Please cite as:
Gosselin Boucher V, Bacon S, Voisard B, Dragomir AI, Gemme C, Larue F, Labbé S, Szczepanik G, Corace K, Campbell T, Vallis M, Garber G, Rouleau C, Diodati JG, Rabi D, Sultan S, Lavoie K, Network For Health Behavior Change And Promotion (CAN-Change)
Assessing Physician's Motivational Communication Skills: 5-Step Mixed Methods Development Study of the Motivational Communication Competency Assessment Test
JMIR Med Educ 2022;8(2):e31489
URL: https://mededu.jmir.org/2022/2/e31489
doi:10.2196/31489
PMID:35749167

Original Paper

Digital Health and Learning in Speech-Language Pathology, Phoniatrics, and Otolaryngology: Survey Study for Designing a Digital Learning Toolbox App

Yuchen Lin[1], MSc, CCC-SLP; Martin Lemos[2], MA; Christiane Neuschaefer-Rube[1], MD

[1]Clinic for Phoniatrics, Pedaudiology & Communication Disorders, University Hospital and Medical Faculty, Rheinisch-Westfaelische Technische Hochschule Aachen, Aachen, Germany

[2]Audiovisual Media Center (AVMZ), University Hospital and Medical Faculty, Rheinisch-Westfaelische Technische Hochschule Aachen, Aachen, Germany

Corresponding Author:
Yuchen Lin, MSc, CCC-SLP
Clinic for Phoniatrics, Pedaudiology & Communication Disorders
University Hospital and Medical Faculty, Rheinisch-Westfaelische Technische Hochschule Aachen
Pauwelsstrasse 30
Aachen, 52074
Germany
Phone: 49 241 80 88954
Email: yuchen.lin@rwth-aachen.de

Abstract

Background: The digital age has introduced opportunities and challenges for clinical education and practice caused by infinite incoming information and novel technologies for health. In the interdisciplinary field of communication sciences and disorders (CSD), engagement with digital topics has emerged slower than in other health fields, and effective strategies for accessing, managing, and focusing on digital resources are greatly needed.

Objective: We aimed to conceptualize and investigate preferences of stakeholders regarding a *digital learning toolbox*, an app containing a library of current resources for CSD. This cross-sectional survey study conducted in German-speaking countries investigated professional and student perceptions and preferences regarding such an app's features, functions, content, and associated concerns.

Methods: An open web-based survey was disseminated to professionals and students in the field of CSD, including speech-language pathologists (SLPs; *German: Logopäd*innen*), speech-language pathology students, phoniatricians, otolaryngologists, and medical students. Insights into preferences and perceptions across professions, generations, and years of experience regarding a proposed app were investigated.

Results: Of the 164 participants, an overwhelming majority (n=162, 98.8%) indicated readiness to use such an app, and most participants (n=159, 96.9%) perceived the proposed app to be helpful. Participants positively rated app functions that would increase utility (eg, tutorial, quality rating function, filters based on content or topic, and digital format); however, they had varied opinions regarding an app community feature. Regarding app settings, most participants rated the option to share digital resources through social media links (144/164, 87.8%), receive and manage push notifications (130/164, 79.3%), and report technical issues (160/164, 97.6%) positively. However, significant variance was noted across professions (H_3=8.006; P=.046) and generations (H_3=9.309; P=.03) regarding a username-password function, with SLPs indicating greater perceived usefulness in comparison to speech-language pathology students (P=.045), as was demonstrated by Generation X versus Generation Z (P=.04). Participants perceived a range of clinical topics to be important; however, significant variance was observed across professions, between physicians and SLPs regarding the topic of diagnostics (H_3=9.098; P=.03) and therapy (H_3=21.236; P<.001). Concerns included technical challenges, data protection, quality of the included resources, and sustainability of the proposed app.

Conclusions: This investigation demonstrated that professionals and students show initial readiness to engage in the co-design and use of an interdisciplinary digital learning toolbox app. Specifically, this app could support effective access, sharing, evaluation, and knowledge management in a digital age of rapid change. Formalized digital skills education in the field of CSD is just a part of the solution. It will be crucial to explore flexible, adaptive strategies collaboratively for managing digital resources and tools to optimize targeted selection and use of relevant, high-quality evidence in a world of bewildering data.

(JMIR Med Educ 2022;8(2):e34042) doi:10.2196/34042

KEYWORDS

digital learning; mLearning; mHealth; speech-language pathology; phoniatrics; otolaryngology; communication disorders; mobile phone

Introduction

Background

Mobile devices are rapidly revolutionizing the means of communication, learning, and health care. Since the emergence of new technologies and devices such as smartphones, tablets, PDAs, smartwatches, and laptop computers among others in the past decades, the adoption of such technologies in the health sciences and medical education has been increasingly explored [1-3]. Specifically, the use of mobile devices for patient care, both by patients (eg, health apps) and by clinical professionals (eg, patient monitoring tools, telemedicine, and teletherapy) and for research and education purposes (eg, reference apps) has increased [4-6]. This has especially been prevalent considering the COVID-19 pandemic, which has pushed investigation into creating and improving such solutions to the forefront [7-9]. In this context, the term *mobile health* or *mHealth* has emerged to describe the broad spectrum of information-communication technology for medical and public health practices supported by mobile devices, whereas the terms *mobile learning* or *mLearning* have evolved to describe the use of mobile devices to deliver educational content for preclinical, clinical, or specialty training and continuing education or professional development [10,11]. Given that most clinical professionals and students own such mobile technology and the evidence that health care students prefer web-based resources as their primary source of clinical information, the use of mobile medical apps as reference tools is becoming the norm as opposed to an exception [12-14]. As implementation of such technologies and apps continues to increase, it is evident that the future of medicine will inevitably require health professionals to flexibly include media literacy, digital knowledge, and skills into part of their professional scopes of practice.

In the field of communication sciences and disorders (CSD), investigation into digital solutions such as mobile health and mobile learning apps is also becoming popular, although at a slower pace than in other medical fields [15,16]. As an interdisciplinary field concerned with treating the estimated 1 billion people worldwide living with a disability often affecting their speech, language, hearing, voice, or ability to functionally communicate, the field strongly relies on the effective, coordinated efforts of speech-language pathologists (SLPs), phoniatricians, and otolaryngologists among others [17]. In the ever-evolving health care environment, digital solutions especially have the potential to optimize interdisciplinary care and collaboration, which has been identified as a key component to *futureproofing* health care, in other words, designing adaptable solutions for even when technology progresses [18-21]. Thus, exploration of digital resources across disciplines can be useful. In addition to improvements to well-established digital technologies in the field such as augmentative and alternative communication devices, hearing aids, and cochlear implants, mobile technologies are beginning to revolutionize alternative methods of service delivery (eg, telerehabilitation and telepractice) and treatment material (eg, digital therapy or medical apps) and are increasingly empowering patients to engage in their own health management to a greater extent [22-25]. Furthermore, there is some evidence that mobile app technology through smartphones and tablets can improve performance in both speech-language pathology graduate students and otolaryngology and phoniatrics residents when explicitly trained or used as knowledge building (eg, case scenarios, simulations, and question banks) or resource sharing tools [24,26-28]. Although the uncountable and increasing number of apps is impressive, it may be beginning to pose a challenge to meaningful, evidence-based, clinical decision-making and learning [24,29]. In response, digital tools are emerging to provide clinical professionals with faster and easier access to preassessed, evidence-based, psychometrically sound assessments for clinical purposes. For example, the NIH Toolbox for the Assessment of Neurological and Behavioral Function from the National Institutes of Health specifically serves as a comprehensive and portable digitized battery of measures for clinicians to assess and track motor, emotional, sensory, and cognitive functions. By using the benefits of big data, the system can transmit 15,000 data points in ≤7 minutes and has been investigated in >600 studies [30]. Similarly, the PROMIS (Patient-Reported Outcomes Measurement Information System) iPad app, which has been investigated in >2000 studies so far, allows for the monitoring of physical, mental, and social health in adults [31]. Similarly, Torous and Vaidyam [32] designed the mindLAMP (learn, assess, manage, and prevent) mental health app with the intent of more comprehensively addressing multiple user needs; importantly, they emphasized that collaboratively designed comprehensive platforms in the form of a *digital health technology toolbox* can help to eliminate the need for single-purpose apps and could potentially maximize utility and user uptake. However, importantly, such comprehensive, data-backed tools specifically within CSD are scarce.

Moreover, even though such tools are available, the influx of digital resources appears to be undermined because many professionals and students are reportedly unfamiliar with such tools and are not confident in their knowledge and skills pertaining to digital health and clinical resources [33-35]. Although >80% of health professionals surveyed in a European Health Parliament questionnaire reported feeling unprepared for technological developments in health care, 60% of students surveyed across 39 countries similarly felt inadequately trained for the digitalizing health care environment, citing lack of digital skills training and knowledge of digital tools and resources as causes [34,35]. Although professionals and students in CSD have demonstrated interest in increased digital topics in clinical training and continuing education, studies have demonstrated that only approximately 36% to 41% of speech-language pathology academic training programs in the United States explicitly incorporated telepractice apps as part of their curricula, and digital skills have not consistently been an integral component of medical otolaryngology or phoniatric specialty training programs [28,36,37]. Moreover, the increasing number of digital clinical resources and tools is accompanied by a concern of information and data overload, which can make judging the relevance and usefulness of information more difficult. It has been suggested that rather than the issue being an influx in digital information and data, it may be that traditional strategies for managing and evaluating information have not progressed at the same pace as the production of information [38,39]. Furthermore, this can make it difficult to assess the quality of digital clinical resources, many of which have not been peer reviewed or have unknown publishers [40,41].

Objectives

Thus, as digital resources continue to grow exponentially [10,11], it is becoming increasingly clear that professionals and students require digital skills and media literacy training and strategies for sorting through and critically evaluating the quality of digital resources that already exist. Therefore, having an up-to-date library of field-relevant, interdisciplinary, digital learning and therapy tools, which could be collaboratively expanded upon and accessed across multiple platforms or devices in the form of an app, could be useful. Expanding upon the concept of a multiple-use *digital health technology toolbox* previously mentioned by Torous and Vaidyam [32], we proposed that a digital resource library app focused on resource sharing rather than clinical assessment—what we have termed a *digital learning toolbox* (DLT)—could be useful. Specifically, such a tool could help to spark discussion regarding quality assessment, usefulness, and areas of need for existing digital resources. To support the future development of such a digital resource library app with maximized user-centered design, our study aimed to gain insights into the perceptions and preferences of valuable stakeholders, specifically professionals and students in speech-language pathology, phoniatrics, and otolaryngology in German-speaking countries (mainly Germany, Austria, and Switzerland). Specifically, we aimed to determine the interest in such an app and identify potential concerns and desired features, functions, or content through a structured questionnaire, which was disseminated to the speech-language pathology, phoniatrics, and otolaryngology professional and academic communities. Differences across professions, generations, and years of experience were also explored.

Methods

Overview

This survey study was conducted in accordance with the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) guidelines [42]. This survey was the second part of a large survey study. In the first part, knowledge, use, attitudes, and preferences toward digital health and learning of current students and professionals across the interdisciplinary fields of speech-language pathology, phoniatrics, and otolaryngology in the German-speaking countries were investigated. The results of the first section have already been published in a separate article to maximize the depth of analysis [43]. This study focuses separately on professional and student attitudes and preferences regarding a proposed DLT app. The target populations of the proposed app were professionals and students in CSD, including physicians (phoniatricians and otolaryngologists), SLPs, medical students, and speech-language pathology students. The app would serve as an interdisciplinary, collaborative library of open-source digital learning and therapy tools. It would include content relating to anatomy and physiology, pathology, diagnostics, therapy, professional practice issues, and networking. Moreover, this proposed app could include functions such as introductory tutorial; filter functions based on content, language, or source among others; tool rating function; glossary; and app community. Additional settings for increased usability such as a tool sharing function, tool organization function, notification management, and technical error reporting could also be incorporated. To co-design the app to be maximally useful, professionals and students were asked to rate and provide their input on desired content, functions, and settings. Survey screens and a narrative explanation are included in Multimedia Appendix 1.

Ethics Approval

The Ethics Committee of the Medical Faculty at the University Hospital of the Rheinisch-Westfälische Technische Hochschule Aachen University (EK 188/20) determined that the study did not require a full data protection impact assessment as the questionnaire used in the study was fully anonymous. Demographic information regarding profession, years of experience, generation, and sex was collected. Participation was voluntary and could be ended at any time.

Participants and Recruitment

An invitational letter and flyer containing a link to the open survey was shared with professional regulating bodies and university clinical programs in speech-language pathology, phoniatrics, and otolaryngology and with relevant open student and professional groups on Facebook within German-speaking countries. To participate in the survey, participants had to be one of the following: (1) physician in phoniatrics or otolaryngology, (2) SLP, (3) medical student, or (4) speech-language pathology student. Before beginning the survey, participants were prompted to read through detailed study background, aims, procedures, anonymous data to be collected, data protection policies, and contact persons and were required to provide informed consent before proceeding. Other than demographic information including profession, years of experience, generation, and sex, no personal information was collected, and no incentives for participation were offered.

Platform

The web-based survey was hosted on university-licensed LimeSurvey (version 4.3.14+200826; LimeSurvey GmbH), a web-based statistical survey web application that conforms to the required data security legislation dictated by the German Federal Data Protection Act, the European Data Protection Directive 95/46/EC, and the European General Data Protection Regulation [44]. To prevent repeated access to the survey, unique survey visitors were tracked by cookies as allowed per the participant's browser settings, but no IP addresses were saved. Cookies were set at the start of the survey and were valid for the LimeSurvey default of 365 days.

Survey Design and Content

An interdisciplinary team (the authors) consisting of an SLP (YL), phoniatrician and otolaryngologist (CNR), and instructional designer (ML) developed, pretested, and cross-checked a semistructured anonymous questionnaire to ensure comprehensibility and appropriateness of the survey questions. This second part of the survey contained 15 questions pertaining to sociodemographic information and attitudes and preferences regarding a proposed digital resource library app, which we have termed a DLT. There were 12 screens with 1 to 4 questions displayed per page, including the initial page with participant information on which the participant had to give consent before proceeding. The survey contained the following question types: yes or no questions, multiple-answer questions (with a free-text response option), arrays with Likert scale ratings, and free-text entries. Array questions contained 5 to 10 features or topics, which the participants rated on a 4-point Likert scale (*translated from German:* not important at all, not important, important, and very important and not useful, minimally useful, useful, and very useful). An even-numbered scale was used to avoid central tendency bias. Free-text entries were conditionally displayed based on the preceding yes or no question; they allowed for expansion upon the chosen answer and additional comments. For each question, directions were provided to aid understanding (eg, "multiple answers may be chosen" and "please rate the following statements"). To ensure common understanding of the purpose of the proposed app, a narrative explanation of the app's purpose and function was provided before any questions were presented. All questions except the free-text entries were mandatory for survey completion and submission. Participants were able to revise their answers using the forward and backward navigation buttons. Surveys were collected from August 2020 to December 2020.

Statistical Analysis

Data from the anonymous surveys were analyzed using SPSS (version 27; IBM Corp) to analyze data in a primarily descriptive manner.

Results

Overview

Of the 213 unique survey visitors, 13 (6.1%) individuals visited the start page containing study information and informed consent but did not start the survey, and 35 (16.4%) individuals started the survey but did not complete it. The participation rate was 93.9% (200/213), and the completion rate was 77.5% (165/213). Only completed questionnaires (optional responses not required) were analyzed. Excluding 0.6% (1/165) of the surveys from a dentistry student, 99.4% (164/165) of the surveys were analyzed. Participant characteristics are summarized in Table 1. Generations were defined according to the divisions defined by the Pew Research Center [45].

Table 1. Participant characteristics (N=164).

Characteristics and subtypes	Participants, n (%)
Sex	
Women	145 (88.4)
Men	19 (11.6)
Profession	
Physician (phoniatrician and ears, nose, and throat specialist)	32 (19.5)
Speech-language pathologist	69 (42.1)
Medical student (German: *Humanmedizin Studierende*)	20 (12.2)
Speech-language pathology student	43 (26.2)
Generation	
Generation Z (1996 and later)	56 (34.1)
Generation Y or Millennials (1980-1995)	62 (37.8)
Generation X (1965-1979)	33 (20.1)
Baby Boomer (1946-1964)	13 (7.9)
Professional experience (years)	
0[a]	60 (36.6)
1-5	38 (23.2)
6-10	15 (9.1)
11-15	11 (6.7)
16-20	18 (10.9)
>20	22 (13.4)

[a]Still studying.

General Interest in a *DLT* App

Of the 164 included participants, 162 (98.8%) participants expressed that they were open to trialing the proposed DLT app for subjects about CSD. In an optional follow-up response, 6.3% (2/32) physicians who indicated no interest in the proposed app cited concerns regarding app data collection. No significant differences across professions, years of experience, or generations were found. Regarding usefulness, 96.9% (159/164) of the participants reported that they found the proposed app helpful. In an optional follow-up free-response question, 3% (1/32) of the physicians indicated concerns about peer review of resources, and another 3% (1/32) of the physicians expressed that the topics were irrelevant to their current work. No significant differences across professions, generations, or years of experience were observed.

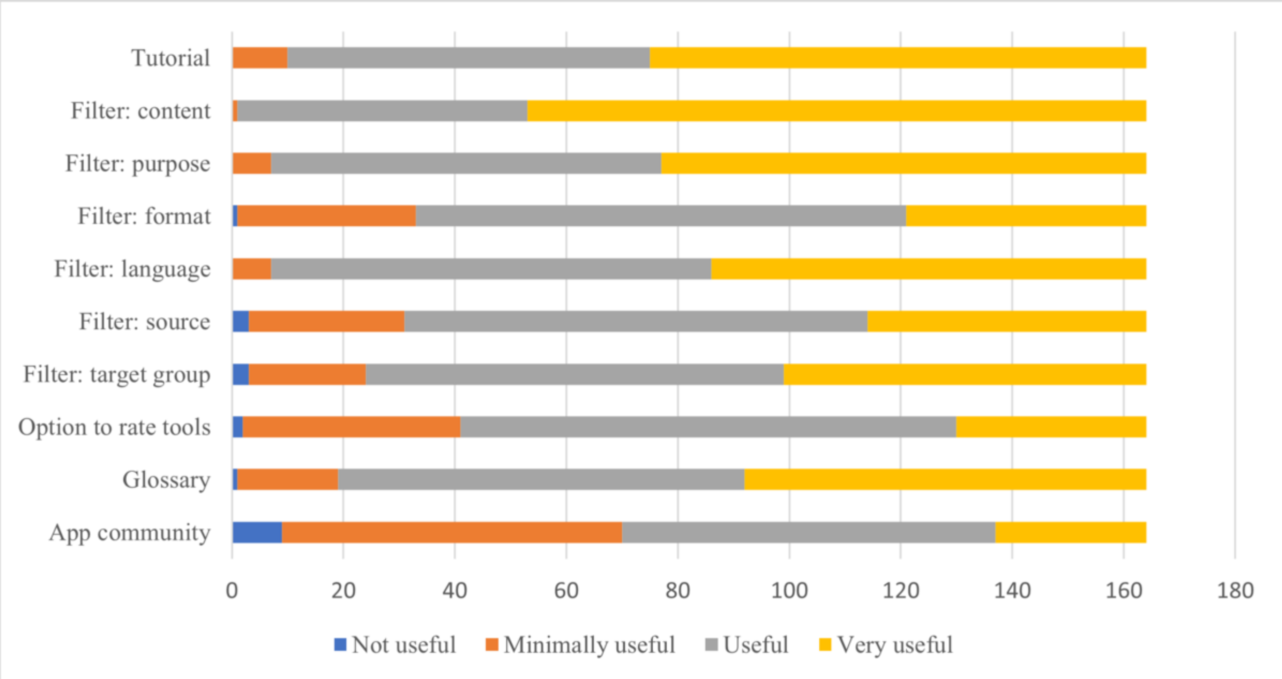
App Functions

Participants were asked to rate selected app functions on a 4-point Likert scale (1=not useful, 2=minimally useful, 3=useful, and 4=very useful; translated from German). This scale was also used to rate app settings in the next section. Participants were asked to rate the following app functions: introductory tutorial; filter functions based on content, purpose, digital format, language, source, and target audience; tool rating function; glossary; and app community function. A summary of the perceived usefulness of these selected app functions is presented in Figure 1. Regarding the introductory tutorial, of the 164 participants, 89 (54.3%) participants rated the function as very useful, 65 (39.6%) rated it as useful, and 10 (6.1%) rated it as minimally useful. Of the 164 participants, 111 (67.7%) participants rated a filter function based on content or topic (eg, anatomy and specific disorder category) as very useful, 52 (31.7%) ranked it as useful, and 1 (0.6%) participant ranked it as minimally useful. Similarly, most participants found a filter function based on the purpose or focus of a digital tool (eg, general disorder overview and clinical measurement) as very useful (87/164, 53%) and useful (70/164, 42.7%), whereas 4.3% (7/164) of the participants rated the function to be minimally useful. Among the 164 participants, a filter function based on digital format was rated as very useful by 43 (26.2%) participants, useful by 88 (53.7%), minimally useful by 32 (19.5%), and not useful by 7 (4.3%) participants. Of the 164 participants, a filter function based on language was rated as very useful by 78 (47.6%) participants, useful by 79 (48.2%), and minimally useful by 7 (4.3%) participants. Of the 164 participants, 50 (30.5%) participants found a filter based on source (eg, digital tool created from a university vs commercial or industry) to be very useful, 83 (50.6%) found it to be useful, 28 (17.1%) found it to be minimally useful, and 3 (1.8%) participants found it to be not useful. Among the 164 participants, a filter function based on target audience (eg, students vs professionals) was rated as very useful by 65 (39.6%) participants to be very useful, useful by 75 (45.7%), minimally useful by 21 (12.8%), and not useful by 3 (1.8%) participants.

Interestingly, although most participants found the option to rate tools as very useful (34/164, 20.7%) or useful (89/164, 54.3%), approximately one-fourth of the participants rated this as minimally (39/164, 23.8%) or not useful (2/164, 1.2%). Regarding the glossary function with digital learning and digital health terminology, of the 164 participants, 72 (43.9%) participants rated the function as very useful, 73 (44.5%) rated it as useful, 18 (10.9%) rated it as minimally useful, and 1 (0.6%) participant rated it as not useful. Finally, when rating the usefulness of an app community function, opinions varied greatly. Among the 164 participants, 27 (16.5%) participants rated the function as very useful, 67 (40.9%) rated it as useful, 61 (37.2%) rated it as minimally useful, and 9 (5.5%) rated it as not useful. Significant differences across professions were found regarding preference for an app community (H_3=9.785; P=.02), specifically between physicians and medical students and between medical students and speech-language pathology students; however, pairwise comparisons were no longer significant when Bonferroni correction was applied.

When asked to provide additional desired functions in an optional free-response follow-up question, participants cited the challenges associated with specific functions and several interesting suggestions. An SLP cited a preference for text tutorials as they felt that video tutorials did not allow sufficient time for processing keywords or skipping irrelevant material. Additional functions suggested by SLPs included the incorporation of newsfeed feature, filter function based on complexity, text-to-speech function, examples of use, individualization options in settings, frequently asked questions (FAQs) function, and the ability to save certain content. Some of these features were presented in the question regarding app setting functions (eg, newsfeed feature and personalization options). Speech-language pathology students additionally suggested the option to track learning progress and link similar content to help with standardization or validation of certain digital tools. A physician also suggested the incorporation of audio or visual aids (eg, text-to-speech and larger font options) for students or professionals who need such supports.

Figure 1. Perceived usefulness of selected app functions.

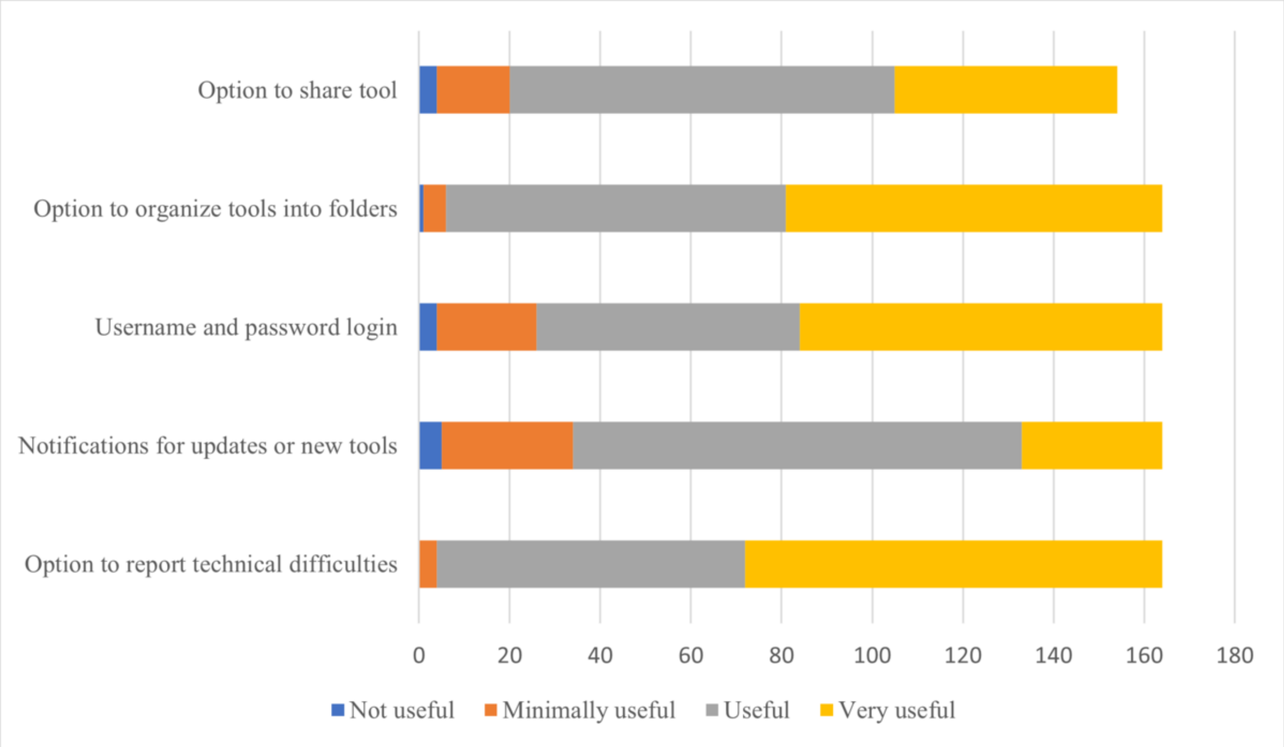


App Settings

Participants rated the app settings in terms of their perceived usefulness. Participants were specifically asked to rate the following app settings: option to share tools, option to organize tools into folders, username and password login, notifications for updates, and setting for reporting technical difficulties. A summary of the perceived usefulness of these app settings is shown in Figure 2. Regarding the option to share digital tools via social media links (eg, through email, WhatsApp, and Facebook), of the 164 participants, 49 (29.9%) participants rated such a function as very useful, 95 (57.9%) rated it as useful, 16 (9.8%) rated it as minimally useful, and 4 (2.4%) participants rated it as not useful. Of the 164 participants, 83 (50.6%) participants rated the settings option to organize and save tools into personalized categories and folders as very useful, 75 (45.7%) rated it as useful, 5 (3%) rated it as minimally useful, and 1 (0.6%) participant rated it as not useful. Regarding the incorporation of a username and password function, of the 164 participants, 80 (48.8%) participants found the setting to be very useful, 58 (35.4%) found it to be useful, 22 (13.4%) found it to be minimally useful, and 4 (2.4%) found it to be not useful. Significant differences were found across professional groups (H_3=8.006; P=.046) regarding the username and password function; specifically, SLPs demonstrated greater preference for such a setting than their speech-language pathology student counterparts (P=.045) when Bonferroni correction was applied. Similarly, significant differences were found across generations (H_3=9.309; P=.03), with Generation Z demonstrating significantly high distribution of opinions (P=.04), whereas Generation X primarily preferred a username and password function. Regarding a setting for receiving notifications for updates or the addition of new digital tools to the app library, of the 164 participants, 31 (18.9%) participants rated it as very useful, most participants (n=99, 60.4%) found it to be useful, 29 (17.7%) found it to be minimally useful, and 5 (3%) participants found it to be not useful. Of the 164 participants, most participants also agreed that a setting for reporting technical issues would be very useful (n=92, 56.1%) or useful (n=68, 41.5%), whereas 4 (2.4%) participants found it to be minimally useful. Significant differences were found across generations (H_3=9.309; P=.02); however, these findings were no longer significant for pairwise comparisons when Bonferroni correction was applied.

When asked to provide additional desired functions in an optional free-response follow-up question, only SLPs and speech-language pathology students made additional comments. SLPs again emphasized the desire for a text-to-speech function and an FAQs section, whereas speech-language pathology students additionally suggested the option for data extraction to or synchronization with Microsoft Office or commonly used programs and the option to synchronize personalized digital libraries across multiple devices.

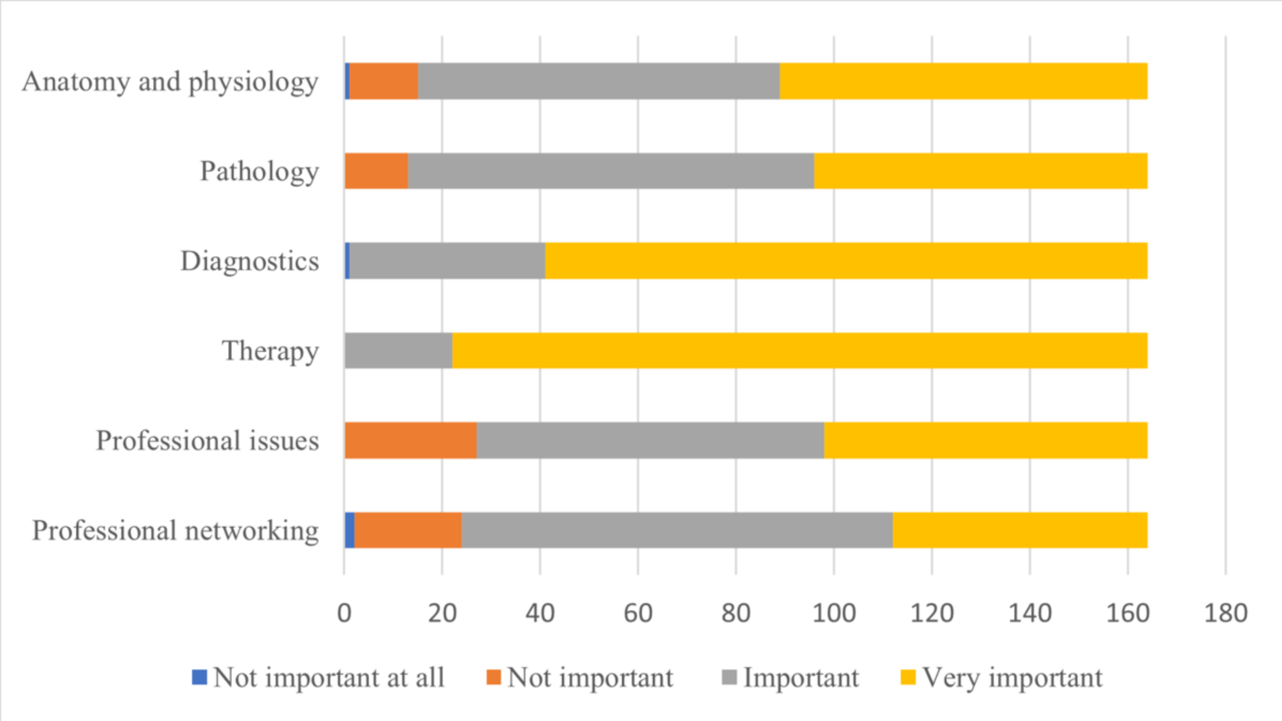
Figure 2. Perceived usefulness of selected setting functions.



Content Areas

Participants were asked to rate the perceived importance of various clinical and professional subjects on a 4-point Likert scale (1=not important at all, 2=not important, 3=important, and 4=very important). Participants were asked to rate the importance of the following content areas: anatomy and physiology, pathology, diagnostics, therapy, professional practice issues, and professional networking. A summary of the perceived importance of the selected clinical content areas is presented in Figure 3. Of the 164 participants, 75 (45.7%) participants found anatomy and physiology to be very important, 74 (45.1%) found them to be important, 14 (8.5%) found them to be not important, and 1 (0.6%) found them to be not important at all. Of the 164 participants, most participants rated the content area of pathology as very important (n=68, 41.5%) or important (n=83, 50.6%), whereas 13 (7.9%) participants found it to be not important. All except 1 participant found the subject of diagnostics to be either very important (123/164, 75%) or important (40/164, 24.4%), with significant difference across professional groups (H_3=9.098; P=.03), specifically with greater variance in perceived importance of diagnostic topics among physicians than among SLPs (P=.02). All participants agreed that therapy was either a very important (142/164, 86.6%) or important content area (40/164, 24.4%); however, interestingly, significant differences were found across professional groups regarding perceived level of importance (H_3=21.236; P<.001). In particular, physicians demonstrated significantly great variance regarding the perceived level of importance for topics related to therapy, whereas SLPs almost unanimously rated therapy as a very important content area (P=.02). Of the 164 participants, 66 (40.2%) participants rated professional issues as very important, 71 (43.3%) rated them as important, and 27 (16.5%) participants rated them as not important. Regarding the content area of professional networks, of the 164 participants, 52 (31.7%) participants perceived it to be very important, 88 (53.7%) perceived it to be important, 22 (13.4%) perceived it to be not important, and 2 (1.2%) participants perceived it to be not important at all.

Figure 3. Perceived importance of selected clinical content areas.



App Concerns

To gain insight into participants' opinions regarding challenges or concerns regarding the proposed app, participants were asked to choose from suggested concerns and were also given the option of a free-response textbox to express their opinions. Although of the 164 participants, 53 (32.3%) participants indicated that they had no concerns at all, most (n=111, 67.7%) participants expressed concerns. Specifically, of the 164 participants, 69 (42.1%) participants expressed concerns about technical difficulties, 67 (40.9%) expressed concerns about data privacy and protection, and 13 (7.9%) participants expressed that they doubted the usefulness of such a proposed app. No significant differences across professions, generations, or years of experience were observed.

In the optional free-response textbox, SLPs expressed concerns regarding quality, the possibility of data overflow, and the need for keeping the app consistently updated. Speech-language pathology students indicated concerns regarding potential difficulties in knowing how to use all the app functions, the quality of the tools included in the library, and how digital tool ratings would be managed. In addition, physicians mentioned concerns about the long-term applicability of some of the tools included in the app library and the potential for limited exchange among users of the app. Notably, 2.4% (4/164) of the participants mentioned confusion and limited understanding of the proposed app and wanted to see a prototype to aid their evaluation.

Discussion

Overview

In this study, participants indicated readiness to trial a DLT app and reported desired functions that would increase utility and ability to share resources and help with knowledge management; they also reported concerns regarding technical barriers, data protection, quality, and sustainability.

Given the exponentially increasing number of digital resources that generates tremendous cognitive load, solutions for better management and evaluation of the vast amounts of incoming digital information are urgently needed. For professionals, lack of time and skills for effective searching has been previously shown to hamper effective integration of information into workflow and clinical practice. Moreover, speed of information access has been shown to take precedence over the quality of information, indicating a problematic lack of prioritization of evidence-based practice in the face of digital information overload [46-48]. For students, it has been demonstrated that providing access to information at the point of need promotes learner-centric knowledge and skill building, which is further facilitated through time management and access to resources other than those offered by their academic institutions [2,3,49,50]. Thus, to improve accessibility, quality, and manageability of digital resources and tools, this survey study sought to investigate the potential usefulness and desired features of a proposed collaborative library of digital tools and resources—a DLT app for CSD.

General Interest in DLT App

The fact that most participants (162/164, 98.8%) indicated interest in trialing the proposed app is encouraging and demonstrates a readiness to engage in digital health topics, as seen in other studies with clinical professionals and students [6,11,23,34]. For the few participants who did not indicate interest in the app, their concerns of data collection are not unfounded. Given that many health apps are free and paid for, often using personal data for personalized advertisement and marketing, such fears are understandable [51]; however, it was not the intention of the proposed app to be used for any commercial purposes. Importantly, a participant indicated feeling that digital resources were not relevant to their current work. Such comments demonstrate that although most professionals and students may be open to digital developments, the degree of digitalization and acceptance varies. Moreover, there is a critical need for explicit digital skills and digital health education—both at the level of academic training and continuing professional development.

App Functions

Regarding app functions, features (introductory tutorial, app rating function, glossary, and app community) and specific filter functions (based on content or topic, purpose, digital format, language, source, and target audience) were investigated. Given that introductory tutorials have been demonstrated to greatly increase the usability of an app, it is unsurprising that most survey participants rated this feature as very useful (89/164, 54.3%) or useful (65/164, 39.6%) [52]. However, notably, in an optional free-response follow-up, a participant mentioned the format of introductory tutorials, specifically pointing to the challenge of insufficient time for processing and understanding, given the increasing shift to video-based as opposed to only text-based tutorials. This aligns with previous literature demonstrating that although video-based tutorials often provided rich visual input and support, text-based introductory materials required less mental effort to comprehend and thus had an efficiency advantage [53]. Thus, although video tutorials organized by functions or settings could be useful for visual input, extraneous redundant processing could be reduced by eliminating unnecessary onscreen text, simple visuals, and the option to pause or skip sections [54].

Regarding tool or resource rating, it was notable that approximately one-fourth of the participants perceived the function to be minimally useful (39/164, 23.8%) or not useful (2/164, 1.2%). Considering the dire need for more quality assessment or peer-review processes for the increasing number of digital resources that continue to be unevaluated, it is interesting that high perceived importance was not found. Although it would be necessary to systematically explore and design the details of such an evaluation or rating function (eg, star ratings, commentaries, and standardized or nonstandardized scale ratings) for such a proposed app with experts, it has been suggested that formalized or standardized checklists, such as the Mobile App Rating Scale or the Interactive Mobile App Review Toolkit, could be used to improve quality assessment [55,56]. Systematically, well-designed review criteria could help professionals and students to have more tangible means of quickly attaining information regarding a digital tool's evidence base, usefulness, problems, and costs among other factors [56]. However, in the long run, it will be important for professional regulating bodies to take a large part in regulating and promoting such review criteria to foster greater standardization of evaluation criteria [40,56].

Regarding glossary function, which defines common digital learning and digital health terminology, most participants rated the function as very useful (72/164, 43.9%) or useful (73/164, 44.5%). As demonstrated by the previously reported lack of confidence in digital health concepts, such a function could serve as an essential foundational reference and also be flexibly modified and adjusted as digitalization continues to advance [28,33-37]. Glossaries are crucial for building a basic common, agreed-upon understanding of specific concepts—which is critical in the ever-evolving digital environment, where new definitions and concepts are frequently emerging and must be adaptively adjusted [57].

Consistent with previous studies, overall, medical students showed more positive views toward an app community [28,58], and in this study more so than their professional counterparts, which, to the best of our knowledge, has not been previously reported. Such web-based learning environments could provide a space for problem solving and developing an interprofessional, collaborative community of team-based practice early in students' academic and clinical careers [59]. For clinical educators, it has been suggested that digital communities in the form of *web-based communities of practice* could also serve as a valuable resource to exchange insights, build professional relationships, foster innovation, and generate successful scholarship of teaching and learning [59]. Moreover, such communities could additionally help to increase user engagement, thereby potentially helping to keep shared information and digital tools up-to-date and serving as a discussion platform for future app improvements. Nevertheless, there are reservations about the benefit of such a function, given that more than one-third of the study participants negatively rated the usefulness of an app community. Although it could be that participants simply prefer to engage in professional exchange elsewhere (eg, social media and existing platforms such as ResearchGate) [60,61], there is evidence that professionals and students are strategic and selective in their use of information-communication technology based on their perception of the extent to which certain tools will meet their operational needs, regardless of their level of digital skills [62,63]. Therefore, investigating the reasons or motivations for perceived utility in future studies would be insightful. In a web-based environment where the establishment of collegiality may be more difficult owing to limited direct social interaction, research has also demonstrated that supportive organizational culture, respect for cultural dimensions of exchange and intellectual insights, presence of personal knowledge–based trust, and availability of adequate exchange tools best foster openness to collaborative knowledge sharing and generation [62,64]. Moving forward, rigorous research into methods to craft such *intellectually safe* and connected web-based spaces most effectively will be critical.

Filter functions were also suggested to help facilitate the targeted identification of relevant tools. As previously mentioned, it has been suggested that the issue of information overload in our digital age may mostly be related to the ability of individuals to concretely use information at their disposal, which can be enhanced through information filters [38,39]. Although information filters based on content or topic (163/164, 99.4%), purpose (157/164, 95.7%), language (157/164, 95.7%), and target audience (140/164, 85.4%) were relatively common and agreed upon to be useful or very useful by most participants, filters for digital format type were considered to be minimally useful by 19.5% (32/164) or not useful by 4.3% (7/164) of the participants. This could be explained by the fact that clinical professionals and students may be more concerned with the content and value of the information itself as opposed to the format in which it is presented. Although incorporation of digital format type as a filter could be useful for identifying information that can be compatibly incorporated into academic or research presentations or even patient's devices, it appears that basic knowledge of digital formats and digital literacy skills continues to be questionable and limited among clinical professionals and students, which again calls for the purposeful integration of health care–relevant digital skills training in academic and continuing education [33,65,66]. On the other hand, the source of a digital resource (eg, created by an academic institution or for commercial purposes), was considered to be very useful (50/164, 30.5%) or useful (83/164, 50.6%) by most participants. Given the currently low barrier for entering the app market, an increasing number of medical or clinical apps or digital resources are being developed by individuals or companies outside the health care sector. Although some have engaged expert opinion, others have not and often do not have the clinical insights to ensure the scientific or clinical quality of the resource, unlike academic institutions where quality evaluation of evidence base may be easier to ensure; however, they may not always be effective [67,68]. Here, it is also useful to mention that under current entrepreneurial models, the creation of *minimum viable products* including for medical apps or digital tools are introduced to the market with the intention of gradual improvement based on user feedback over time. This means that immature or nonoptimized resources, whether for patient or educational use, are being used when their quality or efficacy has not yet been established [69]. This highlights a great challenge of understanding how to shift strategies for evidence-based practices in the context of rapid digital progression—although a filter function based on source will not help to tackle this complex question, it may serve as a first step to encourage professionals and students to be more critical of the resources they choose to use.

App Settings

App setting options were explored to determine personalization functions that could help to enhance the usability of the proposed app. Consistent with findings that suggest that students and clinical professionals are open to and use social media for academic and clinical purposes, most participants (144/164, 87.8%) positively rated the option to share tools via social media [12,70,71]. Importantly, although it has been demonstrated that the use of social media for the sharing of clinical information can be helpful for quick and easy dissemination, it is notable that social networks serve as information filters that may rather highlight digital resources of personal popularity as opposed to true clinical evidence base and quality [38]. Moreover, social media as a digital means of clinical learning has demonstrated questionable educational value and unclear evidence regarding its effect on learning and performance outcomes [70,71].

Regarding personalization functions, an overwhelming majority of the survey participants positively rated the option to organize and save tools into folders. Such a function could serve as an additional layer of information filtering and support the encoding, organization, and synthesis of data as professionals and students try to understand the digital resources and tools that they find more relevant [54]. Therefore, it was encouraging to see the suggestion by several individuals for the option for linking similar content and for extraction to commonly used programs such as Microsoft Office. The suggestion of compatibility options across multiple devices further supported clinicians' and students' desires for easy accessibility, which could potentially increase the uptake of the proposed app. As previously demonstrated, technical software issues including those associated with nontransferability and noncompatibility across different mobile devices were barriers to learning efficacy; individuals were more likely to implement solutions that increased accessibility [5]. The additional suggestion of audiovisual aids (eg, text-to-speech and larger font options) for users with specific needs perhaps reflects the field's general focus on disability supports, given the clinical populations that are typically affected by communication disorders [72].

Regarding personalization of functions and features, the suggestion of a username and password function—which would likely be needed to save such settings—was met with significant variance across professional groups and generations. Specifically, SLPs indicated greater preference in comparison with their speech-language pathology student counterparts, as did Generation X in comparison with Generation Z. Regarding generational differences, although it is particularly difficult to delineate why professional group differences were found among SLPs and speech-language pathology students, it has been previously found that middle-aged individuals (aged 45-60 years, belonging to Generation X) had more negative views regarding data disclosure than their younger counterparts (aged 19-24 years, belonging to Generation Z) and may thus demonstrate greater preference toward password protections [73].

Most participants (130/164, 79.3%) also positively rated notifications for updates and new digital resources and the setting for reporting technical issues. Importantly, notifications, which were also requested by a survey participant in the form of *newsfeed* feature, have been associated with both benefits and drawbacks. Although notifications could help to keep users up-to-date and readily informed, they can be disruptive or further contribute to information overload [74]. Thus, it is important to design setting functions that allow the user to manage or turn off push notifications or to incorporate updates as a newsfeed feature only. Furthermore, the incorporation of a function for reporting technical difficulties has been identified as a critical criterion for app development, maintenance, and improvement [75].

Content Areas

It has been suggested that solutions for information overload can involve, among technological solutions and improved digital literacy, the creation or adaptation of specific content [38]. Subject-focused materials have been demonstrated to improve timely access to relevant material, which in turn can benefit the storage and retrieval of learned information [76]. Thus, perceptions about specific content filters were also explored. The foundational content areas of anatomy and physiology and pathology were rated by most participants to be useful, regardless of their profession, generation, or years of experience. However, physicians demonstrated significantly greater variance in the perceived importance of diagnostics and therapy than their speech-language pathology counterparts, who provided overwhelmingly positive ratings for both topics. Although it is more difficult to explain the significant differences observed in terms of perceived importance of diagnostic content (clinical responsibilities that fall within both physicians' and SLPs' responsibilities, although different in specific scope), differences in the perceived importance of therapy content can potentially be explained by the fact that SLPs take a large part in therapy in their scope of practice [77]. Furthermore, although professional issues and networking may more directly affect working clinical professionals, this study found no significant differences in the perceived importance of such topics among participants across professional groups, generations, or years of experience. This is promising as it suggests that students also value the professional skills and networks that will be crucial to their future work. However, importantly, evidence suggests that such perceived value may not necessarily translate to effective practice without appropriate guidance or explicit teaching [78].

App Concerns

It is well reported that there are clinical professional and student concerns about the quality of digital resources for both patient and academic purposes [40,79,80]. Thus, the concerns of the survey participants were investigated to better understand how to craft more effective future solutions. Although approximately one-third of the participants (53/164, 32.3%) reported no concerns with the suggested app at all, most participants (111/164, 67.7%) highlighted concerns about technical difficulties, data privacy and protection, and utility of the suggested app and several self-reported concerns in the optional free-response follow-up question. Technical difficulties can potentially be addressed through the proposed introductory tutorial, technical difficulty reporting function, or even the participant-suggested FAQs section, which can be a basic troubleshooting page. Data privacy concerns are well reported among health-related apps and digital tools, especially considering the vast amounts of sensitive health-related and personal information that can be collected through such means [81]. Although it was not the intention of the proposed app—which serves as a reference tool as opposed to an app for medical purposes or diagnostics—to collect any personal data other than potential username and password information for app personalization, this concern highlights the importance of ensuring that all apps and digital resource strictly align with current legal frameworks (eg, General Data Protection Regulation) to protect sensitive personal or medical data [82,83]. In addition, it is useful to mention that the concern of reduced exchange and interaction could be addressed through the incorporation of the suggested app community function; however, this would inevitably increase the amount of personal data that would need to be saved or anonymous aliases could be used. Nevertheless, it is worth re-emphasizing that the primary purpose of the proposed app was first and foremost to serve as a collaborative digital library of digital resources for CSD.

Another concern that necessitates further discussion is the need for consistent updates and app maintenance. Lack of app maintenance is reportedly a major reason for the failure of many apps [84]. Thus, to help facilitate the viability of the proposed digital library and reference app, it could be useful to include the option to suggest tools as a part of a collaborative effort to keep the included digital resources and tools up-to-date. It has also been previously suggested that a *curated app repository* that includes apps meeting minimum standards could be managed through *risk-based app triage*, which could be partially automated based on criteria such as the previously mentioned quality checklists (eg, Mobile App Rating Scale and Interactive Mobile App Review Toolkit). Thresholds can be set for determining when apps pose low risk and can undergo a more automated evaluation process, whereas other aspects could be more thoroughly and manually evaluated [40]. Such a system can potentially be applied to the proposed digital tool library that includes resources beyond just mobile apps. Furthermore, as mentioned by a survey participant, it is also important to consider the long-term relevance of digital resources or when a resource should be rendered obsolete. As mentioned previously, in the current age of continuously incoming data, the management of such information requires constant strategizing. Resources can quickly become irrelevant and, for reasons of limited data storage capacity, would have to be removed. To address this challenge, digital resources and tools can similarly undergo the previously mentioned triage process, with a shifted focus toward the timeliness and relevance of the digital resource.

Limitations

Although this study has demonstrated promising interest in the proposed digital reference app for CSD professionals and students, it must be considered with its limitations in mind. First, this study investigated perceptions toward a proposed app based on a narrative description. Although only 2.4% (4/164) of the participants indicated difficulty in understanding the intentions of the app as no prototype was offered as an objective reference, their understanding of the proposed app could have varied and informed their reported perceptions. The decision to implement a questionnaire before the prototype was made in light of previous literature citing lack of knowledge of users' demands and expectations as a key reason for prototype failure [84,85]. Thus, we deliberately chose to implement a co-design approach in which stakeholder insights were incorporated from the forefront to determine whether such a digital app would be desirable and inquired into need areas and preferences that could be used to ideally support more sustainable future development [86,87]. Regarding external validity, it is important to emphasize that the study was conducted in German-speaking countries only, and thus, perceptions and attitudes may likely differ from other cultural or geographical contexts. However, given the global reach of digitalization and the rather comprehensive nature of the proposed digital library reference app, the study findings could help to highlight common or global trends and useful resources across the field of CSD internationally. It is also useful to mention that, as an open survey, a convenience sample was collected; thus, it is possible that individuals who already held greater interest in digital topics were more likely to participate in the survey. Notably, many of the survey participants had ≤5 years of professional experience, which could have certainly affected their perceptions and attitudes toward such an app. In future studies, it would be useful to investigate whether more experienced professionals would demonstrate different preferences or insights into the utility of such a digital resource. Similarly, given this convenience sample, it was difficult to account for differences in group sizes; however, statistical adjustments to data analyses were made as appropriate. The great disparity between male and female participants further highlights the larger trend of a rather female-dominated field of speech-language pathology and the increasing number of women entering the medical specialty of otolaryngology [88,89]. In light of the exponential pace of digital progress, this survey study reveals the current needs, preferences, and perceptions regarding a proposed digital reference tool, which will likely evolve to include other needed supports as new digital resources, formats, and challenges arise.

Future Directions

As this study has only started to explore the potential utility of a collaborative and interdisciplinary digital library reference app, moving forward, it will be critical to further investigate, design, and test desired settings and functions to determine whether stakeholders' perceptions align with the actual use and implementation of such a tool. Even after prototype creation, several rounds of stakeholder evaluation and testing would be preferable to ensure that such an app is not released prematurely without proper initial evidence. Nevertheless, ongoing re-evaluation and improvement would be necessary given the ever-changing digital health care environment [68,90]. As digital learning and health apps continue to emerge, it will be critical that resources are tailored to specific target audiences, these stakeholders are engaged in digital resource creation and evaluation processes, and ongoing technical assistance is explicitly integrated into support tools [5]. As current knowledge and awareness of the range of available digital resources in the interdisciplinary field of CSD is limited, approaching the challenge with a digital *library* or *repository* can help to increase awareness, access, management, sharing, and, ideally, future quality assessment of available and emerging digital resources [90]. Such a learning reference tool could additionally serve as an important prerequisite for investigation into digital tools for clinical use. In the future, investigating the utility of such an app for those working in other related interdisciplinary fields, such as occupational therapists or nurses working in CSD, could also be useful.

Our study has demonstrated that clinical professionals and students in CSD are open to trialing a repository-like, collaborative, interdisciplinary digital library reference app and prefer features and functions that optimize usability, allow personalization, and increase exchanges regarding the quality assessment and evidence base for digital resources and tools. These stakeholders prefer a wide range of content topics and have reasonable concerns about the technical or data privacy challenges associated with app use; however, they are ready to explore new solutions for more efficient and effective knowledge and information management. The digital age is presenting opportunities and challenges for clinical teaching, learning, and practice that "...result in a richer range of resources to support practice and learning, but also creates conflicting evidence, insecurity about the knowledge and greater demands on the professional to identify the appropriate knowledge for their problem in question" [91]. As the digital health care landscape continues to advance at an unpredictable pace, information overload will be inevitable and will require traditional means of collecting, managing, and evaluating clinical information to adaptively evolve through ongoing cycles of evaluation to focus on and improve clinical decision-making, research, and clinical practice.

Acknowledgments

The authors would like to thank the Alexander von Humboldt Foundation for supporting the international and interdisciplinary exchange that made this collaboration possible. The first author, YL, is a former Chancellor Fellow of the Alexander von Humboldt Foundation who worked closely with both coauthors (CNR and ML) and was hosted by the third author CNR.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey screens (in German).
[PDF File (Adobe PDF File), 1004 KB - mededu_v8i2e34042_app1.pdf]

References

1. Ellaway R, Masters K. AMEE Guide 32: e-Learning in medical education Part 1: learning, teaching and assessment. Med Teach 2008 Jun;30(5):455-473. [doi: 10.1080/01421590802108331] [Medline: 18576185]
2. Davies BS, Rafique J, Vincent TR, Fairclough J, Packer MH, Vincent R, et al. Mobile Medical Education (MoMEd) - how mobile information resources contribute to learning for undergraduate clinical students - a mixed methods study. BMC Med Educ 2012 Jan 12;12:1 [FREE Full text] [doi: 10.1186/1472-6920-12-1] [Medline: 22240206]

3. Wallace S, Clark M, White J. 'It's on my iPhone': attitudes to the use of mobile computing devices in medical education, a mixed-methods study. BMJ Open 2012 Aug 24;2(4):e001099 [FREE Full text] [doi: 10.1136/bmjopen-2012-001099] [Medline: 22923627]
4. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012 Sep 27;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]
5. Green BL, Kennedy I, Hassanzadeh H, Sharma S, Frith G, Darling JC. A semi-quantitative and thematic analysis of medical student attitudes towards M-Learning. J Eval Clin Pract 2015 Oct;21(5):925-930. [doi: 10.1111/jep.12400] [Medline: 26153482]
6. Nguyen L, Barton SM, Nguyen LT. iPads in higher education-hype and hope. Br J Educ Technol 2014 Jan 27;46(1):190-203 [FREE Full text] [doi: 10.1111/bjet.12137]
7. Goh P, Sandars J. A vision of the use of technology in medical education after the COVID-19 pandemic. MedEdPublish 2020 Mar 26;9(1):49. [doi: 10.15694/mep.2020.000049.1]
8. Iwai Y. Online learning during the COVID-19 pandemic. Scientific American Blog Network. 2020 Mar 13. URL: https://blogs.scientificamerican.com/observations/online-learning-during-the-covid-19-pandemic/ [accessed 2021-05-20]
9. Ahmed H, Allaf M, Elghazaly H. COVID-19 and medical education. Lancet Infect Dis 2020 Jul;20(7):777-778 [FREE Full text] [doi: 10.1016/S1473-3099(20)30226-7] [Medline: 32213335]
10. mHealth: new horizons for health through mobile technologies. World Health Organization. 2011. URL: http://apps.who.int/iris/bitstream/handle/10665/44607/9789241564250_eng.pdf?sequence=1&isAllowed=y [accessed 2021-05-31]
11. Dunleavy G, Nikolaou CK, Nifakos S, Atun R, Law GC, Tudor Car L. Mobile digital education for health professions: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Feb 12;21(2):e12937 [FREE Full text] [doi: 10.2196/12937] [Medline: 30747711]
12. Bahner DP, Adkins E, Patel N, Donley C, Nagel R, Kman NE. How we use social media to supplement a novel curriculum in medical education. Med Teach 2012;34(6):439-444. [doi: 10.3109/0142159X.2012.668245] [Medline: 22449268]
13. Chase TJ, Julius A, Chandan JS, Powell E, Hall CS, Phillips BL, et al. Mobile learning in medicine: an evaluation of attitudes and behaviours of medical students. BMC Med Educ 2018 Jun 27;18(1):152 [FREE Full text] [doi: 10.1186/s12909-018-1264-5] [Medline: 29945579]
14. Nerminathan A, Harrison A, Phelps M, Alexander S, Scott KM. Doctors' use of mobile devices in the clinical setting: a mixed methods study. Intern Med J 2017 Mar;47(3):291-298. [doi: 10.1111/imj.13349] [Medline: 27925381]
15. Chang CY, Hwang GJ. Trends of mobile technology-enhanced medical education: a review of journal publications from 1998 to 2016. Int J Mob Learn Organ 2018;12(4):373-393. [doi: 10.1504/ijmlo.2018.095153]
16. Thompson K, Zimmerman E. Pediatric speech-language pathologists' use of mobile health technology: qualitative questionnaire study. JMIR Rehabil Assist Technol 2019 Sep 26;6(2):e13966 [FREE Full text] [doi: 10.2196/13966] [Medline: 31573922]
17. Disability and health. World Health Organization. 2020 Dec 1. URL: https://www.who.int/news-room/fact-sheets/detail/disability-and-health [accessed 2021-04-22]
18. Ahmadvand A, Gatchel R, Brownstein J, Nissen L. The biopsychosocial-digital approach to health and disease: call for a paradigm expansion. J Med Internet Res 2018 May 18;20(5):e189 [FREE Full text] [doi: 10.2196/jmir.9732] [Medline: 29776900]
19. Nelson R, Staggers N. Health informatics: an interprofessional approach. St Louis, MO, USA: Elsevier; Dec 8, 2016.
20. Kurpinski K, Johnson T, Kumar S, Desai T, Li S. Mastering translational medicine: interdisciplinary education for a new generation. Sci Transl Med 2014 Jan 08;6(218):218fs2. [doi: 10.1126/scitranslmed.3006858] [Medline: 24401938]
21. Hollweg W, Beck EM, Schulenburg K, Trock S, Räbiger J, Kraus E, et al. Interprofessional health care - field of study with future and challenges / Interprofessionelle Versorgung – Ein Studiengebiet mit Zukunft und Herausforderungen. Int J Health Prof 2016 May 12;3(1):37-46. [doi: 10.1515/ijhp-2016-0009]
22. Theodoros D. A new era in speech-language pathology practice: innovation and diversification. Int J Speech Lang Pathol 2012 Jun;14(3):189-199. [doi: 10.3109/17549507.2011.639390] [Medline: 22563895]
23. Andrews T, Davidson B, Hill A, Sloane D, Woodhouse L. Using students' own mobile technologies to support clinical competency development in speech pathology. In: Kitchenham A, editor. Models for Interdisciplinary Mobile Learning: Delivering Information to Students. Hershey, PA, USA: IGI Global; 2011:247-264.
24. Edwards J, Dukhovny E. Technology training in speech-language pathology: a focus on tablets and apps. Perspect ASHA Spec Interest Groups 2017 Jan;2(10):33-48. [doi: 10.1044/persp2.sig10.33]
25. Leinweber J, Dockweiler C. Perspektiven der Digitalisierung in der Logopädie/Sprachtherapie. forum:logopädie. 2020 May. URL: https://www.dbl-ev.de/fileadmin/Inhalte/Bilder/Service/Forum_Logopaedie/fl_2020_03.pdf [accessed 2022-04-09]
26. Kuperstock JE, Horný M, Platt MP. Mobile app technology is associated with improved otolaryngology resident in-service performance. Laryngoscope 2019 Jan;129(1):E15-E20. [doi: 10.1002/lary.27299] [Medline: 30151970]
27. Hsueh WD, Bent JP, Moskowitz HS. An app to enhance resident education in otolaryngology. Laryngoscope 2018 Jun;128(6):1340-1345. [doi: 10.1002/lary.27040] [Medline: 29214641]
28. Tarpada SP, Hsueh WD, Gibber MJ. Resident and student education in otolaryngology: a 10-year update on e-learning. Laryngoscope 2017 Jul;127(7):E219-E224. [doi: 10.1002/lary.26320] [Medline: 27782300]

29. Rak K, Völker J, Taeger J, Bahmer A, Hagen R, Albrecht UV. Medical apps in oto-rhino-laryngology. Laryngorhinootologie 2019 Mar;98(S 01):S253-S289 [FREE Full text] [doi: 10.1055/a-0740-4866] [Medline: 31096301]
30. Hodes RJ, Insel TR, Landis SC, NIH Blueprint for Neuroscience Research. The NIH toolbox: setting a standard for biomedical research. Neurology 2013 Mar 12;80(11 Suppl 3):S1 [FREE Full text] [doi: 10.1212/WNL.0b013e3182872e90] [Medline: 23479536]
31. Ader DN. Developing the patient-reported outcomes measurement information system (PROMIS). Med Care 2007 May;45(5):S1-S2 [FREE Full text] [doi: 10.1097/01.mlr.0000260537.45076.74]
32. Torous J, Vaidyam A. Multiple uses of app instead of using multiple apps - a case for rethinking the digital health technology toolbox. Epidemiol Psychiatr Sci 2020 Jan 31;29:e100 [FREE Full text] [doi: 10.1017/S2045796020000013] [Medline: 32000876]
33. Lam MK, Hines M, Lowe R, Nagarajan S, Keep M, Penman M, et al. Preparedness for eHealth: health sciences students' knowledge, skills, and confidence. J Inf Technol 2016 Jan 1;15(2016):305-334 [FREE Full text]
34. Machleid F, Kaczmarczyk R, Johann D, Balčiūnas J, Atienza-Carbonell B, von Maltzahn F, et al. Perceptions of digital health education among European medical students: mixed methods survey. J Med Internet Res 2020 Aug 14;22(8):e19827 [FREE Full text] [doi: 10.2196/19827] [Medline: 32667899]
35. Recommendations by the next generation. European Health Parliament. 2020. URL: https://www.healthparliament.eu/wp-content/uploads/2020/09/EHP_Recommendations_by_the_next_Generation.pdf [accessed 2021-05-31]
36. Gagnon MP, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220 [FREE Full text] [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
37. Alwani M, Bandali E, Larsen M, Shipchandler TZ, Ting J. Current state of surgical simulation training in otolaryngology: systematic review of simulation training models. Arch Otolaryngol Head Neck Surg 2019;3(1):5. [doi: 10.24983/scitemed.aohns.2019.00109]
38. Klerings I, Weinhandl AS, Thaler KJ. Information overload in healthcare: too much of a good thing? Z Evid Fortbild Qual Gesundhwes 2015;109(4-5):285-290. [doi: 10.1016/j.zefq.2015.06.005] [Medline: 26354128]
39. Shirky C. It's not information overload. It's filter failure. MAS Context. 2008. URL: https://www.mascontext.com/issues/7-information-fall-10/its-not-information-overload-its-filter-failure/ [accessed 2021-05-31]
40. Wyatt JC. How can clinicians, specialty societies and others evaluate and improve the quality of apps for patient use? BMC Med 2018 Dec 03;16(1):225 [FREE Full text] [doi: 10.1186/s12916-018-1211-7] [Medline: 30501638]
41. Boruff JT, Storie D. Mobile devices in medicine: a survey of how medical students, residents, and faculty use smartphones and other mobile devices to find information. J Med Libr Assoc 2014 Jan;102(1):22-30 [FREE Full text] [doi: 10.3163/1536-5050.102.1.006] [Medline: 24415916]
42. 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 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
43. Lin Y, Lemos M, Neuschaefer-Rube C. Digital health and digital learning experiences across speech-language pathology, phoniatrics, and otolaryngology: interdisciplinary survey study. JMIR Med Educ 2021 Nov 05;7(4):e30873 [FREE Full text] [doi: 10.2196/30873] [Medline: 34738911]
44. Schmitz C. What you need to know about data security in LimeSurvey. LimeSurvey. 2020 Sep 25. URL: https://www.limesurvey.org/blog/20-blog/114-what-you-need-to-know-about-data-security-in-limesurvey [accessed 2021-04-09]
45. Dimock M. Defining generations: where Millennials end and Generation Z begins. Pew Research Center. 2019 Jan 17. URL: https://www.pewresearch.org/fact-tank/2019/01/17/where-millennials-end-and-generation-z-begins/ [accessed 2022-04-09]
46. Clarke MA, Belden JL, Koopman RJ, Steege LM, Moore JL, Canfield SM, et al. Information needs and information-seeking behaviour analysis of primary care physicians and nurses: a literature review. Health Info Libr J 2013 Sep;30(3):178-190 [FREE Full text] [doi: 10.1111/hir.12036] [Medline: 23981019]
47. Fox S. After Dr Google: peer-to-peer health care. Pediatrics 2013 Jun;131 Suppl 4:S224-S225. [doi: 10.1542/peds.2012-3786K] [Medline: 23729765]
48. Duran-Nelson A, Gladding S, Beattie J, Nixon LJ. Should we Google it? Resource use by internal medicine residents for point-of-care clinical decision making. Acad Med 2013 Jun;88(6):788-794. [doi: 10.1097/ACM.0b013e31828ffdb7] [Medline: 23619072]
49. Spector JM, Anderson TM. Integrated and holistic perspectives on learning, instruction and technology: understanding complexity. Dordrecht, The Netherlands: Kluwer Academic Publishers; 2000.
50. Fuller R, Joynes V. Should mobile learning be compulsory for preparing students for learning in the workplace? Br J Educ Technol 2014 Jan 27;46(1):153-158. [doi: 10.1111/bjet.12134]
51. Huckvale K, Prieto JT, Tilney M, Benghozi PJ, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015 Sep 07;13:214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]
52. Xie B, Shabir I, Abelson H. Measuring the usability and capability of app inventor to create mobile applications. In: Proceedings of the 3rd International Workshop on Programming for Mobile and Touch. 2015 Oct Presented at: PROMOTO '15; October 27, 2015; Pittsburgh, PA, USA p. 1-8. [doi: 10.1145/2824823.2824824]

53. Hefter MH, Berthold K. Preparing learners to self-explain video examples: text or video introduction? Comput Human Behav 2020 Sep;110:106404. [doi: 10.1016/j.chb.2020.106404]
54. Clark RC, Mayer RE. e-Learning and the science of instruction: proven guidelines for consumers and designers of multimedia learning. 4th edition. Hoboken, NJ, USA: John Wiley & Sons; 2016.
55. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
56. Maheu MM, Nicolucci V, Pulier ML, Wall KM, Frye TJ, Hudlicka E. The interactive mobile app review toolkit (IMART): a clinical practice-oriented system. J Technol Behav Sci 2016 Dec 15;1(1-4):3-15. [doi: 10.1007/s41347-016-0005-z]
57. Holland PG. The importance of glossaries. Flow Meas Instrum 2002 Dec;13(5-6):299-301. [doi: 10.1016/s0955-5986(02)00058-4]
58. Mehta NB, Hull AL, Young JB, Stoller JK. Just imagine: new paradigms for medical education. Acad Med 2013 Oct;88(10):1418-1423. [doi: 10.1097/ACM.0b013e3182a36a07] [Medline: 23969368]
59. Yarris LM, Chan TM, Gottlieb M, Juve AM. Finding your people in the digital age: virtual communities of practice to promote education scholarship. J Grad Med Educ 2019 Feb;11(1):1-5 [FREE Full text] [doi: 10.4300/JGME-D-18-01093.1] [Medline: 30805087]
60. Curran V, Matthews L, Fleet L, Simmons K, Gustafson DL, Wetsch L. A review of digital, social, and mobile technologies in health professional education. J Contin Educ Health Prof 2017;37(3):195-206. [doi: 10.1097/CEH.0000000000000168] [Medline: 28834849]
61. Manca S. ResearchGate and Academia.edu as networked socio-technical systems for scholarly communication: a literature review. Res Learn Technol 2018 Feb 20;26:1-16 [FREE Full text] [doi: 10.25304/rlt.v26.2008]
62. Gannon-Leary P, Fontainha E. Communities of practice and virtual learning communities: benefits, barriers and success factors. eLearning Papers 2007 Sep:5 [FREE Full text]
63. Smith PJ, Barty K, Stacey EA. Limitations of an established community of practice in developing online innovation. CiteSeerx. 2005. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.84.4912 [accessed 2022-04-09]
64. Ardichvili A. Learning and knowledge sharing in virtual communities of practice: motivators, barriers, and enablers. Adv Dev Hum Resour 2008 Jun 05;10(4):541-554. [doi: 10.1177/1523422308319536]
65. Dattakumar A, Gray K, Henderson KB, Maeder A, Chenery H. We are not educating the future clinical health professional workforce adequately for e-health competence: findings of an Australian study. Stud Health Technol Inform 2012;178:33-38. [Medline: 22797016]
66. Konttila J, Siira H, Kyngäs H, Lahtinen M, Elo S, Kääriäinen M, et al. Healthcare professionals' competence in digitalisation: a systematic review. J Clin Nurs 2019 Mar;28(5-6):745-761. [doi: 10.1111/jocn.14710] [Medline: 30376199]
67. Grundy QH, Wang Z, Bero LA. Challenges in assessing mobile health app quality: a systematic review of prevalent and innovative methods. Am J Prev Med 2016 Dec;51(6):1051-1059. [doi: 10.1016/j.amepre.2016.07.009] [Medline: 27659122]
68. Price L, Kirkwood A. Using technology for teaching and learning in higher education: a critical review of the role of evidence in informing practice. High Educ Res Dev 2013 Nov 09;33(3):549-564. [doi: 10.1080/07294360.2013.841643]
69. What is a minimum viable product (MVP)? Entrepreneur Handbook. 2015 Aug 15. URL: https://entrepreneurhandbook.co.uk/minimum-viable-product/ [accessed 2021-06-05]
70. Sterling M, Leung P, Wright D, Bishop TF. The use of social media in graduate medical education: a systematic review. Acad Med 2017 Jul;92(7):1043-1056 [FREE Full text] [doi: 10.1097/ACM.0000000000001617] [Medline: 28225466]
71. Sutherland S, Jalali A. Social media as an open-learning resource in medical education: current perspectives. Adv Med Educ Pract 2017;8:369-375 [FREE Full text] [doi: 10.2147/AMEP.S112594] [Medline: 28652840]
72. Gillam RB, Marquardt TP. Communication sciences and disorders: from science to clinical practice. 4th edition. Burlington, MA, USA: Jones & Bartlett Learning; 2019.
73. Miltgen CL, Peyrat-Guillard D. Cultural and generational influences on privacy concerns: a qualitative study in seven European countries. Eur J Inf Syst 2014;23(2):103-125. [doi: 10.1057/ejis.2013.17]
74. Westermann T, Möller S, Wechsung I. Assessing the relationship between technical affinity, stress and notifications on smartphones. In: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct. 2015 Aug Presented at: MobileHCI '15; August 24-27, 2015; Copenhagen, Denmark p. 652-659 URL: https://doi.org/10.1145/2786567.2793684 [doi: 10.1145/2786567.2793684]
75. Llorens-Vernet P, Miró J. Standards for mobile health-related apps: systematic review and development of a guide. JMIR Mhealth Uhealth 2020 Mar 03;8(3):e13057 [FREE Full text] [doi: 10.2196/13057] [Medline: 32130169]
76. Swanwick T. Understanding medical education. In: Swanwick T, editor. Understanding medical education: evidence, theory and practice. 2nd ed. Hoboken, NJ, USA: John Wiley & Sons; 2013:1-6.
77. Scope of practice in speech-language pathology. American Speech-Language-Hearing Association (ASHA). 2016 May 9. URL: https://tinyurl.com/m93sbdsa [accessed 2021-06-05]
78. Cruess RL, Cruess SR, Steinert Y. Teaching medical professionalism: supporting the development of a professional identity. Cambridge, UK: Cambridge University Press; 2016.

79. Boulos MN, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014 Feb 5;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
80. McMillan B, Hickey E, Mitchell C, Patel M. The need for quality assurance of health apps. BMJ 2015 Nov 04;351:h5915. [doi: 10.1136/bmj.h5915] [Medline: 26536895]
81. Benjumea J, Ropero J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A. Privacy assessment in mobile health apps: scoping review. JMIR Mhealth Uhealth 2020 Jul 02;8(7):e18868 [FREE Full text] [doi: 10.2196/18868] [Medline: 32459640]
82. Thorarensen B. The processing of health information- protecting the individual right to privacy through effective legal remedies. Health Technol 2017 Mar 7;7(4):401-413. [doi: 10.1007/s12553-017-0184-4]
83. Sokolovska A, Kocarev L. Integrating technical and legal concepts of privacy. IEEE Access 2018 May 14;6:26543-26557. [doi: 10.1109/access.2018.2836184]
84. N Inukollu V, Keshamon DD, Kang T, Inukollu M. Factors influencing quality of mobile apps: role of mobile app development life cycle. Int J Softw Eng Appl 2014 Sep 30;5(5):15-34 [FREE Full text] [doi: 10.5121/ijsea.2014.5502]
85. Deininger M, Daly SR, Lee JC, Seifert CM, Sienko KH. Prototyping for context: exploring stakeholder feedback based on prototype type, stakeholder group and question type. Res Eng Des 2019 Oct;30(4):453-471 [FREE Full text] [doi: 10.1007/s00163-019-00317-5] [Medline: 32863595]
86. Sandholzer M, Rurik I, Deutsch T, Frese T. Medical students' expectations towards an implementation of a family medicine textbook as a comprehensive app in Germany. J Med Syst 2014 Oct;38(10):125 [FREE Full text] [doi: 10.1007/s10916-014-0125-y] [Medline: 25171920]
87. Kildea J, Battista J, Cabral B, Hendren L, Herrera D, Hijal T, et al. Design and development of a person-centered patient portal using participatory stakeholder co-design. J Med Internet Res 2019 Feb 11;21(2):e11371 [FREE Full text] [doi: 10.2196/11371] [Medline: 30741643]
88. Logopädie/Sprachtherapie: Schluss mit der Benachteiligung der Frauenberufe!. Deutscher Bundesverband für Logopädie (DBL). 2019 Mar 8. URL: https://tinyurl.com/43n3w4wr [accessed 2021-04-20]
89. Geschlecht. Kassenärztliche Bundesvereinigung (KBV). 2020. URL: https://gesundheitsdaten.kbv.de/cms/html/16396.php [accessed 2021-04-20]
90. Sousa MJ, Rocha Á. Strategic knowledge management in the digital age: JBR special issue editorial. J Bus Res 2019 Jan;94:223-226. [doi: 10.1016/j.jbusres.2018.10.016]
91. Bullock A. Does technology help doctors to access, use and share knowledge? Med Educ 2014 Jan;48(1):28-33. [doi: 10.1111/medu.12378] [Medline: 24330114]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
CSD: communication sciences and disorders
DLT: digital learning toolbox
FAQ: frequently asked question
PROMIS: Patient-Reported Outcomes Measurement Information System
SLP: speech-language pathologist

Edited by T Leung; submitted 04.10.21; peer-reviewed by MA Joaquin, Y Du, MDG Pimentel, J Ropero; comments to author 11.11.21; revised version received 28.01.22; accepted 18.03.22; published 27.04.22.

Please cite as:
Lin Y, Lemos M, Neuschaefer-Rube C
Digital Health and Learning in Speech-Language Pathology, Phoniatrics, and Otolaryngology: Survey Study for Designing a Digital Learning Toolbox App
JMIR Med Educ 2022;8(2):e34042
URL: https://mededu.jmir.org/2022/2/e34042
doi:10.2196/34042
PMID:35475980

Original Paper

Video-Based Communication Assessment of Physician Error Disclosure Skills by Crowdsourced Laypeople and Patient Advocates Who Experienced Medical Harm: Reliability Assessment With Generalizability Theory

Andrew A White[1], MD; Ann M King[2], MA; Angelo E D'Addario[2], MA; Karen Berg Brigham[3], JD, MPH; Suzanne Dintzis[4], MD, PhD; Emily E Fay[5], MD; Thomas H Gallagher[1,6], MD; Kathleen M Mazor[7], EdD

[1]Department of Medicine, University of Washington School of Medicine, Seattle, WA, United States
[2]National Board of Medical Examiners, Philadelphia, PA, United States
[3]Collaborative for Accountability and Improvement, University of Washington, Seattle, WA, United States
[4]Department of Pathology, University of Washington School of Medicine, Seattle, WA, United States
[5]Department of Obstetrics and Gynecology, University of Washington School of Medicine, Seattle, WA, United States
[6]Department of Bioethics and Humanities, University of Washington, Seattle, WA, United States
[7]Meyers Primary Care Institute, University of Massachusetts Medical School, Worcester, MA, United States

Corresponding Author:
Andrew A White, MD
Department of Medicine
University of Washington School of Medicine
Box 356429
1959 Pacific St
Seattle, WA, 98195
United States
Phone: 1 206 616 1447
Fax: 1 206 221 8732
Email: andwhite@uw.edu

Abstract

Background: Residents may benefit from simulated practice with personalized feedback to prepare for high-stakes disclosure conversations with patients after harmful errors and to meet American Council on Graduate Medical Education mandates. Ideally, feedback would come from patients who have experienced communication after medical harm, but medical researchers and leaders have found it difficult to reach this community, which has made this approach impractical at scale. The Video-Based Communication Assessment app is designed to engage crowdsourced laypeople to rate physician communication skills but has not been evaluated for use with medical harm scenarios.

Objective: We aimed to compare the reliability of 2 assessment groups (crowdsourced laypeople and patient advocates) in rating physician error disclosure communication skills using the Video-Based Communication Assessment app.

Methods: Internal medicine residents used the Video-Based Communication Assessment app; the case, which consisted of 3 sequential vignettes, depicted a delayed diagnosis of breast cancer. Panels of patient advocates who have experienced harmful medical error, either personally or through a family member, and crowdsourced laypeople used a 5-point scale to rate the residents' error disclosure communication skills (6 items) based on audiorecorded responses. Ratings were aggregated across items and vignettes to create a numerical communication score for each physician. We used analysis of variance, to compare stringency, and Pearson correlation between patient advocates and laypeople, to identify whether rank order would be preserved between groups. We used generalizability theory to examine the difference in assessment reliability between patient advocates and laypeople.

Results: Internal medicine residents (n=20) used the Video-Based Communication Assessment app. All patient advocates (n=8) and 42 of 59 crowdsourced laypeople who had been recruited provided complete, high-quality ratings. Patient advocates rated communication more stringently than crowdsourced laypeople (patient advocates: mean 3.19, SD 0.55; laypeople: mean 3.55, SD 0.40; $P<.001$), but patient advocates' and crowdsourced laypeople's ratings of physicians were highly correlated ($r=0.82$, *P*<.001). Reliability for 8 raters and 6 vignettes was acceptable (patient advocates: G coefficient 0.82; crowdsourced laypeople: G coefficient 0.65). Decision studies estimated that 12 crowdsourced layperson raters and 9 vignettes would yield an acceptable G coefficient of 0.75.

Conclusions: Crowdsourced laypeople may represent a sustainable source of reliable assessments of physician error disclosure skills. For a simulated case involving delayed diagnosis of breast cancer, laypeople correctly identified high and low performers. However, at least 12 raters and 9 vignettes are required to ensure adequate reliability and future studies are warranted. Crowdsourced laypeople rate less stringently than raters who have experienced harm. Future research should examine the value of the Video-Based Communication Assessment app for formative assessment, summative assessment, and just-in-time coaching of error disclosure communication skills.

(JMIR Med Educ 2022;8(2):e30988) doi:10.2196/30988

KEYWORDS

medical error disclosure; simulation studies; communication assessment; graduate medical education; crowdsourcing; patient-centered care; generalizability theory; medical education; medical error; communication

Introduction

Poor communication after a medical injury often leaves patients and families feeling alone, afraid, confused, and more likely to seek redress through malpractice claims [1,2]. One cause of this communication gap cited by both practicing and resident physicians is inadequate training on disclosing harmful medical errors [3,4]. Recently, communication and resolution programs have emerged as a framework to enable clinicians and health care institutions to communicate openly with patients and families, apologize, and offer compensation if an error contributed to patient harm [5]. Communication and resolution programs require clinicians, institutional leaders, and liability insurers to collaborate to provide transparent communication and emotional support for harmed patients. Communication and resolution programs align with recent American Council on Graduate Medical Education mandates that require all trainees to participate in real or simulated disclosure of harm events [6]. However, organizations adopting communication and resolution programs may struggle to prepare physicians for these difficult conversations, in part because of challenges in assessing and improving the specific communication skills required [7].

Traditional methods of assessing physician communication are not suitable for this particular type of task. For example, patient surveys can evaluate actual performance on routine communication, but individual physicians disclose harmful errors infrequently, and these high-stakes discussions are difficult to observe or record. As an alternative to real-world practice, educators often use standardized patients (individuals trained to act as a real patient) and simulated encounters for formative and summative assessments [8]. However, standardized patient exams are logistically intensive, expensive to implement at scale, and lack statistical reliability [9-12]. In addition, it is unknown whether standardized patients or peer physician raters adequately approximate the viewpoint of patients who have experienced medical injury. In particular, physicians' viewpoints about ideal disclosure content and performance differ from those of patients, which limits physicians' abilities to assess and coach other physicians' performance [13,14]. Although feedback would ideally come from harmed patients, researchers have found it difficult to reach this community because providers are reluctant to release details about harmed patients, and because patients hesitate to revisit painful events [15]. To make progress, educators and communication and resolution program leaders need a cost-effective and standardized assessment tool that provides actionable, on-demand, high-volume, and patient-centered feedback about physician communication skills after harm.

The National Board of Medical Examiners recently developed the Video-Based Communication Assessment app as an efficient approach to producing timely, specific, and individual feedback about verbal communication [16]. The Video-Based Communication Assessment app displays brief videos of case vignettes and asks users to audiorecord what they would say next to the patient [17]. Recorded responses are rated by web-based panels of analog patients. Analog patients are untrained raters given the task of listening to and rating their impressions of a medical interaction while assuming the patient perspective [18]. Analog patients are typically laypeople recruited via MTurk [19]; MTurk provides access to a very large, diverse population for survey research, and there is extensive proof that MTurk is an inexpensive, rapid, and high-quality data source [20,21]. Users then receive feedback reports with their individual ratings, comparative data on the user's cohort, learning points derived from analysis of crowdsourced raters' comments, and selected highly rated responses from peers. The only study [22] of the Video-Based Communication Assessment app published to date used a variety of 16 typical primary care communication scenarios and found that crowdsourced laypeople can provide high-quality, actionable feedback regarding physician communication skills. Key steps in evaluating the Video-Based Communication Assessment app for error disclosure skill assessment are understanding reliability, educational outcomes, and adoption challenges.

Our aim was to evaluate the reliability of crowdsourced laypeople as raters by comparing their ratings with those of patient advocates who had experienced harm in the course of in their own or a loved one's medical care. We hypothesized that crowdsourced layperson raters could provide reliable ratings of this specific communication skill, given sufficient panel size.

Methods

Overview

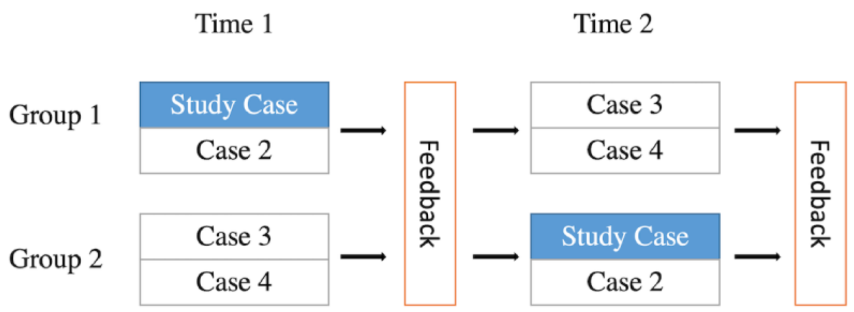
This descriptive study is part of a larger project to develop instruments for assessing resident error disclosure skills. With input from experienced attending physicians, we designed and pilot-tested 4 cases specific to the practice of internal medicine. Each case consisted of 3 or 4 vignettes depicting sequential stages in a conversation (for example, initially sharing information about a mistake, responding to a patient's emotional reaction). We recruited resident physicians at an academic center to use the Video-Based Communication Assessment app. Physicians' disclosure skills were rated by crowdsourced laypeople recruited on MTurk (Mechanical Turk; Amazon) and by a panel of patient advocates.

Participants

We recruited resident physicians in postgraduate years 1 through 3 from the University of Washington academic medical center. We invited all 183 internal medicine residents by email and provided dedicated participation time at a program-wide web-based educational conference (approximate attendance: 40 residents). Residents received a 10-minute orientation to the Video-Based Communication Assessment app and were given class time to participate. Participation was optional. Participants were randomly assigned to 1 of 2 pairs of initial cases to counteract order effects, using a crossover design (Figure 1). After receiving a feedback report, residents were eligible to complete the second 2 cases on their own. Participating residents received a $50 gift card after completing all 4 cases during a 2-month period; however, only 1 case was used in this study.

We used the following inclusion criteria for laypeople: resident of the United States, 18 years or older, and able to speak and read English. Patient advocates were recruited through advertisements with the Patient and Family Advocate Committee of the Collaborative for Accountability and Improvement (a network of health care leaders, attorneys, insurers, and patient advocates who support the development and widespread application of communication and resolution programs). Patient advocates were recruited if they met the following criteria: resident of the United States, 18 years or older, able to speak and read English, not currently or previously employed in health care, and having a personal history of having experienced serious medical injury in their own care or that of a family member. Patient advocates received a US $200 gift card for participation. Crowdsourced raters received variable amounts based on a rate of $0.20 per rating. A crowdsourced rater performing the same total number of ratings as a patient advocate would have received $12.

Figure 1. Crossover study design for 21 internal medicine residents using the Video-Based Communication Assessment app at study start (time 1) and approximately 4 weeks later (time 2). The study case in the blue box (breast cancer misdiagnosis) was selected for further study.



Ethics

The University of Washington Institutional Review Board determined that this study was exempt from review for resident, layperson, and patient advocate participants based on its policies, procedures, and guidance [23].

Video-Based Communication Assessment App

The concept and software of the Video-Based Communication Assessment app have been previously described [16]. The app was used to present vignettes, record user responses, and deliver feedback reports (Figure 2). Instead of a single stand-alone vignette, in this study, cases consisted of a linked series of 3 or 4 vignettes to simulate an unfolding conversation. Because a live conversation might not progress in the same manner or sequence, each vignette after the first was accompanied by text declaring what the patient understood at that point.

We used a case that depicted harm resulting from a delayed diagnosis of breast cancer, which is discovered by a primary care doctor just before the patient returns for an office visit (Table 1). This case was chosen because it has 3 segments, rather than 4, which reduced the time and cognitive demands imposed on the small group of patient advocates.

Figure 2. Screenshot from the Video-Based Communication Assessment app displaying a case of delayed diagnosis of breast cancer and the user controls for playing the vignette video and making an audio response to the patient.

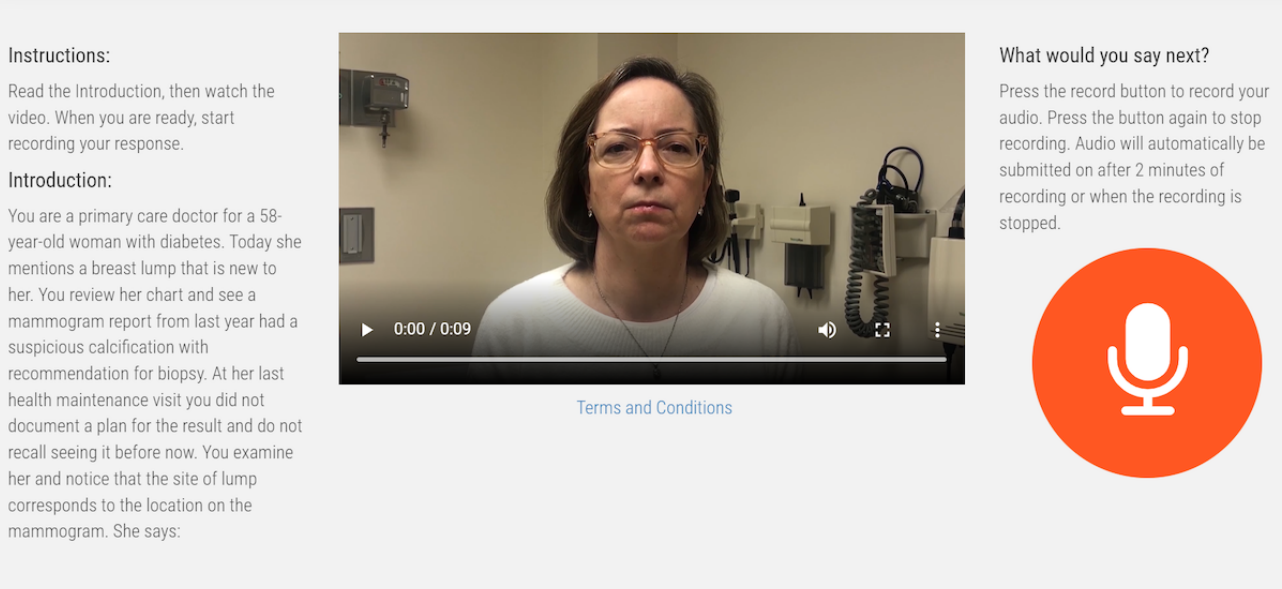


Table 1. Text and scenario (spoken by actors in 3 linked vignettes) presented to users (physicians) and raters (laypeople and patient advocates).

Vignette	Situation description (to physician)	Situation Description (to rater)	What the patient says
1	You are a primary care doctor for a 48-year-old woman with diabetes. Today she mentions a breast lump that is new to her. You review her chart and see a mammogram report from last year had a suspicious calcification with recommendation for biopsy. At her last health maintenance visit you did not document a plan for the result and do not recall seeing it before now. You examine her and the site of the lump corresponds to the location on x-ray. She says:	Lorna Smith visits her primary care doctor to evaluate a new breast lump. She figures it isn't anything serious because she had a mammogram last year and never heard about any abnormal results. The doctor examined her and she changed back to regular clothes. She wants to discuss the lump now and says:	"When I didn't hear from your office about the mammogram, I assumed everything was normal. Was there any sign of this lump on the test last year?"
2	You've told the patient that there were early warning signs of possible breast cancer on her mammogram one year ago. She says:	Lorna has learned that her mammogram last year showed early signs of possible breast cancer, but nothing was done about it. She is feeling panicked and says:	"This is terrible! I've never been more frightened…plus you're telling me that we might have known about it a long time ago!"
3	You've acknowledged how upsetting the error is. The patient now understands that there were early warning signs of possible breast cancer on her mammogram. She says:	Lorna feels like the clinic and her doctor have failed her. She asks:	"How could this happen to me? I feel like I can't trust anyone anymore. How am I supposed to believe your advice in the future?"

Data Collection

Resident physicians participated in the video-based communication assessment and provided audio responses to each vignette. All audio responses to a single case were bundled into rating tasks for the raters, comprising 4 physicians' responses to a case. Raters first completed an audio check and answered questions about demographic characteristics. Raters were asked to read the description of the vignette, view the patient video, listen to each vignette, and rate 6 items (Table 2). Due to the sequential design, we removed raters who did not complete all ratings. We also removed raters who used 2 or fewer response items on the 5-item survey because this may be a sign of inattention and poor rater quality [24]. We defined outliers as raters who reduced the interrater reliability of their task by 0.1 or more.

Table 2. Items to assess error disclosure communication skills.

Item	Response options
Overall this provider's response was	Poor, fair, good, very, good, or excellent
I would feel this provider was accountable for their actions	Not at all, a little, somewhat, very much, or completely
I would feel this provider was being honest about what happened	Not at all, a little, somewhat, very much, or completely
I would feel this provider was sincerely sorry for what happened	Not at all, a little, somewhat, very much, or completely
I would feel the provider understood how I was feeling	Not at all, a little, somewhat, very much, or completely
I would feel this provider cared about me	Not at all, a little, somewhat, very much, or completely
What would you want the provider to say if you were the patient in this situation?	Free text

Analysis

To create vignette-level scores, ratings were aggregated across all items for each vignette. To compare stringency between groups, we employed a 3×2 repeated measures factorial analysis of variance for vignette (1, 2, 3) and rater (patient advocate, crowdsourced layperson). To create overall assessment scores, we aggregated all vignette-level scores for each user (these continuous scores were derived from ordinal approximations of continuous variables, ie, the mean of Likert-scale responses [25,26]). To determine if an individual physician's score would be preserved between groups in relation to their peers, we calculated the Pearson correlation.

Generalizability theory utilizes analysis of variance to parse multiple sources of measurement error and estimate reliability under specific conditions [27]. A generalizability analysis was conducted using GENOVA (version 2.1; University of Iowa) to compute variance components for a fully crossed design utilizing a panel of patient advocates [28]. A separate generalizability analysis was conducted using urGENOVA (version 2.1; University of Iowa) to generate variance components for an unbalanced design utilizing crowdsourced layperson raters [29]. In order to determine the optimal design to achieve sufficient reliability, the estimated variance components were used to conduct multiple decision studies to produce G coefficients corresponding to varying numbers of vignettes and raters for each design.

Patient Advocate Design

To balance consistency and attention span, patient advocates rated batches of 7 physician responses at a time. Batches were block randomized and consisted of physicians' audio responses to all 3 vignettes. In G-theory, this is referred to as fully crossed design—physician crossed with vignette crossed with rater (p × v × λ).

Crowdsourced Layperson Design

Crowdsourced laypeople rated a subset of the physicians. Each crowdsourced layperson rated a single batch of 4 physician responses (all 3 vignettes). In G-theory, this is referred to as a rater nested within physician crossed with vignette ((λ:p) × v) design.

Results

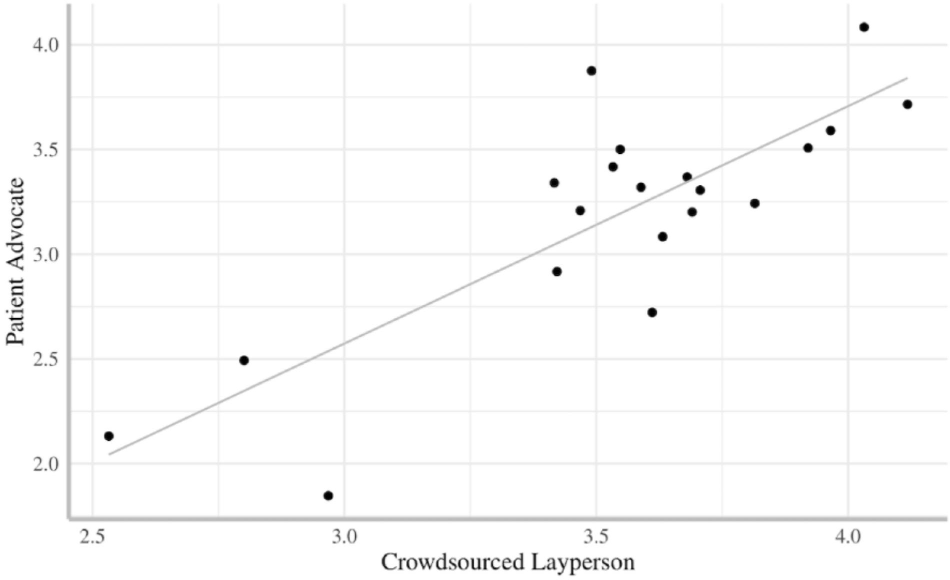
Participant Demographics

Although 21 internal medicine physicians completed all 3 vignettes, one physician was omitted from analyses because of incomplete ratings); therefore, 20 physicians (male: 6/20, 30%; female: 14/20, 70%), with total of 60 audiorecordings, were rated. The patient advocate panel (n=8; male: 2/8, 25%; female: 6/8,75%) had a median age of 57 years (IQR 53-74.3). Patient advocates reported that it took an average of 116 minutes (SD 62) to rate all 20 cases. A total of 59 crowdsourced laypeople were recruited, but 8 were removed because they did not rate all 3 vignettes in the case, 8 were removed for utilizing 2 or fewer response items, and 1 was deemed to be an outlier; thus, 42 crowdsourced layperson raters were included. Of the 42 crowdsourced raters, 16 (38%) were female; 20 individuals (48%) were between 18 and 34 years old, and 22 (52%) individuals were between 35 and 64 years old.

Comparing Crowdsourced Laypeople and Patient Advocates:

There was a significant overall main effect for rater ($F_{1,19}$=24.14, P<.001, d=0.75)—patient advocates (mean 3.19, SD 0.55) rated communication more stringently than crowdsourced laypeople (mean 3.55, SD 0.40) (Multimedia Appendix 1 and Multimedia Appendix 2). Patient advocate ratings were strongly correlated with crowdsourced layperson ratings (r=0.82, P<.001) (Figure 3).

Figure 3. Correlation between ratings of overall communication skill for resident physicians generated by panels of patient advocates and crowdsourced laypeople.



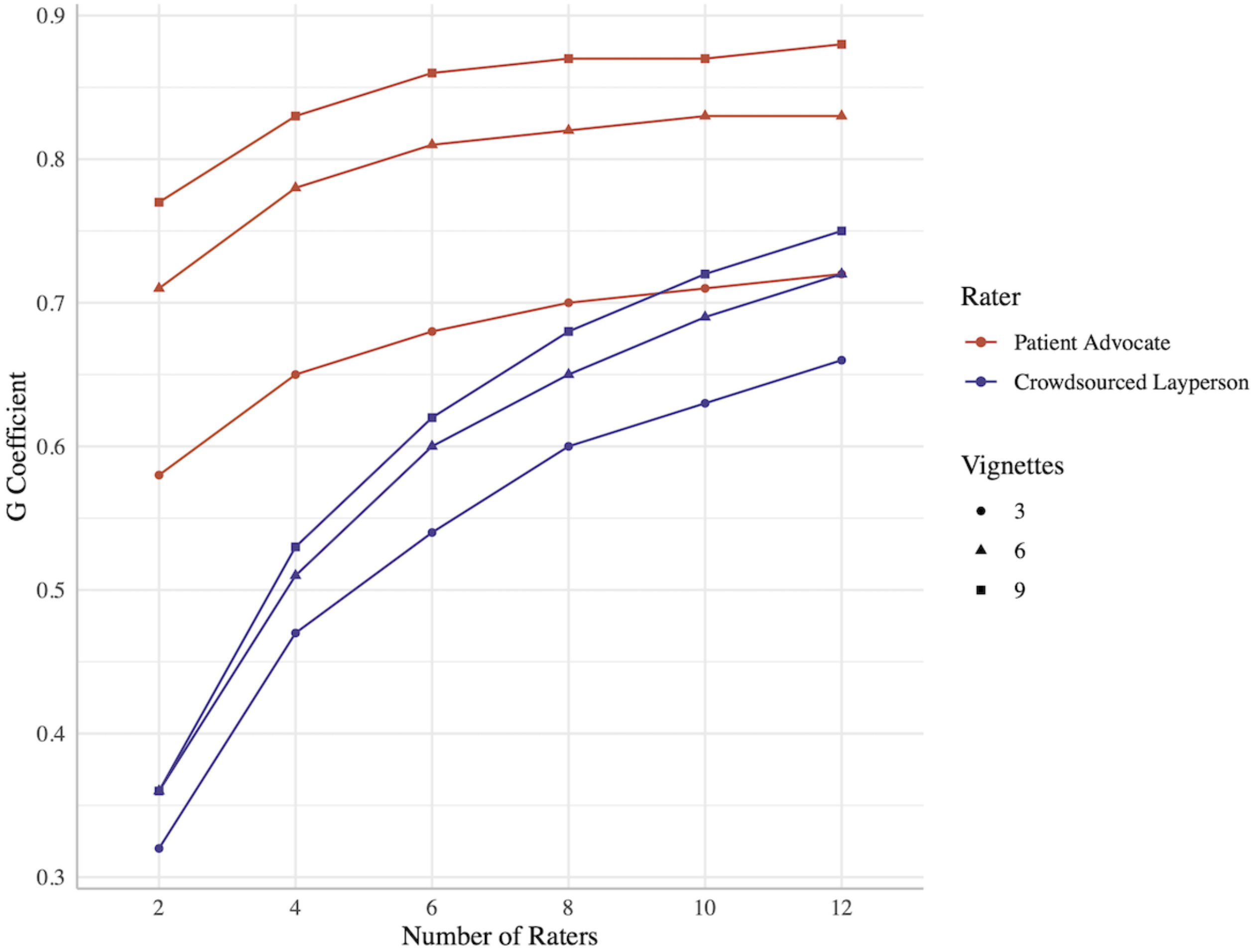
Generalizability

Generalizability analysis yielded the variance attributable to each component (Table 3). The G coefficients for 8 raters and 3 vignettes were 0.7 for patient advocates and 0.6 for crowdsourced laypeople. Maintaining 8 raters and increasing the task to 6 vignettes would increase the G coefficients (patient advocates: 0.82; crowdsourced laypeople: 0.65). Increasing the panels to 12 raters for 6 vignettes would increase the G coefficients (patient advocates: 0.83; crowdsourced laypeople: 0.72). Using 12 raters and 9 vignettes would yield G coefficients of 0.88 and 0.75 for patient advocates and crowdsourced laypeople, respectively (Figure 4).

Table 3. Generalizability study variance components.

Source of variance	Variance component	Variance percentage
Patient advocates (p × v × λ design)		
Physician	0.214	17.979
Rater	0.311	26.105
Vignette	0.017	1.421
Physician × rater	0.008	0.690
Physician × vignette	0.210	17.586
Rater × vignette	0.012	0.986
Residual	0.420	35.232
Crowdsourced laypeople ((λ:p) × v design)		
Physician	0.121	14.564
Vignette	0.007	0.906
Rater:physician	0.368	44.402
Physician × vignette	0.074	8.952
Residual	0.258	31.177

Figure 4. Reliability (G coefficient) models for panels of patient advocates and crowdsourced laypeople, by panel size and number of vignettes rated per user.



Discussion

Principal Findings

Patient advocates rated communication skills more stringently than crowdsourced laypeople, but the correlation between patient advocates' ratings and crowdsourced laypeople's ratings was high. Patient advocates also had higher reliability, but decision studies estimated that panels of crowdsourced laypeople could achieve a G coefficient of 0.75 with 12 raters and 9 vignettes.

These findings demonstrate that crowdsourced laypeople can reliably rate the error disclosure communication skills of physicians using the Video-Based Communication Assessment app. This is encouraging for communication and resolution program leaders and graduate medical educators who require an abundant and affordable pool of raters to support personalized feedback processes in the next generation of physician communication skill training programs. In principle, patient advocates would offer the best possible feedback, but large-scale training efforts would rapidly exhaust the willing and available patient advocate population, given the amount of time that these raters reported spending on this study. Instead, crowdsourced laypeople represent a large and sustainable pool of on-demand raters. Nonetheless, our finding that approximately one-third of crowdsourced laypeople (17/59, 29%) must be removed from analysis to optimize assessment reliability indicates that continuous rater performance monitoring, requirements for raters to complete all vignettes in a series, and a sufficient number of raters would be required for widespread deployment of the Video-Based Communication Assessment app in error disclosure training.

Educators who use the Video-Based Communication Assessment app should understand how crowdsourced raters differ from patient advocates, who represent the gold standard for informed assessment of physician error disclosure skills. Compared with crowdsourced individuals, patient advocates can achieve high reliability with smaller panel sizes and fewer vignettes per physician. This suggests that patient advocates have a common concept of the components of verbal communication that affect the quality of error disclosure and are highly attuned to differences among physicians. Of note, patient advocates assigned lower ratings to resident error disclosure communication than crowdsourced laypeople did. Educators and coaches should recognize that overall scores from crowdsourced laypeople are potentially more generous than those of patients who have experienced harm from medical errors and should note this in reviewing feedback with residents.

Comparison With Prior Work

The Video-Based Communication Assessment app had been previously only used with groups of stand-alone vignettes [22], but this is the first example of a case with sequential vignettes that simulate a longer conversation. The satisfactory reliability should encourage educators to develop cases for other extended exchanges, such as discussions about goals of care, shared decision-making, or new diagnoses of serious illness. However, our need to sacrifice a subset of ratings by crowdsourced laypeople who had not completed all of the vignettes within a case suggests that longer cases would benefit from a modified approach, such as the use of attention checks or restrictions (eg, a high past task acceptance ratio) [30,31]

Although physician educators have been used to evaluate trainee disclosure skills in a prior study [7], our findings suggest that using faculty as raters would be too costly for large training programs. Based on the time estimates in this study, a residency program with 60 residents, each completing 4 cases, would require an educator to allocate approximately 23 hours to listening and rating audio. Rather than finding 6 to 8 faculty to do this task for a single training session, crowdsourcing laypeople appears to be a more viable and rapid solution.

Future Directions

This study sets the stage for investigation of use of the Video-Based Communication Assessment app for error disclosure training, for example, for formative assessment (either for self-directed improvement or in conjunction with coaching from a teacher) or summative assessment and in the identification of struggling learners. Although we did not define a threshold for competency, low performers might warrant additional support from residency leaders, including attention on communication performance in other scenarios. Additional areas to explore include whether the tool can be used in undergraduate medical education, continuing medical education, or in just-in-time scenarios (for physicians to practice and receive feedback just before real-life error disclosure). Future studies should investigate the role of different error types (eg, diagnostic or therapeutic), harm severity, physician and patient identity (eg, gender, race), tone, and accent on ratings. The Video-Based Communication Assessment app could be used to understand the efficacy of training interventions and to study the natural history of communication skill development over time. Finally, future studies should also investigate whether error disclosure performance using the Video-Based Communication Assessment app is associated with other safety behaviors encouraged by communication and resolution programs, such as event reporting, root cause analysis, or physician participation in system redesign to prevent future errors.

Strengths and Limitations

Our work has limitations. We did not assess whether crowdsourced laypeople had personal experience with medical harm and did not measure the amount of time crowdsourced laypeople spent on this evaluation task. Additionally, we recruited patient advocates through their involvement in a national advocacy organization, and their rating behaviors may not generalize to the broader community of patients who have been harmed by care. The convenience sample of patient advocates was not age- and gender-matched to the sample of crowdsourced individuals, and age was not collected as a continuous variable for crowdsourced individuals. The Video-Based Communication Assessment app does not measure nonverbal communication skills, which play an essential role in communication about medical error [32,33]. Finally, this study was conducted using a single case with a breast cancer misdiagnosis and tested with medical residents and may, therefore, not be generalizable to other uses—other unique patient scenarios may require separate validation of crowdsourced laypeople as analog patients. Future research should aim to replicate findings with a more robust sample size.

Conclusion

Crowdsourced laypeople reliably rated error disclosure skills using the Video-Based Communication Assessment app, although reliably distinguishing high and low performers would require larger panels (9-12 raters) and more vignettes per examinee (9 or more). Fortunately, this is readily achievable in error disclosure curricula. Future studies should focus on the educational outcomes achieved by presenting analog patient feedback to resident physicians about their error disclosure communication skills, and the role of the Video-Based Communication Assessment app in other learner groups or just-in-time scenarios.

Acknowledgments

We thank Anders Chen, MD, and Kevin Blau, MD, for their assistance with study enrollment. The study was funded by the University of Washington's Patient Safety Innovations Program. The funder had no role in the study design or manuscript review.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Deidentified data of ratings by patient advocates for residents (user) and vignettes (v1,v2,v3).

[XLSX File (Microsoft Excel File), 14 KB - mededu_v8i2e30988_app1.xlsx]

Multimedia Appendix 2

Deidentified data of ratings by crowdsourced laypeople for residents (user) and vignettes (v1,v2,v3).

[XLSX File (Microsoft Excel File), 14 KB - mededu_v8i2e30988_app2.xlsx]

References

1. Gallagher TH, Garbutt JM, Waterman AD, Flum DR, Larson EB, Waterman BM, et al. Choosing your words carefully: how physicians would disclose harmful medical errors to patients. Arch Intern Med 2006;166(15):1585-1593. [doi: 10.1001/archinte.166.15.1585] [Medline: 16908791]
2. Delbanco T, Bell SK. Guilty, afraid, and alone--struggling with medical error. N Engl J Med 2007 Oct 25;357(17):1682-1683. [doi: 10.1056/NEJMp078104] [Medline: 17960011]
3. White AA, Bell SK, Krauss MJ, Garbutt J, Dunagan WC, Fraser VJ, et al. How trainees would disclose medical errors: educational implications for training programmes. Med Educ 2011 Apr;45(4):372-380 [FREE Full text] [doi: 10.1111/j.1365-2923.2010.03875.x] [Medline: 21401685]
4. Gallagher TH, Waterman AD, Garbutt JM, Kapp JM, Chan DK, Dunagan WC, et al. US and Canadian physicians' attitudes and experiences regarding disclosing errors to patients. Arch Intern Med 2006;166(15):1605-1611. [doi: 10.1001/archinte.166.15.1605] [Medline: 16908793]
5. Mello MM, Kachalia A, Roche S, Niel MV, Buchsbaum L, Dodson S, et al. Outcomes in two Massachusetts hospital systems give reason for optimism about communication-and-resolution programs. Health Aff (Millwood) 2017 Oct 01;36(10):1795-1803. [doi: 10.1377/hlthaff.2017.0320] [Medline: 28971925]
6. ACGME common program requirements (residency): section VI.A.1.a. Accreditation Council for Graduate Medical Education. URL: http://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/CPRResidency2019.pdf [accessed 2022-02-11]
7. Stroud L, Wong BM, Hollenberg E, Levinson W. Teaching medical error disclosure to physicians-in-training: a scoping review. Acad Med 2013 Jun;88(6):884-892. [doi: 10.1097/ACM.0b013e31828f898f] [Medline: 23619064]
8. Scalese RJ. Simulation-based assessment. In: Holmboe ES, Durning SJ, Hawkins RE, editors. Evaluation of Clinical Competence. Amsterdam: Elsevier; 2018:215-255.
9. Stroud L, McIlroy J, Levinson W. Skills of internal medicine residents in disclosing medical errors: a study using standardized patients. Acad Med 2009 Dec;84(12):1803-1808. [doi: 10.1097/ACM.0b013e3181bf9fef] [Medline: 19940591]
10. Patrício MF, Julião M, Fareleira F, Carneiro AV. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? Med Teach 2013 Jun;35(6):503-514. [doi: 10.3109/0142159X.2013.774330] [Medline: 23521582]
11. Adamo G. Simulated and standardized patients in OSCEs: achievements and challenges 1992-2003. Med Teach 2003 May;25(3):262-270. [doi: 10.1080/0142159031000100300] [Medline: 12881047]
12. Jacobs AC, van Jaarsveldt DE. 'The character rests heavily within me': drama students as standardized patients in mental health nursing education. J Psychiatr Ment Health Nurs 2016 Apr;23(3-4):198-206. [doi: 10.1111/jpm.12302] [Medline: 27090349]
13. Martinez W, Browning D, Varrin P, Sarnoff Lee B, Bell SK. Increasing patient-clinician concordance about medical error disclosure through the patient TIPS model. J Patient Saf 2019 Dec;15(4):305-307. [doi: 10.1097/PTS.0000000000000284] [Medline: 28492422]
14. Gallagher TH. A 62-year-old woman with skin cancer who experienced wrong-site surgery: review of medical error. JAMA 2009 Aug 12;302(6):669-677. [doi: 10.1001/jama.2009.1011] [Medline: 19584321]
15. Iedema R, Allen S, Britton K, Piper D, Baker A, Grbich C, et al. Patients' and family members' views on how clinicians enact and how they should enact incident disclosure: the "100 patient stories" qualitative study. BMJ 2011 Jul 25;343:d4423 [FREE Full text] [doi: 10.1136/bmj.d4423] [Medline: 21788260]
16. Mazor KM, King AM, Hoppe RB, Kochersberger AO, Yan J, Reim JD. Video-based communication assessment: development of an innovative system for assessing clinician-patient communication. JMIR Med Educ 2019 Feb 14;5(1):e10400 [FREE Full text] [doi: 10.2196/10400] [Medline: 30710460]
17. National Board of Medical Examiners. Introducing the VCA. YouTube. URL: https://www.youtube.com/watch?v=ZqdZKVk7aak [accessed 2022-02-11]
18. Blanch-Hartigan D, Hall JA, Krupat E, Irish JT. Can naive viewers put themselves in the patients' shoes?: reliability and validity of the analogue patient methodology. Med Care 2013 Mar;51(3):e16-e21. [doi: 10.1097/MLR.0b013e31822945cc] [Medline: 22498688]
19. Sheehan K. Crowdsourcing research: Data collection with Amazon's Mechanical Turk. Communication Monographs 2018;85(1):140-156 [FREE Full text] [doi: 10.1080/03637751.2017.1342043]
20. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? Perspect Psychol Sci 2011 Jan;6(1):3-5. [doi: 10.1177/1745691610393980] [Medline: 26162106]
21. Mortensen K, Hughes TL. Comparing Amazon's Mechanical Turk platform to conventional data collection methods in the health and medical research literature. J Gen Intern Med 2018 Apr;33(4):533-538 [FREE Full text] [doi: 10.1007/s11606-017-4246-0] [Medline: 29302882]
22. Mazor K, King A, Hoppe R, D'Addario A, Musselman T, Tallia A, et al. Using crowdsourced analog patients to provide feedback on physician communication skills. Patient Educ Couns 2021 Sep;104(9):2297-2303 [FREE Full text] [doi: 10.1016/j.pec.2021.02.047] [Medline: 33715944]
23. Guidance exempt research. UW Research. URL: https://www.washington.edu/research/policies/guidance-exempt-research-2/ [accessed 2022-04-28]

24. Vriesema C, Gehlbach H. Assessing survey satisficing: the impact of unmotivated questionnaire responding on data quality. Educ Res 2021 Aug 24;50(9):618-627. [doi: 10.3102/0013189X211040054]
25. Norman G. Non-cognitive factors in health sciences education: from the clinic floor to the cutting room floor. Adv Health Sci Educ Theory Pract 2010 Mar;15(1):1-8. [doi: 10.1007/s10459-009-9204-0] [Medline: 19924553]
26. Sullivan GM, Artino AR. Analyzing and interpreting data from likert-type scales. J Grad Med Educ 2013 Dec;5(4):541-542 [FREE Full text] [doi: 10.4300/JGME-5-4-18] [Medline: 24454995]
27. Gallagher TH, Garbutt JM, Waterman AD, Flum DR, Larson EB, Waterman BM, et al. Choosing your words carefully: how physicians would disclose harmful medical errors to patients. Arch Intern Med 2006;166(15):1585-1593 [FREE Full text] [doi: 10.1001/archinte.166.15.1585] [Medline: 16908791]
28. Crick J, Brennan R. Manual for GENOVA: A Generalized Analysis of Variance System. Iowa City: ACT Inc; 1983.
29. Brennan R. Generalizability Theory. New York: Springer; 2001.
30. Chandler J, Shapiro D. Conducting clinical research using crowdsourced convenience samples. Annu Rev Clin Psychol 2016;12:53-81. [doi: 10.1146/annurev-clinpsy-021815-093623] [Medline: 26772208]
31. Peer E, Vosgerau J, Acquisti A. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. Behav Res Methods 2014 Dec;46(4):1023-1031. [doi: 10.3758/s13428-013-0434-y] [Medline: 24356996]
32. Riess H, Kraft-Todd G. E.M.P.A.T.H.Y.: a tool to enhance nonverbal communication between clinicians and their patients. Acad Med 2014 Aug;89(8):1108-1112 [FREE Full text] [doi: 10.1097/ACM.0000000000000287] [Medline: 24826853]
33. Drukteinis DA, O'Keefe K, Sanson T, Orban D. Preparing emergency physicians for malpractice litigation: a joint emergency medicine residency-law school mock trial competition. J Emerg Med 2014 Jan;46(1):95-103. [doi: 10.1016/j.jemermed.2013.08.017] [Medline: 24071029]

Edited by T Leung; submitted 27.06.21; peer-reviewed by L Suppan, H Wolfe; comments to author 04.02.22; revised version received 19.02.22; accepted 31.03.22; published 29.04.22.

Please cite as:
White AA, King AM, D'Addario AE, Brigham KB, Dintzis S, Fay EE, Gallagher TH, Mazor KM
Video-Based Communication Assessment of Physician Error Disclosure Skills by Crowdsourced Laypeople and Patient Advocates Who Experienced Medical Harm: Reliability Assessment With Generalizability Theory
JMIR Med Educ 2022;8(2):e30988
URL: https://mededu.jmir.org/2022/2/e30988
doi:10.2196/30988
PMID:35486423

Original Paper

An Alternative to Traditional Bedside Teaching During COVID-19: High-Fidelity Simulation-Based Study

Shereen Ajab[1], MBBS, MRCS; Emma Pearson[1], MBBS; Steven Dumont[1], BSc, MSc; Alicia Mitchell[1], MBBS; Jack Kastelik[1], BSc, MBChB, MD; Packianathaswamy Balaji[1], MBBS, MD, DNB, EDRA; David Hepburn[1], BMEDSC, MBChB, MD

Hull York Medical School, Hull University Teaching Hospitals NHS Trust, Hull, United Kingdom

Corresponding Author:
Shereen Ajab, MBBS, MRCS
Hull York Medical School
Hull University Teaching Hospitals NHS Trust
Allam Medical Building, Cottingham Road
Hull, HU6 7RX
United Kingdom
Phone: 44 1482875875
Email: shereen.ajab@nhs.net

Abstract

Background: Bedside teaching is integral to medical education and has been highlighted to improve clinical and communication skills, as well as clinical reasoning. Despite the significant advantages of bedside teaching, its usage within medical education has been declining, and COVID-19 has added additional challenges. The pandemic has resulted in a significant reduction in opportunities to deliver bedside teaching due to risk of viral exposure, patients declining student interactions, and ward closures. Educators have therefore been required to be innovative in their teaching methods, leading to the use of online learning, social media platforms, and simulation. Simulation-based education allows for learning in a low-risk environment and affords the opportunity for deliberated repeated practice with case standardization. The results demonstrate that simulation-based training can increase students' confidence, increase the rates of correct clinical diagnoses, and improve retention of skills and knowledge when compared with traditional teaching methods.

Objective: To mitigate the impact of COVID-19 upon bedside teaching for third year students at Hull York Medical School amid closure of the cardiorespiratory wards, a high-fidelity simulation-based model of traditional bedside teaching was designed and implemented. The objectives of the teaching session were to enable students to perform history taking and a focused cardiorespiratory clinical examination in a COVID-19–safe environment using SimMan 3G.

Methods: Four clinical teaching fellows with experience of simulation-based medical education scripted histories for 2 common cardiorespiratory cases, which were asthma and aortic stenosis. The simulation sessions were designed for students to take a focused cardiorespiratory history and clinical examination using SimMan 3G. All cases involved dynamic vital signs, and the simulator allowed for auscultation of an ejection systolic murmur and wheezing in accordance with the cases chosen. Key aspects of the pathologies, including epidemiology, differential diagnoses, investigations, and management, were summarized using an interactive PowerPoint presentation, followed by a debriefing session.

Results: In total, 12 third year medical students undertook the sessions, and overall feedback was highly positive. Of the 10 students who completed the feedback questionnaires, 90% (n=9) felt more confident in their clinical examination skills following the teaching; 100% (n=10) of the students responded that they would recommend the session to a colleague; and implementation of regular simulation was frequently requested on feedback. These results are in keeping with the current literature.

Conclusions: Bedside teaching continues to face ongoing challenges from the COVID-19 pandemic as well as declining patient recruitment and fluctuations in clinical findings. The support for simulation-based medical education is derived from high-quality studies; however, studies describing the use of this technology for bedside teaching in the undergraduate curriculum are limited. The authors describe a highly effective teaching session amid the pandemic, which allowed for maintenance of staff and student safety alongside continued education during a challenging time for educators globally.

(JMIR Med Educ 2022;8(2):e33565) doi:10.2196/33565

KEYWORDS

simulation; high fidelity; low fidelity; COVID-19; bedside teaching; undergraduate medical education; fidelity; medical education; medical student; review; innovation; risk; design; implementation

Introduction

The COVID-19 pandemic has posed significant challenges to undergraduate medical education. Teaching methods with patient-facing encounters such as bedside teaching have raised numerous difficulties with regards to exposure and testing of students, staff and patients, limited access to personal protective equipment, and strict social distancing requirements. The hurdles associated with conducting clinical placements amid the pandemic have been acknowledged by the UK Medical Schools Council [1]. In an attempt to ensure ongoing education during this time, web-based learning platforms have been increasingly adopted; however, not all areas of the undergraduate medical curriculum are best suited to this form of learning [2].

Bedside teaching has commonly been used to teach valuable core skills such as history taking and clinical examination, fostering effective communication skills and enabling professional relationships with patients and other health care professionals to be established [3]. Solutions to the difficulties associated with bedside teaching in the pandemic have been sought through the use of available technologies, including simulators. Simulation has been defined as a person, device, or set of conditions that attempts to present education and evaluation authentically [4]. Simulation-based teaching in health professions education has seen tremendous growth over the past 20 years, driven by factors including patient safety and the requirement for standardization of both training and assessment [5]. Simulation-based education, through various high-quality randomized controlled trials, has been shown to accelerate knowledge and skill acquisition including nontechnical skills, engage learners in deliberate practice, and provide a controlled low-risk learning environment [6-8].

Simulators can be classified according to the degree of realism known as fidelity. High-fidelity simulators provide the user with immersive and often complex scenarios with a high degree of realism [9]. Low-fidelity simulators replicate the real world to a lesser extent and include part-task trainers, for example, an inanimate model of a limb to allow learners to practice venipuncture skills [9]. In addition, medium- or intermediate-fidelity simulators provide greater authenticity than low-fidelity simulators but may require instructors to produce physiological signals displayed on monitors and therefore lack authenticity compared with high-fidelity simulators [10]. High- and low-fidelity simulators have been shown to have various advantages with recognized limitations. Weller has demonstrated that students have also derived benefit from medium-fidelity simulators, with feedback suggesting students found it useful to apply their knowledge in a safe environment using a structured approach, as well as developing their team-working skills [11]. High fidelity simulation can provide students with exposure to relevant clinical signs without requiring patient contact, while maintaining a high degree of realism [1]. Studies have demonstrated that students reported higher satisfaction and self- rated confidence scores when using high-fidelity simulators compared with low-fidelity models [12]. The advantages of simulation-based education have resulted in its incorporation into some of the medical school curriculums. Hull York Medical School has adopted simulation-based education for final year medical students and physician associate students using the high-fidelity simulator SimMan 3G.

SimMan 3G is an adult-size full body mannikin, which can be preprogrammed with adjustable parameters representing vital functions that can be visualized on a display monitor. Vital sign monitoring through the application of a saturation probe, electrocardiogram leads, and a blood pressure cuff onto the mannikin will allow for the parameters to be displayed on a monitor connected to the mannikin. The monitor displays dynamic heart rate, 3- and 12-lead electrocardiograms, blood pressure, mean arterial pressure, oxygen saturations, respiratory rate, and end tidal carbon dioxide values, allowing students immediate feedback when interventions are performed. The simulator allows for speech using a microphone and speaker, which is connected wirelessly to a separate control room where an instructor can simulate the patient’s voice. The technology audibly simulates cardiac murmurs, pathological respiratory sounds, and chest wall motion abnormalities as well as peripheral and central pulse palpation. Visually, the mannikin can simulate cyanosis, pupillary changes, diaphoresis, tongue oedema, pharyngeal swelling, trismus, and seizures.

Hull York Medical School delivers weekly bedside teaching for third year undergraduate students, facilitated by a team of clinical teaching fellows. During the academic year, students rotate through the 4 blocks of cardiorespiratory, gastroenterology, metabolic, and mental health. The focus of these teaching sessions is enabling students to develop their history taking and clinical examination skills in clinic environments. During the pandemic, several wards including the cardiology, cardiothoracic surgery, and respiratory medicine wards were closed to students due to outbreaks and to minimize risk of viral spread. The greatest impact was subsequently for those students undertaking their cardiorespiratory module. COVID-19 has therefore demanded educators to be innovative in their teaching methods, and the authors describe their personal experience of implementing novel simulation-based “bedside” teaching sessions to address the forementioned issues. It is likely that in the postpandemic era, technology will continue to play an important role in education [13]. The aim of this paper is to describe our experience of designing and delivering high-fidelity simulation-based teaching for history taking and clinical examination for simulated cases of asthma and aortic stenosis.

Methods

Materials

The team of 4 clinical teaching fellows, with experience in delivering simulation-based medical education at Hull York Medical school, designed and implemented a novel approach to bedside teaching. The aim of the teaching sessions was to provide continued high-quality education during the COVID-19 pandemic to third year medical students. The group of students that had already undertaken their cardiorespiratory block were identified, as this rotation received significant impact due to the pandemic, resulting in limited student exposure to patients from this specialty.

Two cases were chosen for the teaching sessions—asthma and aortic stenosis—due to their relative epidemiological prevalence and the ability to replicate clinical signs in the simulation suite (wheeze and ejection systolic murmur, respectively). Undergraduate medical students in their third year of education at Hull York Medical School were recruited via email by a student coordinator and offered electronic sign-up dates and times. The teaching sessions were delivered to pairs of students to allow sufficient time for each student to take a history and perform a focused respiratory or cardiac clinical examination. In addition, the small group sizes allowed for each student to observe their peers while maintaining compliance with social distancing requirements.

The format of the teaching sessions involved 2 students entering a simulation suite, accompanied by a teaching fellow to guide and support the session. In the suite, each student had the opportunity to take a history and perform a focused cardiorespiratory examination using the high-fidelity simulator SimMan 3G. The patient history was provided by a clinical teaching fellow in the control room in real time via a microphone linked to speakers in the simulation suite. The histories had been prescripted in order to allow for standardization of the cases and for learning objectives to be met. One teaching fellow was responsible for programming the simulator’s vital signs, providing the students with dynamic heart rate, blood pressure, oxygen saturations, respiratory rate, and temperature measurements. For each case the relevant positive clinical findings were simulated allowing for wheezing and an ejection systolic murmur to be auscultated. The limitations of the simulator were identified and therefore, to provide greater realism to the cases, visual aids were used in the form of printed photographs to demonstrate clubbing and a thoracotomy scar as well as props including a salbutamol inhaler. Each simulated session lasted approximately 45 minutes, and on completion of both cases, the students exited the suite and entered the debriefing room.

In the debrief room, a teaching fellow led a discussion of both cases with an emphasis on both individual and peer reflection and provided an opportunity for the students to ask questions. To summarize the key aspects of the topics covered during the simulation session, an interactive presentation was then delivered. The students spent approximately 45 minutes in the debrief room, which afforded the opportunity for the next pair of students to enter the simulation suite simultaneously. The equipment in the simulation room was cleaned between each pair of students, and all students performing the simulation wore appropriate personal protective equipment. Following the session, all students were sent a web-based feedback form and asked to rate their session in usefulness and relevance using the 5-point Likert scale (1=very poor, 5=very good), with additional white space fields to provide comments for qualitative feedback.

The web-based form included standardized questions that Hull York Medical School uses to collect qualitative feedback, including learners who undertake regular simulation-based teaching in the fifth year of the medical program, as shown in Multimedia Appendix 1.

Ethics Approval

This study was part of a quality improvement pilot project, and therefore no formal ethics board approval was required.

Results

In total, 12 students completed the simulation-based teaching sessions, with 10 completing the feedback questionnaires. Analysis of the feedback demonstrated a very positive experience with an overall student self-rating score of 4.5 on a 5-point rating scale (1=very poor, 5=very good). The majority of students (n=9, 90%) felt more confident after the simulated bedside teaching, predominately with regards to clinical examination skills. Feedback also demonstrated that students found the combination of simulated scenarios alongside interactive presentations useful for their learning, and 100% (n=10) of students were keen to recommend the session to a colleague.

Student comments from qualitative feedback included “great opportunity for hands-on learning for practical skills” and “allows directed bedside teaching that is otherwise not available or not as easy to do with a real patient.” Moreover, 40% (n=4) of the students requested further simulation-based teaching sessions to be conducted using different clinical scenarios, and 1 (10%) student suggested “maybe we can have session like this in each block to practice for our end of block Objective Structured Long Examination Record.”

Discussion

Our experience of delivering high-fidelity simulation-based teaching for third year undergraduate medical students demonstrated that the sessions were well received by students with high levels of learner satisfaction. In addition, the majority of students’ self-rated scores of confidence following the simulation sessions were high. Padilha et al [14] describe the development of knowledge as influenced by both student’s intrinsic factors as well as extrinsic factors such as satisfaction. Our experience is in keeping with studies described in the literature including that by Meyers et al [15], an observational pilot study that showed supplemental simulation-based training using a high-fidelity manikin improved overall satisfaction in preclinical medical students.

Physical examination is a vital skill for clinicians and is an essential component of high-quality patient care [16]. Traditionally, this skill has been taught in the clinical environment through bedside teaching; however, there are several challenges to this modality including declining patient recruitment, fluctuations in clinical findings during the course of treatment, and most recently the COVID-19 pandemic [16]. A recent systematic review by Dedeilia et al [12] analyzed these challenges imposed on medical and surgical education and summarized the innovations that have enabled the continuation of education in the era of COVID-19. Innovations include teleconferences and webinars, online learning, social media platforms, virtual consultations, virtual reality, and simulation [12].

Simulators vary with regards to fidelity. High-fidelity simulation, as described in our teaching, has several advantages including the ability to control physiological parameters, standardization of cases, and the relative ease of accessibility, and it allows students to contact with rare or life-threatening situations in a low-risk environment [17,18]. This can be contrasted with bedside teaching when patients who are available can be very variable and the “positive” examination findings can also be limited [17]. In addition, rarer and more complex patients may be too unwell to consult with students, or due to their inherent epidemiological rarity, may not be present within the hospital setting [19]. Simulation allows for the replication of these clinical profiles, providing students with hands-on exposure they may not otherwise gain [19]. Simulation has also been shown to reduce anxiety levels among medical students as shown by Yu et al [20]. The latter propose that students need to be repeatedly exposed to simulation for psychological stability and to develop competence [20]. These findings are further supported by Zheng et al [21], where structural integration of high-fidelity simulation in the cardiovascular physiology curriculum for undergraduate medical students proved successful with regards to student’s learning experiences and learning outcomes.

The most widely studied high-fidelity simulator is Harvey, a life-size manikin that simulates 27 cardiac conditions, which was introduced in 1968 by the University of Miami [18,22]. Giovanni et al [23] randomized 37 students to Harvey compared with CD tuition and assessed students via a 6-station objective structured clinical examination (OSCE), 6 weeks post teaching. The authors reported moderate (though not statistically significant) advantage in interpreting clinical signs in real patients in students trained with the simulator [23], as shown in Multimedia Appendix 2. There was no difference in diagnostic accuracy, which the authors postulated could be due to learning decay, resulting from the 6-week delay in testing. Giovanni et al [23] concluded that low-fidelity simulators also have a role and are associated with greatly reduced costs compared with high-fidelity simulators. However, studies by Anastakis et al [24] and Matsumo et al [17] have demonstrated that low-fidelity simulators had smaller gains than the high-fidelity simulator group (though not statistically significant).

O’Flynn [18] also notes that simulation training is able to increase student’s confidence but recognizes the risk of skill decay. In order to overcome this, Kneebone et al [19] propose distributed learning resources, allowing learners to access a range of simulation suitable for their level of training. Learners involved in simulation-based education, when compared with traditional learning, have shown that greater retention and simulation can provide valuable opportunities for interdisciplinary interactions [25]. Reed et al [26] demonstrated the long-term benefits of simulation-based training when 98% of students scored at or above the minimum passing standards on retesting 1 to 9 months after receiving teaching on core emergency medicine skills using the Laerdal SimMan.

Simulation-based medical education has been shown to have long-term beneficial outcomes by reducing inherent risk to patients and reducing the frequency of medical error, thereby improving patient care [19]. Bernardi et al [22] used the Kyoto-Kagaku patient simulator to train a group of fifth year medial students on cardiac auscultation. Simulation exposure significantly improved heart auscultation skills with mitral regurgitation being correctly identified by 87.9% of students versus 71.4% of non–simulation trained students (P=.02) [22]. Increased diagnostic accuracy following simulation training was also demonstrated by Perlini et al [16] after incorporating a 10-hour teaching session using the Harvey simulator for medical students and residents. They found that after simulation-based teaching, learners had a greater ability to recognize the correct cardiac diagnoses (from 11% to 72% P<.001) compared with baseline [16].

Gauthier et al [11] randomized 32 first year medical students to teaching modules with standardized patients or Harvey simulators. The authors found no difference in mean OSCE scores but a higher frequency of correct diagnoses among the students trained with standardized patients [11]. However, student feedback revealed Harvey offered superior clinical findings, and the authors concluded a combined teaching program would be ideal for transferability to patients [11]. Butter et al [27] have demonstrated transfer of skills and knowledge learnt through simulation training to the clinical environment. Learners having undergone simulation training accurately assessed 93.8% of simulated heart sounds (P<.001) compared with 73.9% accuracy among untrained students [28]. The authors subsequently advocate for simulation-based mastery learning programs, contending that they are a practice and feasible modality, allowing sufficient time for practice and compliant with competency-based accreditation requirements [28].

Studies using high-fidelity simulators other than Harvey have also shown favorable results. In 2019, Arangalage et al [29] delivered 28 hours per year of simulation-based teaching to over 400 students by incorporating a simulation-based course into the undergraduate medical curriculum. The authors used the Lifeform Auscultation Trainer and Smartscope to teach cardiac auscultation, blood pressure measurements, peripheral arterial examination, and the clinical examination of heart failure [29]. In keeping with the results of our study, Arangalage et al [29] reported that the majority of students provided positive feedback and found the teaching useful. They concluded that the simulation-based teaching facilitated educator-student and student-student interactivity with fulfilment of pedagogical objectives [29].

Simulation has also been used effectively to teach specialist skills to medical students. Scholz et al [30] compared a high-fidelity simulator to a wood-and-leather phantom to teach intrapartum care to 46 undergraduate students. Students using the high-fidelity simulator felt better prepared for obstetric house jobs and performed better in obstetric skills evaluations [30]. Siassakos et al [25] conducted an exploratory randomized controlled trial involving 24 fourth year medical students randomizing 1:1 to hybrid simulation training or small-group tutorials to teach management of shoulder dystocia. The authors' results demonstrated that students having undergone hybrid simulation training had significantly higher median total patient perception scores (11 vs 9, respectively; *P*=.02) and significantly higher median communication scores (4 vs 3; *P*=.01) compared with those who underwent small-group tutorials [25]. The time dedicated to debriefing and the provision of immediate feedback are also considered a significant strength of simulation-based teaching, an opportunity that is often lacking in the clinical setting [9]. According to Riaz et al [31], over 90% of students found debriefing a useful component.

A significant barrier to using high-fidelity simulation is the high costs involved. Karnath et al [32] overcame this by using transportable simulators (blood pressure simulator and palpable pulse simulator) to effectively deliver a cardiopulmonary module for second year medical students [32]. Students were assessed through an OSCE; 80% of the students accurately measured the blood pressure, and cardiopulmonary auscultation proficiency showed average recognition of 60% for cardiac abnormalities and 88% for pulmonary sounds [32].

The limitations of simulation-based teaching are recognized. This method of teaching is resource intensive, both with regards to staff and the inherent high costs involved with high-fidelity simulators [33]. Simulation-based teaching has a higher staff and technological requirement, in addition to requiring technical knowledge to run the simulation effectively [33]. These factors may explain why simulation-based education has not been as widely adopted in undergraduate education compared with postgraduate training. Hull York Medical School has already incorporated simulated-based teaching into the undergraduate program for final year students. The authors propose that the use of simulation for other year groups may also be advantageous as demonstrated by the self-rated responses from third year students and allows for more efficient use of this expensive resource.

Limitations of the simulation technology also require consideration. In clinical practice, patients may present atypically, and a disadvantage of simulation-based bedside teaching is that these subtle nuisances and atypical presentations are not conveyed as well as when compared with a true patient presentation [15]. Simulation-based teaching may therefore not always represent a suitable alternative, particularly when there is advocation for preserving bedside teaching even in the face of new technologically assisted learning methods, with beliefs being held that nothing can simulate real patient encounters [20].

The support for simulation-based medical education is derived from high-quality studies (Multimedia Appendix 2); however, studies describing the use of this technology for bedside teaching in the undergraduate curriculum are limited [9,18,22,28,31,33-37]. We have described a successful teaching session, well received and enjoyed by the students with increased self-rated confidence scores in keeping with other studies [22,29,33]. During the unprecedented times of the pandemic, alternatives to bedside teaching were grossly limited. The authors describe a highly effective teaching session amid the pandemic, which allowed for maintenance of staff and student safety alongside continued education during a challenging time for educators globally. The teaching sessions allowed for learning in a safe controlled learning environment while meeting learning objectives. The cases chosen represent common pathologies, and with careful design and planning, future scenarios could incorporate more complex and rarer patients to allow for a more diverse learning experience. Simulation-based education is a useful adjunct to traditional teaching modalities providing an immersive and highly interactive learning environment that more accurately reflects the clinical experience [9]. The use of emergent technology is most likely to be an indispensable component of post–COVID-19 undergraduate medical education [21].

Authors' Contributions

SA: manuscript writing, data collection, data analysis, and critical revision; EP: manuscript writing and data collection; SD: manuscript writing and data analysis; AM: manuscript writing and data collection; JK: critical revision of the manuscript; PB: conception and critical revision of the manuscript; DH: critical revision of the manuscript. All authors read, revised, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Year 3 simulation-based bedside teaching session feedback questionnaire.

[DOCX File , 16 KB - mededu_v8i2e33565_app1.docx]

Multimedia Appendix 2

Studies using simulation in undergraduate medical education.

[DOCX File , 18 KB - mededu_v8i2e33565_app2.docx]

References

1. Caplan J, Clements R, Chadwick C, Kadirgamar J, Morgan J, Rao A. Medical education in 2020: developing COVID secure undergraduate hospital placements. Med Sci Educ 2020 Dec;30(4):1677-1683 [FREE Full text] [doi: 10.1007/s40670-020-01080-2] [Medline: 32929391]
2. Papapanou M, Routsi E, Tsamakis K, Fotis L, Marinos G, Lidoriki I, et al. Medical education challenges and innovations during COVID-19 pandemic. Postgrad Med J 2021 Mar 29:Online ahead of print. [doi: 10.1136/postgradmedj-2021-140032] [Medline: 33782202]
3. Keifenheim KE, Teufel M, Ip J, Speiser N, Leehr EJ, Zipfel S, et al. Teaching history taking to medical students: a systematic review. BMC Med Educ 2015 Sep 28;15:159 [FREE Full text] [doi: 10.1186/s12909-015-0443-x] [Medline: 26415941]
4. Motola I, Devine LA, Chung HS, Sullivan JE, Issenberg SB. Simulation in healthcare education: a best evidence practical guide. AMEE Guide No. 82. Med Teach 2013 Oct;35(10):e1511-e1530. [doi: 10.3109/0142159X.2013.818632] [Medline: 23941678]
5. Acton RD. The evolving role of simulation in teaching surgery in undergraduate medical education. Surg Clin North Am 2015 Aug;95(4):739-750. [doi: 10.1016/j.suc.2015.04.001] [Medline: 26210967]
6. Shore EM, Davidson A, Arnason M, Kara H, Shah A, Shah R. Bridging the gap: incorporating simulation into obstetrics and gynaecology undergraduate medical education. J Obstet Gynaecol Can 2019 Feb;41(2):191-196.e2. [doi: 10.1016/j.jogc.2018.03.016] [Medline: 30316714]
7. McGaghie WC, Siddall VJ, Mazmanian PE, Myers J, American College of Chest Physicians HealthScience Policy Committee. Lessons for continuing medical education from simulation research in undergraduate and graduate medical education: effectiveness of continuing medical education: American College of Chest Physicians Evidence-Based Educational Guidelines. Chest 2009 Mar;135(3 Suppl):62S-68S. [doi: 10.1378/chest.08-2521] [Medline: 19265078]
8. Weller JM. Simulation in undergraduate medical education: bridging the gap between theory and practice. Med Educ 2004 Jan;38(1):32-38. [doi: 10.1111/j.1365-2923.2004.01739.x] [Medline: 14962024]
9. Massoth C, Röder H, Ohlenburg H, Hessler M, Zarbock A, Pöpping DM, et al. High-fidelity is not superior to low-fidelity simulation but leads to overconfidence in medical students. BMC Med Educ 2019 Jan 21;19(1):29 [FREE Full text] [doi: 10.1186/s12909-019-1464-7] [Medline: 30665397]
10. Goh P, Sandars J. A vision of the use of technology in medical education after the COVID-19 pandemic. MedEdPublish 2020 Mar 26;9:49. [doi: 10.15694/mep.2020.000049.1]
11. Gauthier N, Johnson C, Stadnick E, Keenan M, Wood T, Sostok M, et al. Does cardiac physical exam teaching using a cardiac simulator improve medical students' diagnostic skills? Cureus 2019 May 07;11(5):e4610 [FREE Full text] [doi: 10.7759/cureus.4610] [Medline: 31312537]
12. Dedeilia A, Sotiropoulos MG, Hanrahan JG, Janga D, Dedeilias P, Sideris M. Medical and surgical education challenges and innovations in the COVID-19 era: a systematic review. In Vivo 2020 Jun;34(3 Suppl):1603-1611 [FREE Full text] [doi: 10.21873/invivo.11950] [Medline: 32503818]
13. Bokken L, Rethans J, Scherpbier AJJA, van der Vleuten CPM. Strengths and weaknesses of simulated and real patients in the teaching of skills to medical students: a review. Simul Healthc 2008;3(3):161-169. [doi: 10.1097/SIH.0b013e318182fc56] [Medline: 19088660]
14. Padilha JM, Machado PP, Ribeiro A, Ramos J, Costa P. Clinical virtual simulation in nursing education: randomized controlled trial. J Med Internet Res 2019;21(3):e11529. [doi: 10.2196/11529] [Medline: 30882355]
15. Meyers L, Mahoney B, Schaffernocker T, Way D, Winfield S, Uribe A, et al. The effect of supplemental high Fidelity simulation training in medical students. BMC Med Educ 2020 Nov 10;20(1):421 [FREE Full text] [doi: 10.1186/s12909-020-02322-y] [Medline: 33172450]
16. Perlini S, Salinaro F, Santalucia P, Musca F. Simulation-guided cardiac auscultation improves medical students' clinical skills: the Pavia pilot experience. Intern Emerg Med 2014 Mar;9(2):165-172. [doi: 10.1007/s11739-012-0811-z] [Medline: 22767224]
17. Matsumoto ED, Hamstra SJ, Radomski SB, Cusimano MD. The effect of bench model fidelity on endourological skills: a randomized controlled study. J Urol 2002 Mar;167(3):1243-1247. [Medline: 11832706]
18. O'Flynn S, Shorten G. Simulation in undergraduate medical education. Eur J Anaesthesiol 2009 Feb;26(2):93-95. [doi: 10.1097/EJA.0b013e32831a47df] [Medline: 19142080]
19. Kneebone RL, Scott W, Darzi A, Horrocks M. Simulation and clinical practice: strengthening the relationship. Med Educ 2004 Oct;38(10):1095-1102. [doi: 10.1111/j.1365-2929.2004.01959.x] [Medline: 15461655]
20. Yu JH, Chang HJ, Kim SS, Park JE, Chung WY, Lee SK, et al. Effects of high-fidelity simulation education on medical students' anxiety and confidence. PLoS One 2021;16(5):e0251078 [FREE Full text] [doi: 10.1371/journal.pone.0251078] [Medline: 33983983]
21. Zheng J, Lapu R, Khalid H. Integrating high-fidelity simulation into a medical cardiovascular physiology curriculum. Adv Med Educ Pract 2020;11:41-50 [FREE Full text] [doi: 10.2147/AMEP.S230084] [Medline: 32021542]
22. Bernardi S, Giudici F, Leone MF, Zuolo G, Furlotti S, Carretta R, et al. A prospective study on the efficacy of patient simulation in heart and lung auscultation. BMC Med Educ 2019 Jul 23;19(1):275 [FREE Full text] [doi: 10.1186/s12909-019-1708-6] [Medline: 31337361]

23. de Giovanni D, Roberts T, Norman G. Relative effectiveness of high- versus low-fidelity simulation in learning heart sounds. Med Educ 2009 Jul;43(7):661-668. [doi: 10.1111/j.1365-2923.2009.03398.x] [Medline: 19573189]
24. Anastakis DJ, Regehr G, Reznick RK, Cusimano M, Murnaghan J, Brown M, et al. Assessment of technical skills transfer from the bench training model to the human model. Am J Surg 1999 Feb;177(2):167-170. [doi: 10.1016/s0002-9610(98)00327-4] [Medline: 10204564]
25. Siassakos D, Draycott T, O'Brien K, Kenyon C, Bartlett C, Fox R. Exploratory randomized controlled trial of hybrid obstetric simulation training for undergraduate students. Simul Healthc 2010 Aug;5(4):193-198. [doi: 10.1097/SIH.0b013e3181d3ee0b] [Medline: 21330796]
26. Reed T, Pirotte M, McHugh M, Oh L, Lovett S, Hoyt AE, et al. Simulation-based mastery learning improves medical student performance and retention of core clinical skills. Simul Healthc 2016 Jun;11(3):173-180. [doi: 10.1097/SIH.0000000000000154] [Medline: 27093509]
27. Butter J, McGaghie WC, Cohen ER, Kaye M, Wayne DB. Simulation-based mastery learning improves cardiac auscultation skills in medical students. J Gen Intern Med 2010 Aug;25(8):780-785 [FREE Full text] [doi: 10.1007/s11606-010-1309-x] [Medline: 20339952]
28. Gray-Miceli D, Aselage M, Mezey M. Teaching strategies for atypical presentation of illness in older adults. J Gerontol Nurs 2010 Jul;36(7):38-43. [doi: 10.3928/00989134-20100601-01] [Medline: 20608591]
29. Arangalage D, Abtan J, Gaschignard J, Ceccaldi P, Remini S, Etienne I, et al. Implementation of a large-scale simulation-based cardiovascular clinical examination course for undergraduate medical students - a pilot study. BMC Med Educ 2019 Sep 18;19(1):361 [FREE Full text] [doi: 10.1186/s12909-019-1750-4] [Medline: 31533700]
30. Scholz C, Mann C, Kopp V, Kost B, Kainer F, Fischer MR. High-fidelity simulation increases obstetric self-assurance and skills in undergraduate medical students. J Perinat Med 2012 Nov;40(6):607-613. [doi: 10.1515/jpm-2012-0052] [Medline: 23093253]
31. Riaz S, Jaradat AAK, Gutierrez R, Garadah TS. Outcome of undergraduate medical education using medical simulation according to students' feedback. Sultan Qaboos Univ Med J 2020 Aug;20(3):e310-e315 [FREE Full text] [doi: 10.18295/squmj.2020.20.03.010] [Medline: 33110646]
32. Karnath B, Thornton W, Frye AW. Teaching and testing physical examination skills without the use of patients. Acad Med 2002 Jul;77(7):753. [doi: 10.1097/00001888-200207000-00044] [Medline: 12114177]
33. Qureshi Z, Maxwell S. Has bedside teaching had its day? Adv Health Sci Educ Theory Pract 2012 May;17(2):301-304. [doi: 10.1007/s10459-011-9308-1] [Medline: 21681592]
34. Okuda Y, Bryson EO, DeMaria S, Jacobson L, Quinones J, Shen B, et al. The utility of simulation in medical education: what is the evidence? Mt Sinai J Med 2009 Aug;76(4):330-343. [doi: 10.1002/msj.20127] [Medline: 19642147]
35. Gaur U, Majumder MAA, Sa B, Sarkar S, Williams A, Singh K. Challenges and opportunities of preclinical medical education: COVID-19 crisis and beyond. SN Compr Clin Med 2020;2(11):1992-1997 [FREE Full text] [doi: 10.1007/s42399-020-00528-1] [Medline: 32984766]
36. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. A critical review of simulation-based medical education research: 2003-2009. Med Educ 2010 Jan;44(1):50-63. [doi: 10.1111/j.1365-2923.2009.03547.x] [Medline: 20078756]
37. Ntlokonkulu ZB, Rala NM, Goon DT. Medium-fidelity simulation in clinical readiness: a phenomenological study of student midwives concerning teamwork. BMC Nurs 2018;17:31 [FREE Full text] [doi: 10.1186/s12912-018-0303-1] [Medline: 30065617]

Abbreviations

OSCE: objective structured clinical examination

Edited by N Zary; submitted 13.09.21; peer-reviewed by M Mosher; comments to author 20.11.21; revised version received 13.03.22; accepted 07.04.22; published 09.05.22.

Please cite as:
Ajab S, Pearson E, Dumont S, Mitchell A, Kastelik J, Balaji P, Hepburn D
An Alternative to Traditional Bedside Teaching During COVID-19: High-Fidelity Simulation-Based Study
JMIR Med Educ 2022;8(2):e33565
URL: https://mededu.jmir.org/2022/2/e33565
doi:10.2196/33565
PMID:35404828

Original Paper

Freestyle Deliberate Practice Cadaveric Hand Surgery Simulation Training for Orthopedic Residents: Cohort Study

Hannah K James[1], MBChB, MMedEd, MRCS, PhD; Ross A Fawdington[2], Dip Hand Surg, MBChB, MFST Ed, FRCS

[1]Clinical Trials Unit, Warwick Medical School, University of Warwick, Coventry, United Kingdom

[2]Department of Trauma and Orthopaedics, Queen Elizabeth Hospital Birmingham, Birmingham, United Kingdom

Corresponding Author:
Hannah K James, MBChB, MMedEd, MRCS, PhD
Clinical Trials Unit
Warwick Medical School
University of Warwick
Gibbet Hill Road
Coventry, CV4 7AL
United Kingdom
Phone: 44 02476574880
Email: hsmith22@doctors.org.uk

Abstract

Background: Cadaveric simulation training may be part of the solution to reduced quantity and quality of operative surgical training in the modern climate. Cadaveric simulation allows the early part of the surgical learning curve to be moved away from patients into the laboratory, and there is a growing body of evidence that it may be an effective adjunct to traditional methods for training surgical residents. It is typically resource constrained as cadaveric material and facilities are expensive. Therefore, there is a need to be sure that any given cadaveric training intervention is maximally impactful. Deliberate practice (DP) theory as applied to cadaveric simulation training might enhance the educational impact.

Objective: The objectives of this study were (1) to assess the impact of a freestyle DP cadaveric hand surgery simulation training intervention on self-reported operative confidence for 3 different procedures and (2) to assess the subjective transfer validity, perceived educational value, and simulation fidelity of the training.

Methods: This study used validated questionnaires to assess the training impact on a cohort of orthopedic residents. The freestyle course structure allowed the residents to prospectively define personalized learning objectives, which were then addressed through DP. The study was conducted at Keele Anatomy and Surgical Training Centre, a medical school with an integrated cadaveric training laboratory in England, United Kingdom. A total of 22 orthopedic surgery residents of postgraduate year (PGY) 5-10 from 3 regional surgical training programs participated in this study.

Results: The most junior (PGY 5-6) residents had the greatest self-reported confidence gains after training for the 3 procedures (distal radius open reduction internal fixation, flexor tendon repair, ulnar shortening osteotomy), and these gains diminished with resident seniority. The confidence gains were proportional to the perceived procedural complexity, with the most complex procedure having the lowest pretraining confidence score across all experience levels, and the greatest confidence increase in posttraining. Midstage (PGY 7-8) residents reported receiving the highest level of educational benefit from the training but perceived the simulation to be less realistic, compared to either the junior or senior residents. The most senior residents (PGY 9-10) reported the greatest satisfaction with the self-directed, freestyle nature of the training. All groups reported that they were extremely likely to transfer their technical skill gains to their workplace, that they would change their current practice based on these skills, and that their patients would benefit as a result of their having undertaken the training.

Conclusions: Freestyle, resident-directed cadaveric simulation provides optimum DP conditions whereby residents can target their individualized learning needs. By receiving intensive, directed feedback from faculty, they can make rapid skill gains in a short amount of time. Subjective transfer validity potential from the training was very high, and objective, quantitative evidence of this is required from future work.

(JMIR Med Educ 2022;8(2):e34791) doi:10.2196/34791

KEYWORDS

simulation; high fidelity simulation; orthopedic residency; surgical training; postgraduate education; medical education; medical student; surgeon; hand; hand surgery; surgery; orthopedic; cadaver; cadaveric simulation; cadaveric; training; cadaveric training; DP; deliberate practice

Introduction

Cadaveric training is rapidly gaining popularity as the ultimate surgical simulation [1]. Human cadavers accurately represent anatomy as seen in the operating room, which allows residents to appreciate neurovascular and soft tissue relationships and the associated hazards with unparalleled realism [2,3]. Modern fresh-freezing techniques preserve the soft tissue handling characteristics, meaning the intraoperative "haptic feedback" when operating on cadavers is highly realistic [4]. Furthermore, when the environmental and psychological fidelities are optimized by dressing the simulation as a real operating theatre, it leads to the acquisition of both technical and nontechnical skills in a complete training package [5]—residents are not just "learning surgical skills but learning to be surgeons" 6.

Cadaveric training may be part of the solution to the joint problems of reduced quantity and quality of surgical training in the United Kingdom. The European Working Time Directive has dramatically reduced the number of hours available for surgical training [6,7], and the time that is available is not being used to best effect [8]. This is because junior residents are increasingly spending their time doing administrative and other tasks that offer less training value at the expense of attending the operating theatre [9]. Such tasks might include requesting investigations, writing discharge summaries, and other nonsurgical tasks required for their professional development such as participation in audit and quality improvement work. A large 2016 study by the Royal College of Surgeons of 990 residents found that in the average 12-hour shift, 218 minutes were spent on administrative tasks compared to just 34 minutes operating [10].

These challenges of delivering training have led to concerns about the possible patient safety implications [11]. Cadaveric simulation allows the early part of the surgical learning curve to be moved away from patients and into the laboratory so that patient safety can be assured [12]. There is a growing body of evidence that cadaveric simulation is effective for training across a wide range of specialties [1].

One known problem is that cadaveric simulation is expensive to provide [2] and is necessarily restricted to specialized wet-laboratory facilities [13]. When designing a cadaveric training course, residents will be limited to 1 attempt at each procedure [14]. It is therefore essential to maximize the impact of that training opportunity to allow for the greatest educational gains in the most cost- and time-efficient way.

Deliberate practice (DP) is an educational theory–driven way to maximize the efficiency of surgical simulation training. DP theory says that attainment of expert performance results from a continued process of targeted practice of tasks with immediate feedback, which allows learners to focus on their weaknesses while also refining other aspects of their performance [15]. It is this process, rather than merely "time on the job," that leads to expertise, and the level of proficiency that can be attained through DP is independent of innate ability [16,17].

The aim of this study was to evaluate a "freestyle" DP cadaveric training intervention for hand surgery, where residents prospectively identified their individual learning needs using SMART (specific, measurable, achievable, realistic, time-related) objectives [18]. The course was freestyle in the sense that there was no didactic, taught element and no prescribed timetable of procedures to be performed. We hypothesize that this would provide optimum conditions for DP and would maximize and expedite the learning gains from the training.

Methods

Ethical Considerations

This work comes under the remit of course evaluation and therefore formal ethical board approval was not deemed necessary. The surgical training center holds the appropriate licenses to host cadaveric simulation training [13].

Recruitment

The study was designed as a prospective cohort study. Participants were recruited via an email invitation sent to all orthopedic residents (approximately 80) in 3 regional training programs in the United Kingdom. All specialist training grades from postgraduate year (PGY) 3-10 were eligible. In total, 22 participants were recruited and completed the training.

The Cadaveric Surgical Simulation Training Course

The training course took place over 1 day at Keele Anatomy and Surgical Training Centre. Fresh frozen whole cadaveric arms were used, obtained from the local body donation program. Instruments and implants were provided by Trimed (Trimed Inc). A large C-arm and radiographer were available to participants. The attending hand and wrist surgeon faculty each supervised 2 pairs of residents. The costs were funded by Health Education West Midlands, and the course was delivered free of cost to the participants.

Participants were asked to complete prelearning from a reading list and to write and submit bespoke SMART objectives of what they planned to achieve from the course before attending.

Participants were self-paired during the cadaveric sessions, with 2 residents to 1 cadaveric arm. Equipment to perform any or all of the procedures—distal radius open reduction internal fixation (ORIF), ulnar shortening osteotomy (USO), and flexor tendon repair (FTR)—was made available. Participants decided among themselves which procedures (or parts of procedures) they would perform. Participants were asked to pay specific attention to their SMART objectives. Attending faculty were circulating closely and were on hand to provide immediate feedback on performance. Importantly, there was no demonstration or guidance provided and no prescriptive structure to the session—the wet lab time was entirely free for the participants to explore the anatomy and perform the procedure at their own pace. This was done consciously to allow for the maximum time to be devoted to DP and is different from the usual provision in cadaveric simulation, where typically a guided demonstration is followed by participants performing all parts of all procedures in a sequential rotational manner, regardless of individual learning needs.

As part of the structured feedback and to self-audit against achievement of their SMART objectives, participants were offered the opportunity for procedure-based assessments (PBAs) to be completed by the attending faculty. PBAs are a framework for residents to receive structured feedback and allow for personal reflection.

Data Collection and Analysis

Data were collected using prepiloted questionnaires that were designed to provide a sophisticated, subjective, and principally qualitative assessment of cadaveric simulation training 20. A Likert scale of 1-10 was used, with no middle descriptive anchor to avoid response centralization. Some questions were deliberately negativized to encourage thoughtful completion.

Demographic details and assessment of pretraining procedural confidence scores were obtained at registration before the start of the course. Posttraining confidence scores and assessment of educational value, simulator fidelity, and transfer validity potential were made at the end of the course before debriefing. Data analysis was undertaken using IBM SPSS Statistics for Windows (version 26; IBM Corp).

Results

Overview

There were 22 participants in the study, from PGY 5-10. Of them, 19 participants were male and 3 were female. Participants were divided into 3 subgroups for analysis, which correspond to the stages of UK higher surgical training: early (PGY 5-6), mid (PGY 7-8), and late stage (PGY 9-10). Participant demographics by subgroup are shown in Table 1. A total of 6 (28%) participants were cadaveric simulation naive, and the likelihood of past exposure to cadaveric simulation did not relate to seniority level.

Table 1. Participant demographics.

Characteristics	Stage of training			
	Early (PGY[a] 5-6)	Mid (PGY 7-8)	Late (PGY 9-10)	Total
Gender, n (%)				
Male	13 (93)	2 (100)	4 (67)	19 (86)
Female	1 (7)	0 (0)	2 (33)	3 (14)
Total	14 (100)	2 (100)	6 (100)	22 (100)
Cadaveric simulation naive, n (%)				
Yes	5 (36)	0 (100)	1 (17)	6 (28)
No	9 (64)	2 (100)	5 (83)	16 (72)
Total	14 (100)	2 (100)	6 (100)	22 (100)

[a]PGY: postgraduate year.

Procedural Confidence

Procedural confidence increased for all procedures and within all subgroups following the DP cadaveric training. Pretraining procedural confidence was lowest for all groups for the least frequently performed procedure (USO). Mean reported confidence levels were 1.8, 3.5, and 5.2 for early-, mid-, and late-stage residents, respectively, on a Likert scale of 1-10 (where 1=not at all confident and 10=extremely confident). Posttraining confidence increased by +4.4, +4.5, and +2.5 points by subgroup (Figure 1). Pretraining confidence was highest across all subgroups for the procedure perceived to be most straightforward (distal radius ORIF), at 5.2, 7.0, and 8.0 for early-, mid-, and late-stage residents. There were confidence gains of +2.6, +2.5, and +1.7 points, respectively, after training (Figure 2). Confidence gains for FTR are shown in Figure 3.

The size of confidence gain by procedure was inversely proportional to the stage of training, with the largest gains seen in the most junior, early-stage residents (+2.3, +4.4, and +2.6 points for distal radius ORIF, USO, and FTR), moderate gains seen in the midstage residents (+2.0, +4.5, and +2.5 points), and the smallest gains seen in the most senior residents in all 3 procedures (+0.8, +2.5, and +1.7 points) (Table 2).

There were significant differences in between-group mean confidence gains for USO (P=.02), using 1-way ANOVA, with the most junior residents yielding the greatest gain. The between-group differences in confidence gains in distal radius ORIF and FTR were not statistically significant.

There was a significant correlation between specialist training year and mean change in confidence after training for distal radius ORIF (P=.01) and USO (P=.004) but not for FTR (Pearson test).

Figure 1. Box plot showing confidence change for ulnar shortening osteotomy. PGY: postgraduate year.

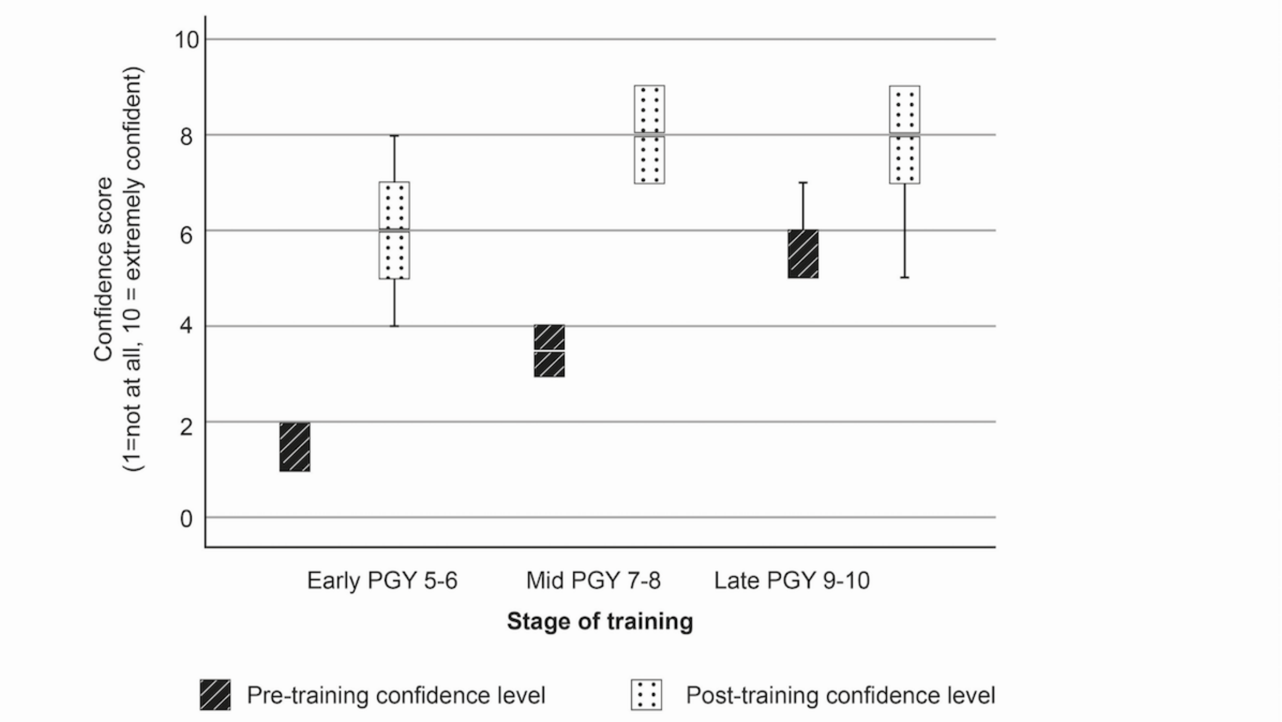


Figure 2. Box plot showing confidence change for distal radius open reduction internal fixation. PGY: postgraduate year.

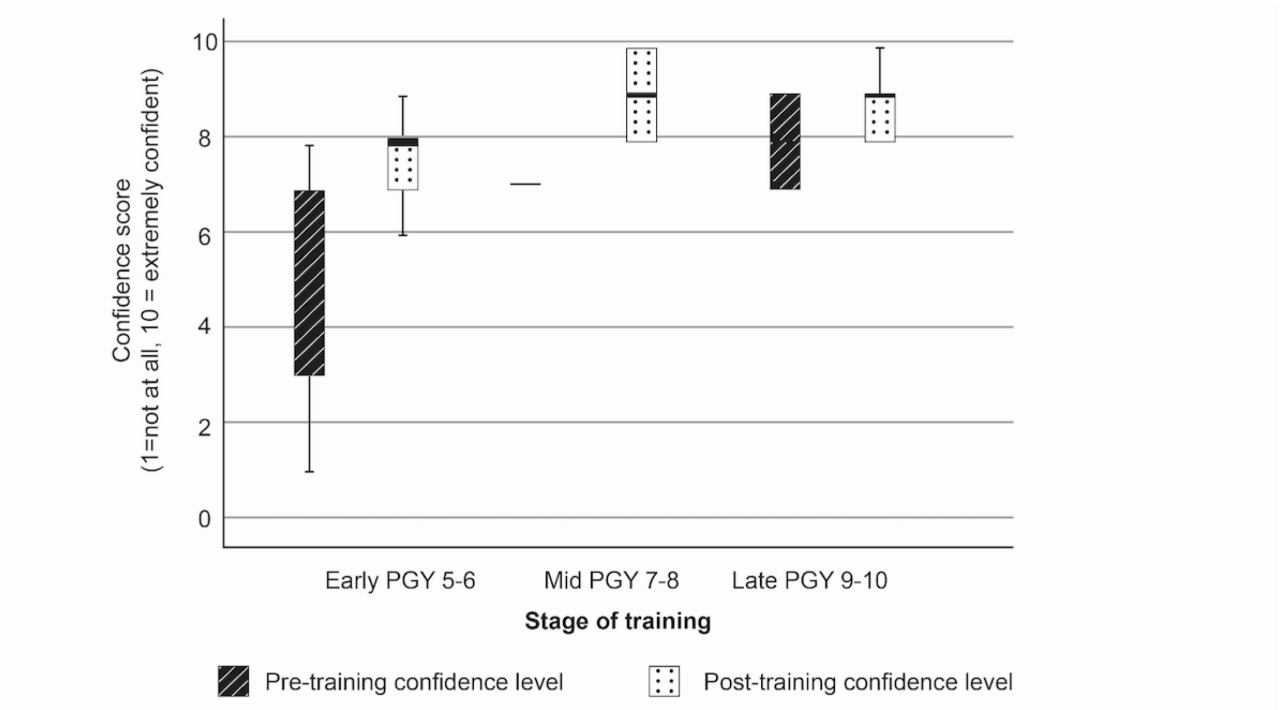


Figure 3. Box plot showing confidence change for flexor tendon repair. PGY: postgraduate year.

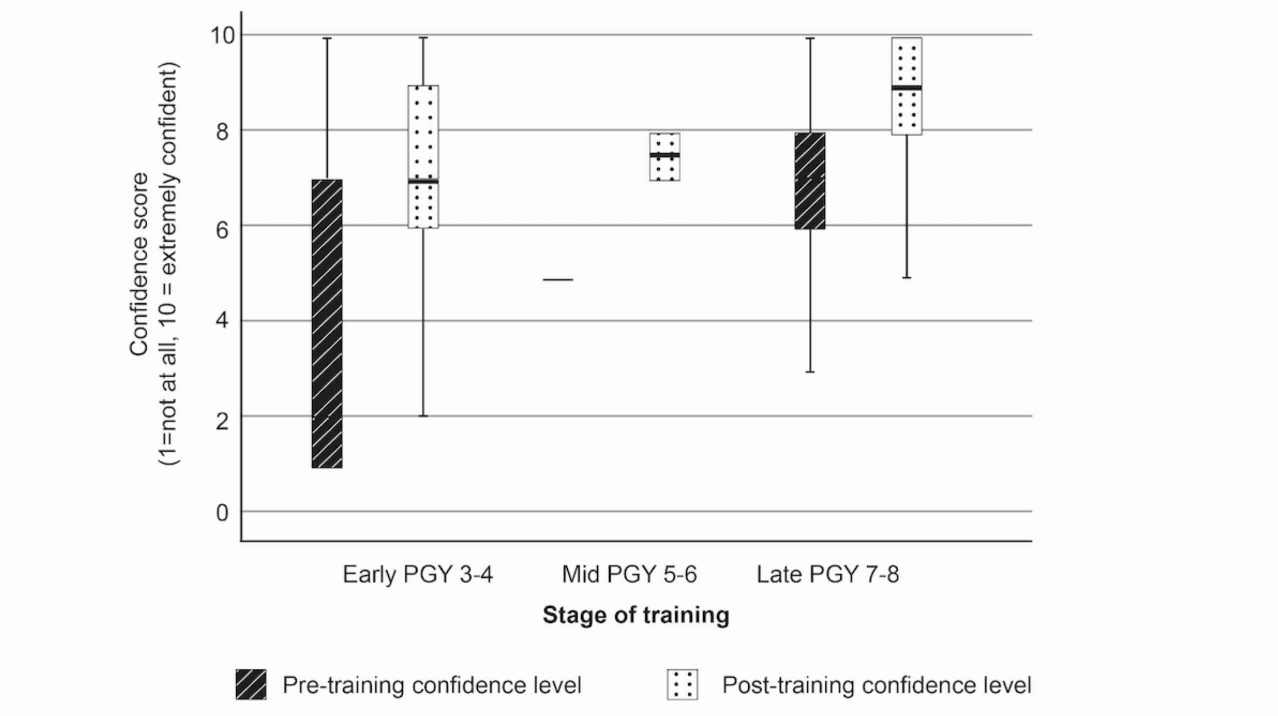


Table 2. Mean confidence gains by procedure and stage of training.

Procedure	Stage of training								
	Early (PGY[a] 5-6)			Mid (PGY 7-8)			Late (PGY 9-10)		
	Pre	Post	Change	Pre	Post	Change	Pre	Post	Change
Distal radius ORIF[b]	5.2	7.5	+2.3	7.0	9.0	+2.0	8.0	8.8	+0.8
Ulnar shortening osteotomy	1.8	6.2	+4.4	3.5	8.0	+4.5	5.2	7.7	+2.5
Flexor tendon repair	4.1	6.7	+2.6	5.0	7.5	+2.5	6.8	8.5	+1.7

[a]PGY: postgraduate year.

[b]ORIF: open reduction internal fixation.

Educational Value of DP Cadaveric Simulation

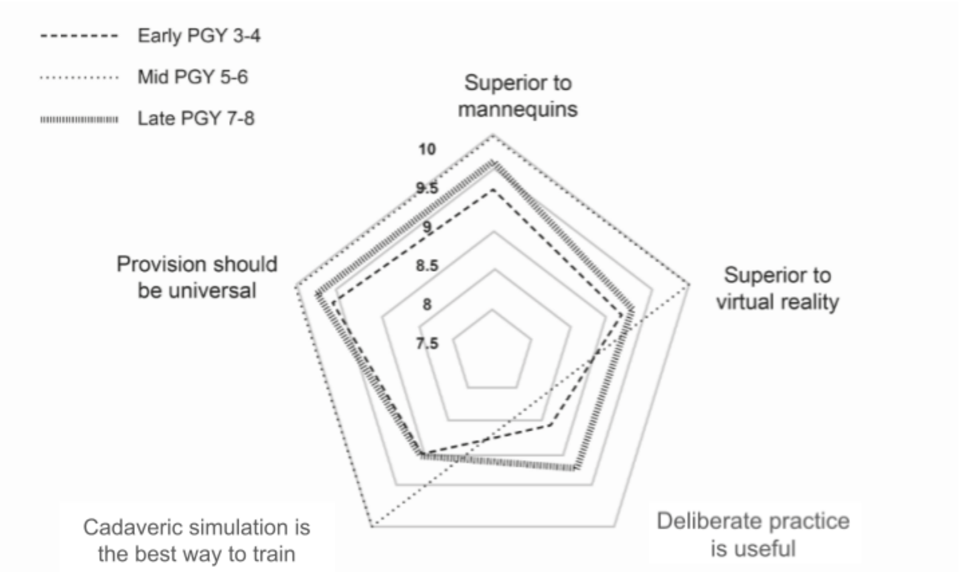
The perceived educational value of the training was assessed across 5 domains. All participants strongly agreed that the cadaveric training was superior to training on mannequins (mean 9.64, range 7-10 on a Likert scale of 1-10, where 1=strongly disagree and 10=strongly agree), and it was superior to training by virtual reality (mean 9.27, range 6-10). The majority of participants believed the freestyle DP nature of the course enhanced their learning, although this was not universal (mean 8.77, range 3-10). The late-stage residents were most enthusiastic about the DP design (Table 3). The participants strongly believed that cadaveric simulation training should be more widely provided to orthopedic residents (mean 9.59, range 8-10). Subgroups scores by domain are shown in Table 3 and Figure 4.

Table 3. Participant perception of educational value, simulator fidelity, and transfer validity of cadaveric training (scale 1-10, where 10 is considered the best score).

Participant perception	Stage of training			
	Early (PGY[a] 5-6), mean	Mid (PGY 7-8), mean	Late (PGY 9-10), mean	Total participants, mean (range)
Educational value				
Superior to mannequins	9.5	10	9.8	9.6 (7-10)
Superior to virtual reality	9.1	10	9.3	9.3 (6-10)
Deliberate practice is useful	8.6	8.5	9.2	8.8 (3-10)
Cadaveric simulation is the best way to train	9	10	9	9.1 (9-10)
Provision should be universal	9.5	10	9.7	9.6 (8-10)
Simulator fidelity				
Cadavers as patients	8.7	9	8.8	8.8 (6-10)
Surgical anatomy	9	9	9.3	9.1 (7-10)
Hospital environment	7.2	4.4	8	7.3 (3-10)
Multidisciplinary team	6.4	4	7.3	6.4 (1-10)
Psychological stress	5	4.5	5.7	5.1 (1-10)
Transfer fidelity				
Will take new technical skills back to workplace	9.2	10	9.8	9.5 (8-10)
Will take new nontechnical skills back to workplace	7.1	4.5	8.2	7.1 (1-10)
Will change current practice	8.8	9.5	9.5	9.1 (7-10)
My future patients will benefit	9.1	9.5	9.3	9.2 (7-10)

[a]PGY: postgraduate year.

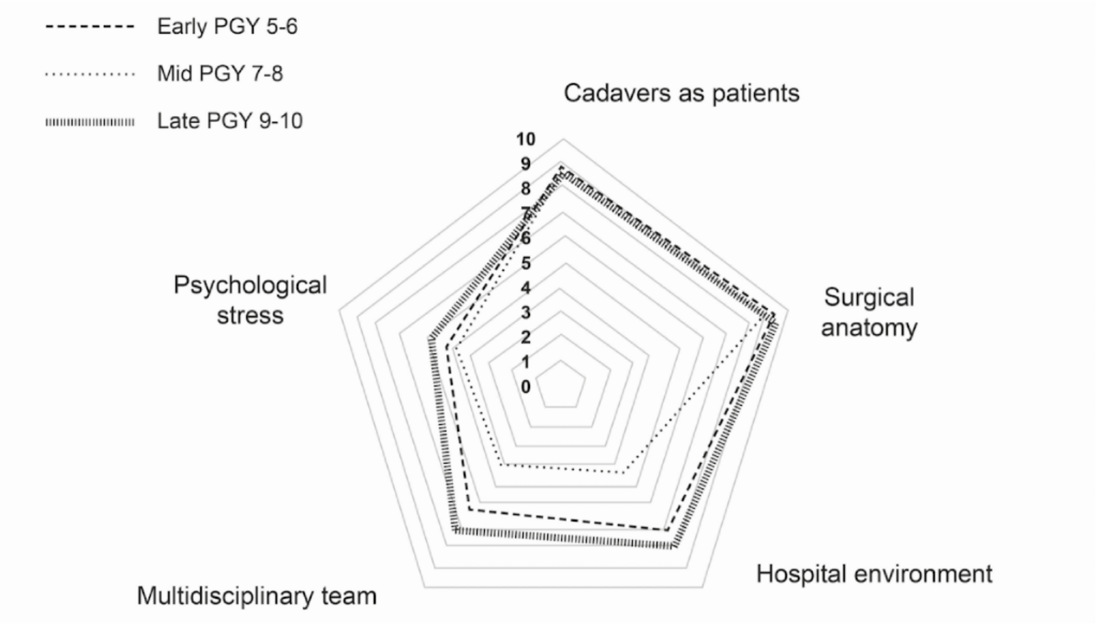
Figure 4. Radar plot showing educational value domain scores by training level. PGY: postgraduate year.



Simulator Fidelity

The fidelity of the simulation was considered across physical, environmental, and psychological domains. To assess physical fidelity, participants were asked about their perception of the realism of the cadaver as a patient and the realism of the surgical anatomy. Both were reported as being highly realistic (mean 8.77, range 6-10; and mean 9.09, range 7-10 for patient and anatomy, respectively). Environmental fidelity was assessed by asking participants about their perception of the realism of the hospital environment and multidisciplinary team. The environmental fidelity of the simulation was reported as being less than the physical fidelity but still reasonably high (mean 7.27, range 3-10; and mean 6.41, range 1-10 for hospital environment and multidisciplinary team, respectively). The psychological fidelity was assessed by asking the participants if they felt the simulation accurately recreated the emotional stress of performing real surgery. On average, the participants felt that the psychological stress of the DP cadaveric simulation was only moderately realistic, but there was a wide range of opinion on this (mean 5.14, range 1-10). The perception of psychological stress did not correlate with the stage of training. Assessment of simulator fidelity by resident stage of training is shown in Table 3 and Figure 5.

Figure 5. Radar plot showing simulator fidelity domain scores by training level. PGY: postgraduate year.



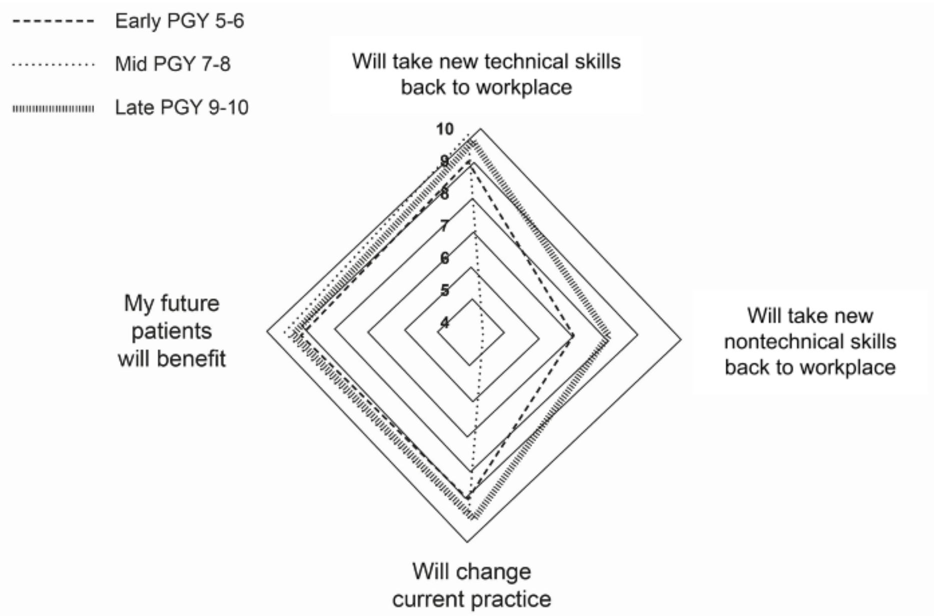
Skill Transfer Following Training

Transfer fidelity was examined in 4 areas: transfer of technical skills and nontechnical skills to the workplace, likelihood of changing current practice following training, and belief that their future patients would benefit from participants having done the training.

All participants strongly agreed they had learned technical skills during the training that they would transfer into their surgical practice (mean 9.45, range 8-10). There was moderate agreement that nontechnical skills had been gained from the training that would transfer to the workplace, with a wide range of views (mean 7.14, range 1-10).

Participants strongly agreed they would change 1 or more aspects of their current practice based on what they had learned during the training (mean 9.05, range 7-10), and they felt that their patients would benefit from their having done the training (mean 9.18, range 7-10). The late-stage trainees overall reported the highest likelihood of skill transfer following the training (Table 3 and Figure 6).

Figure 6. Radar plot showing transfer fidelity domain scores by training level. PGY: postgraduate year.



Discussion

Principal Findings

The evidence base for cadaveric simulation training is growing, with an increasing number of studies showing that it can induce short-term behavioral change when measured by objective means [1]. There is still uncertainty around the timing of delivery of training and how to optimize the learning impact given the known resource limitations. We sought to assess the latter question in this study to see if a DP-style cadaveric simulation course would expedite learning and “make the most” of the single procedural attempt that is typically available in cadaveric training courses.

Procedural confidence gains following training increased with procedural complexity and were inversely proportional to experience level, with the most junior residents reporting the greatest procedural confidence gains for all 3 procedures. This is not surprising and is in line with other studies assessing the impact of cadaveric simulation on junior residents [19-29].

The most senior residents reported the greatest enthusiasm for the DP-style training. This may be because they have greater insight into their learning needs and may have greater confidence and autonomy in pursuing independent practice when compared to more junior trainees. A comparative study of standard cadaveric versus DP cadaveric training would be needed to explore this topic further.

The physical and environmental fidelities of the simulation were reported to be high by all groups, and the psychological fidelity was less so. This may be because in the cadaveric simulation laboratory, the real-world pressures of unwell patients and other clinical commitments and service pressures are absent. This lack of psychological stress with concomitant high physical and environmental fidelity has actually been shown to be a key driver of learning in cadaveric simulation [5], as participants can take time to refine their skills and learn from their mistakes in a manner that is impossible to safely replicate in the real-life operating room.

Regarding transfer fidelity, all groups reported a very high likelihood that they would take technical skills back to their workplace, but it was less so with nontechnical skills. This is not surprising as we did not design the training to develop nontechnical skills; however, evidence does show that nontechnical skills learning during cadaveric simulation may occur passively and “unnoticed” as a result of immersion in the high fidelity, “symbolically structured environment,” which exerts an “anonymous, pervasive, pedagogic action” [30]. It may, therefore, be that the participants underreported their nontechnical skills acquisition following training. There is often a considerable bias toward assessing purely technical skills following simulation training [31]; however, given that previous ethnographic studies in surgery have shown that technical skill is only thought to be around 20% of the required skill set of a competent surgeon [32,33], consideration ought to be given to the role of simulation in addressing other dimensions of competence as well.

Our study has several strengths. To our knowledge, it is the first report in the literature of the application of DP theory specifically to cadaveric simulation training. We had the full range of resident experience levels included in our study cohort, which makes subgroup assessment of impact possible. We used a sophisticated prepiloted questionnaire instrument to evaluate the course in a high level of qualitative detail across domains that are grounded in educational theory. Our study also has several weaknesses. There was no comparator group receiving “standard”’ structured cadaveric training, so any inferences we make about the likely superiority of a DP-style cadaveric training are inherently speculative. Another weakness, in common with much of the existing evidence base on cadaveric simulation, is that we used subjective, Kirkpatrick Level 1 [34] measures of impact. Objective, quantitative assessment of performance and outcome following training may provide more compelling evidence of impact, but it was not possible to do that in our study. The cohort of residents in our study were also self-selected and so may be a particularly motivated group, and hence, it is difficult to know how generalizable these results are to the orthopedic resident population as a whole. Only one-quarter of our eligible resident cohort (22 of approximately 80 residents) participated, and it is impossible to know if this represents a particular sector—perhaps those of particularly high or conversely low ability and confidence. There were only 2 residents in the midstage group, which limits the inferences that can be drawn from quantitative analysis. We chose to present 3 categories of resident seniority rather than combine groups to increase the generalizability of our results where training programs typically consider training to be in 3 phases. There was a skew toward younger and male residents in our study population, which may impact the generalizability of the results to other groups. We did not attempt to explore participants’ motivation for joining the study; it is possible that we attracted a particularly motivated cohort of participants, or the reverse may be true—individuals with low confidence in their skills may have been more likely to participate.

Conclusions

Freestyle DP cadaveric simulation allows training efficiency and educational impact to be maximized when there are inevitable resource constraints on repeated procedural attempts. The most senior residents reported greatest enthusiasm for the DP style of training, and this may be because of a greater awareness of their own learning needs and confidence in addressing them independently. All participants reported the course to be an extremely valuable training opportunity with a very high likelihood of skill transfer to the workplace and resultant patient benefit.

Acknowledgments

Funding for the training course was provided by Health Education England West Midlands.

Conflicts of Interest

None declared.

References

1. James H, Chapman A, Pattison G, Griffin DR, Fisher JD. Systematic review of the current status of cadaveric simulation for surgical training. Br J Surg 2019 Dec;106(13):1726-1734 [FREE Full text] [doi: 10.1002/bjs.11325] [Medline: 31573088]
2. Gilbody J, Prasthofer A, Ho K, Costa M. The use and effectiveness of cadaveric workshops in higher surgical training: a systematic review. Annals 2011 Jul;93(5):347-352. [doi: 10.1308/147870811x582954]
3. Reznick RK, MacRae H. Teaching surgical skills—changes in the wind. N Engl J Med 2006 Dec 21;355(25):2664-2669. [doi: 10.1056/NEJMra054785] [Medline: 17182991]
4. Benkhadra M, Faust A, Ladoire S, Trost O, Trouilloud P, Girard C, et al. Comparison of fresh and Thiel's embalmed cadavers according to the suitability for ultrasound-guided regional anesthesia of the cervical region. Surg Radiol Anat 2009 Aug 19;31(7):531-535. [doi: 10.1007/s00276-009-0477-z] [Medline: 19225711]
5. James HK, Pattison GT, Griffin DR, Fisher JD. How does cadaveric simulation influence learning in orthopedic residents? J Surg Educ 2020 May;77(3):671-682 [FREE Full text] [doi: 10.1016/j.jsurg.2019.12.006] [Medline: 32147464]
6. Marron C, Byrnes C, Kirk S. An EWTD-compliant shift rota decreases training opportunities. Bulletin 2005 Jul 01;87(7):246-248. [doi: 10.1308/147363505x46880]
7. Bates T, Cecil E, Greene I. The effect of the EWTD on training in general surgery: an analysis of electronic logbook records. Bulletin 2007 Mar 01;89(3):106-109. [doi: 10.1308/147363507x177045]
8. Temple J. Time for training: a review of the impact of the European Working Time Directive on the quality of training. SILO. 2010 May. URL: https://silo.tips/download/time-for-training-a-review-of-the-impact-of-the-european-working-time-directive [accessed 2022-05-04]
9. James HK, Fisher JD, Griffin DR, Pattison G. If we can't get to theatre, we can't learn to operate : a study of factors influencing core trainee access to the operating theatre in trauma and orthopaedics. Bulletin 2021 Jan;103(1):32-37. [doi: 10.1308/rcsbull.2021.10]
10. Improving surgical training: proposal for a pilot surgical training programme. Royal College of Surgeons of England. 2015 Oct. URL: https://www.rcseng.ac.uk/-/media/files/rcs/careers-in-surgery/improving_surgical_training_text.pdf [accessed 2022-05-05]
11. McCaskie AW, Kenny DT, Deshmukh S. How can surgical training benefit from theories of skilled motor development, musical skill acquisition and performance psychology? Med J Aust 2011 May 02;194(9):463-465. [doi: 10.5694/j.1326-5377.2011.tb03062.x] [Medline: 21534904]
12. Aggarwal R, Mytton OT, Derbrew M, Hananel D, Heydenburg M, Issenberg B, et al. Training and simulation for patient safety. Qual Saf Health Care 2010 Aug 06;19(Suppl 2):i34-i43 [FREE Full text] [doi: 10.1136/qshc.2009.038562] [Medline: 20693215]
13. Human tissue act 2004. Human Tissue Authority. URL: https://www.hta.gov.uk/guidance-professionals/hta-legislation/human-tissue-act-2004 [accessed 2020-02-12]
14. Blaschko SD, Brooks HM, Dhuy SM, Charest-Shell C, Clayman RV, McDougall EM. Coordinated multiple cadaver use for minimally invasive surgical training. JSLS 2007;11(4):403-407 [FREE Full text] [Medline: 18237501]
15. Hashimoto DA, Sirimanna P, Gomez ED, Beyer-Berjot L, Ericsson KA, Williams NN, et al. Deliberate practice enhances quality of laparoscopic surgical performance in a randomized controlled trial: from arrested development to expert performance. Surg Endosc 2015 Nov 25;29(11):3154-3162 [FREE Full text] [doi: 10.1007/s00464-014-4042-4] [Medline: 25539697]
16. Ericsson K. Deliberate practice and acquisition of expert performance: a general overview. Acad Emerg Med 2008 Nov;15(11):988-994 [FREE Full text] [doi: 10.1111/j.1553-2712.2008.00227.x] [Medline: 18778378]
17. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. Psychol Rev 1993;100(3):363-406. [doi: 10.1037/0033-295x.100.3.363]
18. Bjerke MB, Renger R. Being smart about writing SMART objectives. Eval Program Plann 2017 Apr;61:125-127. [doi: 10.1016/j.evalprogplan.2016.12.009] [Medline: 28056403]
19. Gunst M, O'Keeffe T, Hollett L, Hamill M, Gentilello LM, Frankel H, et al. Trauma operative skills in the era of nonoperative management: the trauma exposure course (TEC). J Trauma 2009 Nov;67(5):1091-1096. [doi: 10.1097/TA.0b013e3181bc77ba] [Medline: 19901673]
20. 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] [Medline: 24051881]
21. Dunnington GL. A model for teaching sentinel lymph node mapping and excision and axillary lymph node dissection. J Am Coll Surg 2003 Jul;197(1):119-121. [doi: 10.1016/S1072-7515(03)00231-X]

22. Sheckter CC, Kane JT, Minneti M, Garner W, Sullivan M, Talving P, et al. Incorporation of fresh tissue surgical simulation into plastic surgery education: maximizing extraclinical surgical experience. J Surg Educ 2013 Jul;70(4):466-474. [doi: 10.1016/j.jsurg.2013.02.008] [Medline: 23725934]
23. Reed AB, Crafton C, Giglia JS, Hutto JD. Back to basics: use of fresh cadavers in vascular surgery training. Surgery 2009 Oct;146(4):757-763. [doi: 10.1016/j.surg.2009.06.048] [Medline: 19789036]
24. Pham M, Kale A, Marquez Y, Winer J, Lee B, Harris B, et al. A perfusion-based human cadaveric model for management of carotid artery injury during endoscopic endonasal skull base surgery. J Neurol Surg B Skull Base 2014 Oct 2;75(5):309-313 [FREE Full text] [doi: 10.1055/s-0034-1372470] [Medline: 25301092]
25. Aydin A, Ahmed K, Khan M, Dasgupta P, McCabe J. The role of human cadaveric procedural simulation in urology training. J Urol 2015 Apr 1;193(4S):e273. [doi: 10.1016/j.juro.2015.02.1259]
26. Winer JL, Kramer DR, Robison RA, Ohiorhenuan I, Minneti M, Giannotta S, et al. Cerebrospinal fluid reconstitution via a perfusion-based cadaveric model: feasibility study demonstrating surgical simulation of neuroendoscopic procedures. JNS 2015 Nov;123(5):1316-1321. [doi: 10.3171/2014.10.jns1497]
27. Mahon N, Hanu-Cernat L, Thanabalan N, AlAli S, Virdi M. Effectiveness of cadaveric surgical simulation as perceived by participating OMFS trainees. Br J Oral Maxillofac Surg 2016 Dec;54(10):e126. [doi: 10.1016/j.bjoms.2016.11.172]
28. Weber EL, Leland HA, Azadgoli B, Minneti A, Carey JN. Preoperative surgical rehearsal using cadaveric fresh tissue surgical simulation increases resident operative confidence. Ann Transl Med 2017 Aug;5(15):302-302 [FREE Full text] [doi: 10.21037/atm.2017.06.28] [Medline: 28856142]
29. Chouari TAM, Lindsay K, Bradshaw E, Parson S, Watson L, Ahmed J, et al. An enhanced fresh cadaveric model for reconstructive microsurgery training. Eur J Plast Surg 2018 Apr 25;41(4):439-446 [FREE Full text] [doi: 10.1007/s00238-018-1414-3] [Medline: 30100677]
30. Prentice R. Drilling surgeons: the social lessons of embodied surgical learning. ST&HV 2016 Aug 18;32(5):534-553. [doi: 10.1177/0895904805303201]
31. Aggarwal R, Moorthy K, Darzi A. Laparoscopic skills training and assessment. Br J Surg 2004 Dec;91(12):1549-1558. [doi: 10.1002/bjs.4816] [Medline: 15547882]
32. Agha RA, Fowler AJ, Sevdalis N. The role of non-technical skills in surgery. Ann Med Surg (Lond) 2015 Dec;4(4):422-427 [FREE Full text] [doi: 10.1016/j.amsu.2015.10.006] [Medline: 26904193]
33. Cleland J, Walker KG, Gale M, Nicol LG. Simulation-based education: understanding the socio-cultural complexity of a surgical training 'boot camp'. Med Educ 2016 Aug 11;50(8):829-841. [doi: 10.1111/medu.13064] [Medline: 27402043]
34. James H. Measuring the educational impact of simulation training in trauma and orthopaedics. JTO 2019;7(3):54-56 [FREE Full text]

Abbreviations

DP: deliberate practice
FTR: flexor tendon repair
ORIF: open reduction internal fixation
PBA: procedure-based assessment
PGY: postgraduate year
SMART: specific, measurable, attainable, relevant, time-based
USO: ulnar shortening osteotomy

Edited by N Zary; submitted 09.11.21; peer-reviewed by A Stephens, R McKeown; comments to author 06.12.21; revised version received 08.12.21; accepted 15.12.21; published 29.06.22.

Please cite as:
James HK, Fawdington RA
Freestyle Deliberate Practice Cadaveric Hand Surgery Simulation Training for Orthopedic Residents: Cohort Study
JMIR Med Educ 2022;8(2):e34791
URL: https://mededu.jmir.org/2022/2/e34791
doi:10.2196/34791
PMID:35767315

Review

Mobile-Social Learning for Continuing Professional Development in Low- and Middle-Income Countries: Integrative Review

Dominique Guillaume[1,2], MSN, AGPCNP-BC, ACRN; Erica Troncoso[1], MSc; Brenice Duroseau[2], MSN, FNP-C, AAHIVS; Julia Bluestone[1], CNM; Judith Fullerton[1], CNM, PhD

[1]Jhpiego, Baltimore, MD, United States

[2]School of Nursing, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
Dominique Guillaume, MSN, AGPCNP-BC, ACRN
Jhpiego
1615 Thames St # 200
Baltimore, MD, 21231
United States
Phone: 1 410 537 1800
Email: dominique.guillaume@jhpiego.org

Abstract

Background: Access to continuing professional development (CPD) for health care workers in low- and middle-income countries (LMICs) is severely limited. Digital technology serves as a promising platform for supporting CPD for health care workers by providing educational content virtually and enabling virtual peer-to-peer and mentor interaction for enhanced learning. Digital strategies for CPD that foster virtual interaction can increase workforce retention and bolster the health workforce in LMICs.

Objective: The objective of this integrative review was to evaluate the evidence on which digital platforms were used to provide CPD to health care workers and clinical students in LMICs, which was complemented with virtual peer-to-peer or mentor interaction. We phrased this intersection of virtual learning and virtual interaction as *mobile-social learning*.

Methods: A comprehensive database and gray literature search was conducted to identify qualitative, quantitative, and mixed methods studies, along with empirical evidence, that used digital technology to provide CPD and virtual interaction with peers or mentors. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were followed. Eligible articles were written in English, conducted in an LMIC, and used a mobile device to provide CPD and facilitate virtual peer-to-peer or mentor interaction. Titles, abstracts, and full texts were screened, followed by an assessment of the quality of evidence and an appraisal of the articles. A content analysis was then used to deductively code the data into emerging themes.

Results: A total of 750 articles were identified, and 31 (4.1%) were included in the review. SMS text messaging and mobile instant messaging were the most common methods used to provide continuing education and virtual interaction between peers and mentors (25/31, 81%). Across the included articles, participants had high acceptability for using digital platforms for learning and interaction. Virtual peer interaction and mentorship were found to contribute to positive learning outcomes in most studies (27/31, 87%) through increased knowledge sharing, knowledge gains, improved clinical skills, and improved service delivery. Peer-to-peer and mentor interaction were found to improve social support and reduce feelings of isolation (9/31, 29%). There were several challenges in the implementation and use of digital technology for mobile-social learning, including limited access to resources (eg, internet coverage and stable electricity), flexibility in scheduling to participate in CPD, and sociobehavioral challenges among students.

Conclusions: The summary suggests that mobile-social learning is a useful modality for curriculum dissemination and skill training and that the interface of mobile and social learning serves as a catalyst for improved learning outcomes coupled with increased social capital.

(JMIR Med Educ 2022;8(2):e32614) doi:10.2196/32614

KEYWORDS

digital learning; continuing medical education; mHealth; peer learning; mentorship; health systems; global health; mobile phone

Introduction

Background

The shortage of health care providers in low- and middle-income countries (LMICs) places an insurmountable strain on health care systems. The World Health Organization estimates that nearly 57 countries lack an approximate 4.2 million health care providers (physicians, nurses, midwives, and allied health professionals) [1,2]. The strengthening of health care systems in LMICs requires multifaceted approaches to the training and retention of health care workers to meet clinical needs [3]. Continuing professional development (CPD) for health care providers is essential for the development and application of health care practices and policies necessary for health promotion, disease prevention and management, and fostering sustainable health systems. Although many high-income countries require health care workers to participate in regular CPD, many LMICs do not have such regulations or policies. Traditionally, in-person CPD training has been the primary method of providing health care education for health care providers in LMICs [4]. However, health care workers in low-resource settings, especially those in rural areas, face substantial logistical barriers to accessing in-person CPD programs (eg, cost of travel and inflexibility in scheduling). Thus, access to such programs is remarkably limited, especially when there is a lack of provider engagement [4]. Research has demonstrated that health care workers in LMICs are more likely to have higher motivation, satisfaction, and retention when they are provided with access to continuing education [3,5-7]. In countries with limited resources, addressing health care worker shortages and service needs requires tailored, cost-effective approaches for training, supervising, and mentoring health care workers. It is imperative for such approaches to minimize strain on already burdened health care systems while simultaneously providing instructional experiences that trainees need to successfully perform their jobs [8].

Mobile-Social Learning to Support CPD

Digital education strategies have gained momentum over the last decade in low-resource settings for the provision of CPD to health care workers. Digital education encompasses various modalities of learning, including but not limited to offline and web-based computer-based education, gamification, massive open web-based courses, virtual reality environments, augmented reality, virtual patient simulations, and mobile digital education [9]. Given the expansion of mobile phone use in LMICs, leveraging digital health strategies through the use of mobile technology has the potential to alleviate significant health system challenges [10,11]. CPD that is provided through mobile devices can support workplace-based practical training, reduce in-person instruction time, support social peer learning, and allow programs to reach a greater number of providers, especially those practicing in remote locations [4]. Virtual training is affordable and can be easily adapted as new information is discovered to update providers with the latest information and clinical developments. Furthermore, virtual CPD that is provided through mobile technology has demonstrated high feasibility and acceptability among health care providers in LMIC settings [4,12-14]. Mobile platforms can be used to not only provide CPD education but also complement CPD through peer-to-peer and mentorship engagement and interaction. Although there have been a number of studies that have evaluated the efficacy of digital and mobile health (mHealth) technology for providing CPD to health care workers, few studies have evaluated the potential of digital health, specifically mHealth, in providing peer interaction and mentorship for sustained education and training. The investigation of mobile-networked communication technology for health care provider CPD has been relatively understudied in low-income settings [15].

Jhpiego (a Johns Hopkins University affiliate) has supported CPD for health care providers and capacity building in multiple low-resource settings for nearly 50 years. In leveraging strategies to increase the accessibility of CPD programs, a heightened focus has been placed on using digital training approaches, including the provision of CPD through mobile phones [16-18]. Jhpiego recognizes that delivering lean, just-in-time learning via mobile devices can support workplace-based practical training, reduce in-person instruction time, support social peer learning, and allow programs to reach greater numbers of providers. As access to mobile technology continues to increase among health care providers in LMICs, so does access to platforms that foster virtual interaction and communication, such as mobile instant messaging services (eg, WhatsApp) and social media (eg, Facebook, Instagram, and Twitter). We believe that health workers should receive greater social support to improve retention rates, improve morale, and accelerate the potential of social learning. Such findings have been demonstrated in high-income countries, where social support and virtual interaction were found to foster understanding and learning among health care professionals and clinical students [19,20]. Social and mobile platforms can be used for learning and support broader, facility-based quality improvement efforts with scalable efficiency. For these purposes, we have combined the terms into a single phrase, *mobile-social*, to describe this key intersection that warrants further investment for achieving greater capacity-building impact. We are defining mobile-social learning as a new methodology that is powerful for supporting health care providers to improve their clinical capacity, learning, and performance. A mobile-social learning approach incorporates the following two aspects: the use of mobile technology to increase access to digital learning opportunities and social platforms that encourage the social aspect of learning by facilitating professional networks for the sharing of experiences and exchange of knowledge through virtual communication.

Through mobile-social learning, mobile distance education is provided to users and supplemented by web-based real-time discussions and a collaborative learning approach. This allows for opportunities for students to collaborate to construct knowledge while promoting the development of learning communities and supporting the learning process [21]. As part of Jhpiego's commitment to make learning available and convenient to health workers anytime, anywhere, whether in the workplace, on the road, or at home, we see fostering this modality for learning and supporting peer learning as a great investment. Given the dearth of literature that has assessed the efficacy and outcomes of mobile-social learning in LMICs, the purpose of this integrative review was to explore the potential of mobile-social learning to support capacity building and improve the quality of CPD for health care providers in LMICs.

Methods

Search Strategy

A comprehensive literature search was undertaken using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) recommendations to guide the search and review process. Peer-reviewed literature published between January 2016 and March 2021 was searched for on the PubMed, CINAHL, and Embase bibliographic databases. A gray literature search was also undertaken using Digital Square and the US Agency for International Development mHealth database. Hand searches of references from articles that were populated from the search were also conducted to identify relevant articles that may not have been identified using the search strategy. Searches were conducted separately on each database using controlled vocabulary supplemented with keywords and Medical Subject Headings terms combined with the Boolean operators *OR* and *AND*. The key terms and medical subject heading terms included concepts pertaining to continuing education, virtual training, health care providers, e-learning, mentorship, peer-to-peer interaction, and LMICs (Multimedia Appendix 1).

Eligibility Criteria

Studies were included if they were written in English, took place in an LMIC, and used mobile devices (eg, cellular phones, smartphones, tablets, palmtops, and pocket PCs) to provide continuing education and skill training to health care providers or preprofessional students. We included studies in which mobile platforms that fostered the concept of mobile-social learning were used either solely or in conjunction with a traditional face-to-face learning approach (eg, blended learning approach) [9]. Interventions using mobile-social learning were defined as any teaching, learning, or training intervention along with virtual interaction delivered using wireless networking, mobile telecommunication technology, multimedia messaging services, or SMS text messaging through a mobile device [9]. Virtual interaction among the eligible articles had to consist of either peer-to-peer interactions or mentor interactions that bolstered learning and professional support. Articles that did not evaluate student learning outcomes from virtual continuing education programs and that solely focused on program design and feasibility were excluded.

Peer-reviewed studies were included in our synthesis in addition to gray literature that was not peer-reviewed. Gray literature included relevant programmatic reports, case reports, research reports, presentations, and issue papers published by government entities, nongovernmental organizations, and private organizations. We included non–peer-reviewed gray literature given the possibility of limited published research on this topic coupled with the recent shift toward digital learning modalities concurrent with the COVID-19 pandemic. Thus, including gray literature allowed for an opportunity to highlight relevant work that may not have otherwise been identified while providing a balanced view of the evidence given the lack of peer-reviewed studies on this topic [22]. Furthermore, this method mitigated the risk of publication bias, which could potentially limit the availability of research in a field that is novel and rapidly growing [11,22].

Study Identification and Selection

The search yielded a total of 750 articles (PubMed: n=484, 64.5%; CINAHL: n=202, 26.9%; Embase: n=31, 4.1%; Digital Square: n=22, 2.9%; US Agency for International Development mHealth database: n=10, 1.3%) in addition to 2 articles that were identified from a hand search of previously published systematic reviews that assessed the use of digital health for educating health care providers. Articles were imported into Mendeley (Elsevier) and subsequently uploaded to the systematic review tool *Rayyan QCRI* (Rayyan Systems Inc). Duplicates were identified and excluded (104/750, 13.9%), thus resulting in a total of 646 articles that were screened. Authors DG, JB, and ET conducted title and abstract screening and full-text screening. After title and abstract screening of the 646 articles, a total of 602 (93.2%) were excluded, with 44 (6.8%) remaining for full-text screening. Disagreements among the articles were identified and resolved through consensus (DG, ET, and JB). Articles were then graded on their quality of evidence using the Johns Hopkins Nursing Evidence Level and Quality Guide [23]. This tool was selected as it has been extensively used in the literature and provides grading criteria for peer-reviewed studies along with experiential, nonresearch evidence (eg, case reports, quality improvement guidelines, and programmatic reports). The following factors were assessed in evaluating the studies and determining the quality of evidence: generalizability of the results, sample size, control, consistency of the results, methodology, limitations, conclusions, and recommendations [23]. After the articles were evaluated, the research team made final agreements for inclusion.

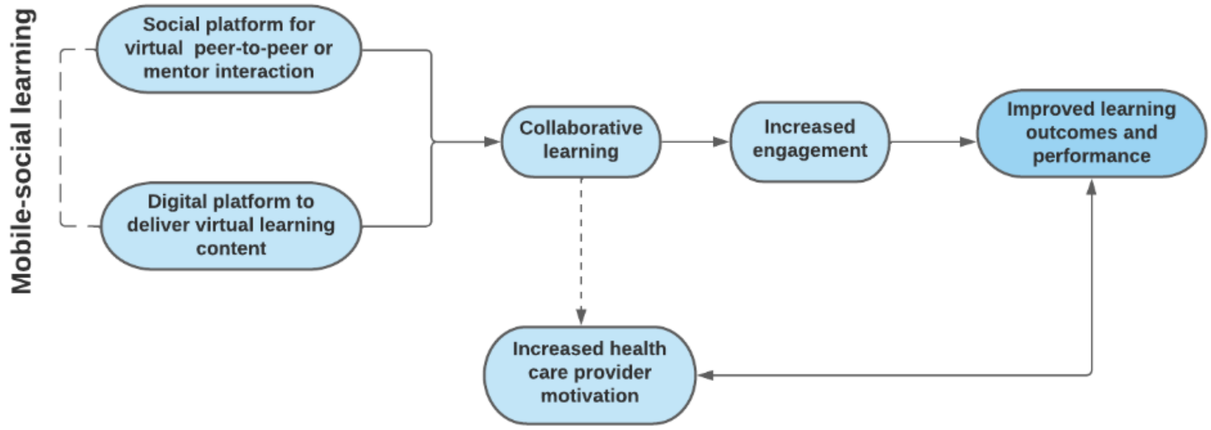
Data Extraction and Analysis

Once the articles were agreed upon for inclusion by the research team (DG, JB, and ET), data from the included articles were extracted (DG). The extracted data from the included studies consisted of the following categories: the country in which the study was conducted, study design and methods, study aims, participant characteristics, clinical focus, intervention and comparisons, outcome measures, key findings, and limitations. An integrated approach using data conversion was used, in which quantitative and qualitative data were used to address the research question. Quantitative data were transformed into qualitative themes as described in Tashakkori and Teddlie [24]. The qualitization of quantitative data involves converting quantitative data into qualitative themes in which data are theoretically grouped based on concepts measured from cross-sectional survey data [25-28]. Quantitative data were operationalized based on the concepts that were measured, and clusters of numeric data were transformed into qualitative themes. Quantitative and qualitative data were inductively coded line-by-line followed by the codes being categorized into broader themes [28,29]. The identified themes included the following: student perspectives on mobile-social learning, forms of interaction and communication within mobile-social platforms, learning outcomes, and challenges to and facilitators of mobile-social learning.

Conceptual Framework

Findings from the synthesis of the included studies were used to develop a conceptual framework depicting mobile-social learning (Figure 1). In this conceptual framework, virtual learning and virtual engagement with peers or mentors foster collaborative learning. Collaborative learning results in increased engagement followed by improved learning outcomes and performance. Collaborative learning can also increase health care provider motivation and lead to improved outcomes and performance. This relationship can also be reciprocal to improved learning outcomes, contributing to increased motivation among health care providers. We anticipate that this model will be used in providing CPD for clinical professionals, community health workers, and preclinical and clinical students on a variety of platforms in low-resource settings where there may be significant challenges in implementing and accessing face-to-face CPD.

Figure 1. Mobile-social learning conceptual framework.



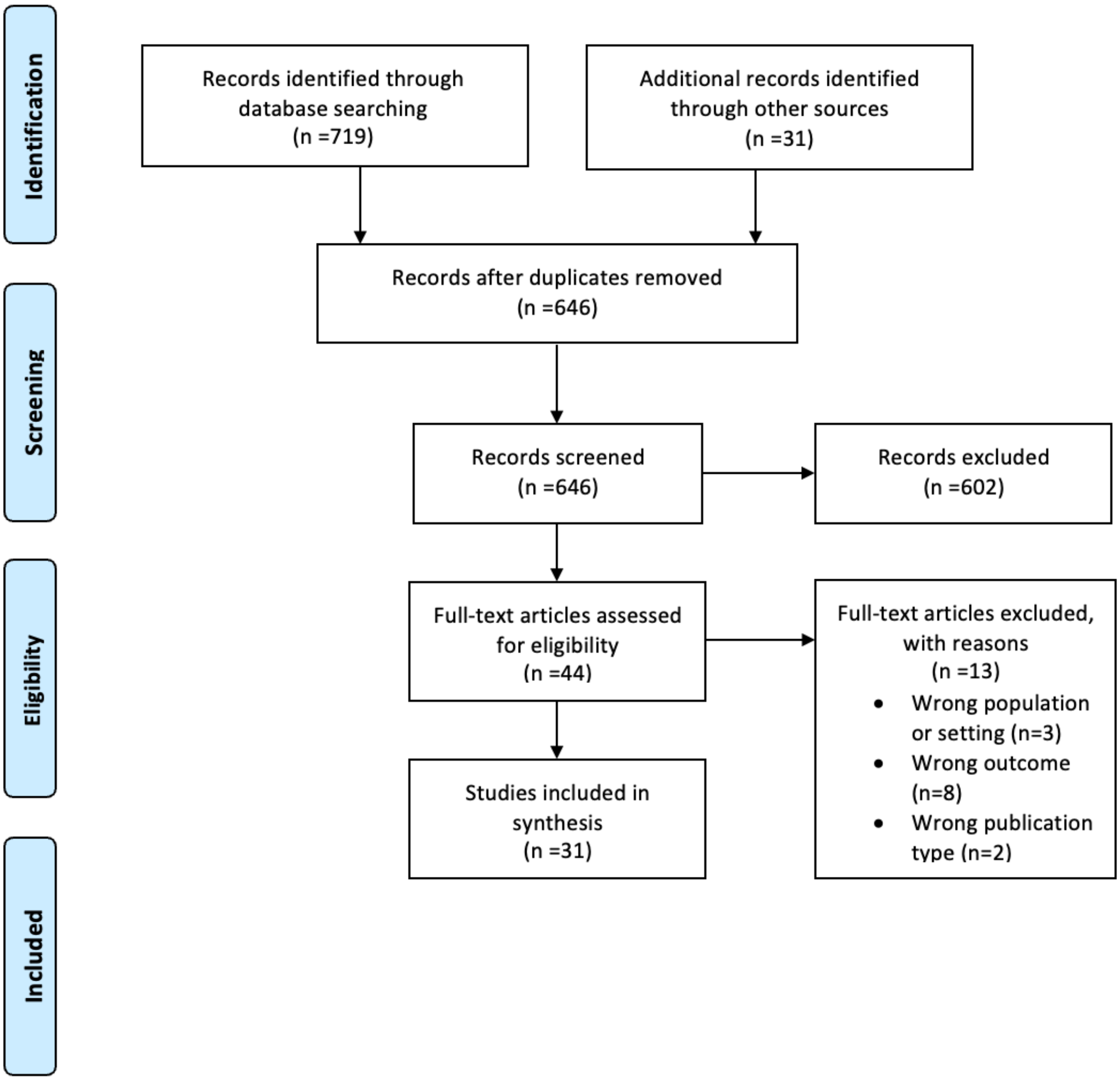
Results

Included Studies

A total of 31 articles were selected for inclusion (Figure 2). Among the included articles, sub-Saharan Africa (28/31, 90%) was the most represented geographical region [1,8,15-17,30-52]. Countries in Southeast Asia were represented in 10% (3/31) of the studies [45,53,54], and 3% (1/31) of the studies included participants across Africa, Asia, and South America [1]. A wide array of health care cadres was targeted across the studies, including nurses, physicians, midwives, community health workers, public health specialists, hospital administrators, and health officers. In addition, several studies (7/31, 23%) included preprofessional clinical trainees (eg, nursing and medical students) [32-34,37,50-52]. Of the 31 studies, 9 (29%) were qualitative in design and were conducted using descriptive thematic analysis, in-depth interviews, or focus groups [8,15,30,32,34,36,40,41,50]. A total of 39% (12/31) of the studies [1,16,31,33,35,37,45-48,51,54] were quantitative, with most quantitative studies (10/12, 83%) having observational cross-sectional designs. Of the 31 studies, 10 (32%) used both quantitative and qualitative data collection [17,38,39,42-44,49,52,53,55] (Multimedia Appendix 2 [1,8,15-17,29-54]).

There was diversity in the clinical topics that were represented among the continuing education interventions. Nearly half (14/31, 45%) of the interventions focused on clinical topics pertaining to sexual and reproductive health (ie, maternal health, basic emergency obstetric and newborn care, HIV and AIDS prevention and treatment, treatment of sexually transmitted infections, cervical cancer screening, and family planning) [1,8,16,17,30,31,40-42,47,48,52,54,55]. Additional clinical topics included primary care, general nursing practice and skills, research, malnutrition, anesthesia, pediatric hematology-oncology, integrated management of childhood illnesses, nephrology, and orthopedics (Multimedia Appendix 3 [1,8,15-17,29-54]).

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.



Interventions

Most articles (25/31, 81%) used texting, mobile instant messaging, or SMS text messaging to provide continuing education content either as a stand-alone intervention or in addition to other digital learning tools and blended learning formats (eg, coupling SMS text messaging with traditional face-to-face learning) [8,15-17,30-38,40,41,43,45,46,48-52,55,56]. WhatsApp was the most commonly used platform among the studies that used messaging services (17/31, 55%) [8,15,30-34,36,37,43,46,48,50-52,55,56]. Telephone calls were included in 16% (5/31) of the studies and were primarily used between students and mentors to discuss and reinforce learning content and for students to obtain feedback from mentors [16,41,45,46,53]. Web-based courses that facilitated web-based interactive discussions and commentary were used in 19% (6/31) of the studies [1,42-46]. Of these 6 studies, 3 (50%) coupled web-based courses with social media platforms such as Facebook, Twitter, and Google groups, which served as an adjunct to the course in facilitating learning and peer interaction and engagement [1,43,44]. A total of 3% (1/31) of the studies used Facebook as the sole platform in both providing educational content and facilitating discussions between peers and mentors [39] (Multimedia Appendix 3).

Acceptability

Across the studies that assessed student acceptability and student satisfaction with the mobile-social interventions (16/31, 52%), most participants strongly endorsed using mobile-social learning for continuing education and were highly satisfied with the platform [1,8,15,32,36-42,46,49,51,52,54,55]. Students regarded the learning platforms as easily accessible, informative, and user-friendly, with the social components improving communication and knowledge sharing and fostering real-time feedback. Students in a number of studies (12/16, 75%) specifically highlighted the ability to receive real-time feedback from peers or mentors as beneficial to clinical practice, knowledge gains, and team building [8,15,31,34-37,39,40,50,52,53].

Student Engagement

Articles that measured student engagement (16/31, 52%) within the mobile-social learning platforms reported high levels of use among students [1,15,31,33,36-39,43-45,48,50-52,55]. However, the association between levels of engagement and learning outcomes varied among the studies. Woods et al [31] noted that students who regularly followed the WhatsApp learning group had a clinically significant increase in the odds of having higher confidence in managing their patients (odds ratio 8.44, 95% CI 2.33-35.23; $P<.05$). Abiodun et al [37] noted that, although students were highly engaged within WhatsApp groups, there were no significant associations between the frequency of reading messages and social-professional outcome measures. Among the studies that coupled web-based learning modules with social media interaction and discussions through Facebook and WhatsApp, module completion rates were significantly higher compared with standard massive open web-based course completion [1,30,43,44].

Interaction With Peers

High levels of peer engagement and interaction were reported among the included studies (19/31, 61%) [1,15,30,32,34,38-40,42-44,46,49-52,54,57]. A dominant theme that arose among the studies that assessed peer interaction was the ability to engage in active knowledge sharing. Interventions that fostered peer engagement and interaction through messaging and discussion forums found that participants viewed the interactive component and commentary as beneficial [1,15,30,32,34,38-40,42-44,46,49-52,54,57]. CPD programs that were augmented with social media found that peer support was facilitated through social media platforms, which allowed for real-time interactions and peer feedback and greater understanding of course material [1,30,43,44]. The inclusion of a peer support network was suggested by students in a study that solely focused on virtual mentorship for cervical cancer screening among nurses and did not include a virtual peer-to-peer interaction component [40]. Students in this study stated that, although they were actively engaged with mentors virtually, they strongly believed that including a peer support network or chat room to actively discuss clinical cases and medical imaging in real time would be beneficial [40].

Interaction With Mentors

Mentorship occurred at varying degrees (eg, pairing a student with a master mentor, assigning groups of students to a mentor, and facilitators providing feedback to students in discussion groups). Among the studies that included virtual mentorship (24/31, 77%), students reported positive interactions with mentors [1,8,15-17,34,35,38-41,43-54]. Mentorship occurred through phone calls, texting, and video calls in which mentors were able to provide remote support, real-time feedback, and guidance to students along with reinforcement of key learning messages and skills gained from educational modules and training sessions. In studies where a blended training approach was used, students emphasized the importance of having virtual access to an expert for further questions and case discussions after the initial on-site training [40,45]. Asiedu et al [37] noted that, despite receiving positive feedback on the mobile mentoring component from students, several students and mentors voiced concerns that mobile mentoring on its own was insufficient for posttraining follow-up and support. A few mentors noted that it was difficult to sustain the process of repeated telephone calls to students and that certain students were not being honest about progress [37]. The intervention implemented by Feldacker et al [42] did not contain a virtual mentorship component; however, students highly suggested the inclusion of virtual mentoring opportunities to supplement learning. Similarly, an intervention in Kenya that used WhatsApp mobile messaging for family planning learning did not include a mentorship component; however, students voiced the need for more mentorship support [55].

Learning Outcomes

Students who used mobile-social platforms reported positive learning outcomes in most of the included articles (27/31, 87%) [1,8,15,17,30-39,41-48,50,52,54,55,58]. Learning outcomes included knowledge gains, improved clinical skills, positive influences on clinical practice, and improved quality of service delivery. Students emphasized the educational benefits of having live case discussions with peers or mentors via the mobile-social learning platforms, in which patient-related questions could be addressed in real time along with consulting with peers or mentors about complex clinical cases [8,15, 30,36,38,46,50,52,54]. Across the studies that combined face-to-face learning with mobile-social learning (8/31, 26%), digital platforms that supported mobile-social learning were found to be instrumental in bridging the periods between face-to-face meetings through the ability of the participants to engage in continuous communication and feedback [52]. In 13% (4/31) of the studies where mobile messaging through peer groups was used, participants stated that previous case discussions with peers were used as a resource to which they referred when presented with a complicated clinical case in their practice. In addition, students cited using old case discussions saved in group chats for self-study [31,34,38,41].

Of the 31 studies, 3 (10%) [16,17,49] compared mobile-social learning platforms with traditional face-to-face interventions, but these studies yielded varied results regarding learning outcomes. Yigzaw et al [16] and Muhe et al [49] found no significant differences between groups in the gains in knowledge scores, whereas Ugwa et al [17] reported that the virtual learning arm, which included mobile mentoring, led to better skill performance at all assessment points compared with the traditional arm, with the virtual arm performing better in all competencies at 3 and 12 months after training.

Social Capital and Professional Integration

Social capital characterizes the relationships and interactions between members of a social group [59]. Social capital encompasses a culture of trust and tolerance in which extensive networks of voluntary associations emerge that facilitate coordination and cooperation for mutual benefit [60,61]. In medicine, social capital has been tied to the realization of lifelong learning opportunities, with digital engagement contributing to the development and maintenance of social capital [51,52]. Social capital typologies among the included articles included emotional support, the formation of deeper social connections with peers and colleagues, networking, and non–work-related communication [8,15,30,33,37,39,42,51,52]. Students in the studies that measured social capital (9/31, 29%) cited that virtual interaction with peers or mentors reduced feelings of isolation. Participants stated that interactions helped them maintain existing relationships while also developing and strengthening new social ties, thus promoting professional integration and improving team dynamics [33,34,37,51]. Pimmer et al [33] and Abiodun et al [37] noted that participants with higher levels of active engagement also felt less isolated professionally and had higher levels of social capital.

Challenges and Success Factors

Although positive feedback was elicited from students regarding mobile-social learning, there were several challenges in the implementation and use of the learning modality. Resource challenges related to power cuts and maintenance of uninterrupted internet access were primary barriers to the use of the mobile-social learning platforms in several studies (14/31, 45%) [8,15,31,32,38,40,42-46,50,52,54]. Ugwa et al [17] cited additional resource challenges pertaining to the unavailability of equipment for students to translate what they learned on the web into practice. The lack of time to participate in learning programs because of conflicts with clinical schedules was a challenge noted in 13% (4/31) of the studies [8,17,42,44]. However, 19% (6/31) of the studies highlighted that mobile-social programs allowed for increased flexibility of learning that could easily fit into students' schedules and reduce the need for travel to access face-to-face continuing education programs, which could be challenging for health care providers located in rural areas [1,32,42,43,49,54]. The costs of data bundles and messaging services to support the programs were cited as barriers in 10% (3/31) of the studies [41,49,52]. However, a few studies (6/31, 19%), including those that mentioned cost as a barrier, stated that, over the long term, distance learning approaches were more cost-effective compared with traditional face-to-face programs [8,16,41,43,48,49]. Sociobehavioral challenges were also noted, such as certain users engaging in inappropriate non–work-related discussions on the learning platforms, low levels of active participation, and low digital literacy [15,38,45,50]. In addition, hesitancy among students to interact on platforms because of fear, embarrassment, or lack of knowledge or awareness was cited as a barrier [38,40]. Success factors that were noted among the interventions included tailor-made content to meet the needs of the participants, the flexibility of being able to access learning content anytime and anywhere, and features that promoted ongoing engagement and personal interactivity with peers and mentors [1,8,15,17,30,32-35,37-39,42-47,49,51,52,54]. Peer and mentor interaction allowed for direct personalized feedback, the application of educational content from theory to clinical practice by the students, self-direction for learning, and active learning. Interaction with peers and mentors heightened confidence and promoted empowerment among students, thus leading to increased motivation to maintain newly acquired knowledge and skills [31,40,41].

Discussion

Principal Findings

The use of a wide variety of mobile platforms has become a common adjunct to traditional classroom-based pedagogy and, in many cases, is the default strategy for curriculum dissemination in many countries. However, even in the era of increased uptake of mobile platforms for learning, multifaceted approaches are needed for the diversity of learning needs and preferences [42,62-64]. This integrative review demonstrates that mobile platforms that foster mobile-social learning can serve as an innovative method for providing health care workers and preprofessional students with skill training and education along with virtual interactive components. A major advantage of these strategies is the ability to reach a widely dispersed learner population residing in diverse geographic locations [15,30,40,54]. Access to learning can be independent of time specificity and open to the availability of repetitive exposure to content in support of mastery learning. This evidence synthesis demonstrates the added value of social learning networks as interventions that enhance the utility and effectiveness of internet-based learning platforms whether as independent dissemination strategies or as part of blended learning (hybrid) teaching and learning approaches. Social networks serve as the interface between teachers, mentors, independent learners, and the network of learners engaged in the exploration of any single topic [65,66]. Social networks through mobile platforms can be used to facilitate mobile-social learning as a means of deeper exploration of the understanding of content through peer engagement and interactive discussion and the giving and receiving of feedback [66,67]. In addition, this review highlighted the impact of mobile-social learning on reducing feelings of professional isolation and increasing social capital among health care workers. Increased social connectedness and social presence can play a key role in team building and increasing health care worker motivation and confidence [68]. Improved health care worker motivation may serve as a critical component in reducing attrition in LMICs, thus improving workforce capacity within health systems.

The studies included in this review offered evidence for the acceptability of the use of mobile platforms as an instructional modality, viewing them as accessible and user-friendly. Learners recognized and valued the addition of social interaction strategies via these mobile platforms as measures to receive affirmation of learning. This reinforcement offered by peers and mentors affected the confidence of students when translating learning into clinical practice. The interface of mobile learning with social engagement in the learning process offered the opportunity for meaningful interaction with a wider network of learners and health workers. The ability to engage in real-time discussion of clinical topics and resolution of clinical challenges diminished feelings of isolation, extended the professional network for collaboration when the next need might arise, and expanded the opportunity to remain current with emerging evidence underpinning clinical practice, thus affecting the quality of service delivery.

The evidence synthesis also highlighted the challenges that are inherent in the use of mobile-social learning platforms—infrastructure challenges that are often more complex in lower-resource settings. These include the tangible costs associated with hardware, software, and network connections. The intangible costs of the *learning curve* that must be faced by each participant must also be considered in the implementation of mobile-social learning programs as some platforms are inherently more complex and less user-friendly than others.

Implications for Future Research and Practice

Several key areas of digital health research are needed to support the efficacy of mobile-social learning. The use of digital platforms is well-suited to measuring the utility of the approach given that the platforms allow for the tracking of time and frequency of use by individual learners, including the degree to which these learners take advantage of the opportunities for social interaction as an enhancement of learning. These data are critical for the evaluation of the effectiveness of the mobile-social teaching and learning strategy. In addition, although mobile-social learning offers various advantages and benefits, certain studies in our review (3/31, 10%) did not yield significant learning outcomes among participants, with participants in some studies citing that they preferred face-to-face learning [16,17,49]. Thus, this underscores the importance of providing multifaceted options in CPD to meet the diversity of learning needs and preferences. Although most studies that included virtual mentorship noted positive results and favorability among participants, the health care infrastructure and human resource constraints in many LMICs may limit access and availability of mentors [69]. Thus, further research is needed to assess how mentorship can be provided without being an additional burden on health care systems and personnel. More research is also needed on developing mobile-social learning programs designed to fit local contexts. Future studies should evaluate the use of recent technological innovations such as augmented reality and virtual reality for mobile-social learning. The included studies did not evaluate the use of these platforms; however, these interventions are gaining traction for the training of health care professionals and preclinical students in high-income countries [70]. As technology advances at a rapid pace, there is a need to explore how such technological innovations can be accessible for health care providers and students in lower-resource settings. Although none of our included studies applied a gender lens to the observed outcomes, it is generally acknowledged as a moderating influence in health and education and, thus, should be considered in the implementation of programs [64,71]. More evidence generated through studies that are methodologically rigorous while simultaneously allowing for lean, iterative, and rapid-paced development and evaluation is needed to thoroughly assess the benefits of mobile-social learning in comparison with traditional learning modalities [72].

Central to the success of digital health interventions is the knowledge of health informatics challenges that may be experienced, particularly in environments where instability in the digital and health infrastructure is common [73]. Developing interventions that are designed to meet learning needs and preferences will entail more representation of individuals from LMICs in the technology development sector to inform the development of digital tools that fit local country contexts [74,75]. Although the included studies discussed challenges in internet access as a barrier to accessing mobile-social learning programs, there was limited discussion on the digital environment in which these studies were implemented. When developing, implementing, and expanding the use of digital health programs that focus on mobile-social learning in LMICs, it is essential to foster strong digital health ecosystems by building and promoting partnerships between the relevant public and private sectors [10,76,77]. Communication channels such as WhatsApp are increasingly being used as a simple, low-cost, and effective means of learning and communication within the clinical health sector. However, more attention must be paid to confidentiality, consent, and data security if individual client data are being shared through these channels [78]. The roles and responsibilities of medical professionals when using digital platforms for mobile-social learning must be outlined along with the development of guidelines and protocols to facilitate the integration of mobile-social learning within digital and health care infrastructures [78]. Thus, the standardization of policies in the exchange and use of information between systems will be critical in ensuring the usability and sustainability of mobile-social learning programs [73,78,79].

Limitations

This study is not without limitations. Although we used a comprehensive search strategy, we cannot guarantee that it identified all relevant studies. However, the incorporation of gray literature in our review reduced publication bias and denoted experiential evidence that supported mobile-social learning. Our search yielded studies that were predominantly focused in sub-Saharan Africa (28/31, 90%), with only a few studies (3/31, 10%) being conducted outside of that global region. Thus, the findings we report cannot be generalized to all LMIC settings given that sociocultural contexts, subjective norms, health system contexts, and digital environments may vary in different country settings. Finally, we only included articles that were published in English; therefore, we could have missed relevant articles published in other languages.

Conclusions

Mobile-social learning is a particularly useful modality for curriculum dissemination and skill training, and the interface of mobile and social learning offers an interaction effect that can serve as a catalyst for improved learning outcomes coupled with increased social capital. The mobile-social approach is by its nature conducive to the dissemination of shorter segments of key content packed in social platform formats that allow for peer and mentor engagement and interactivity. The concurrent enhancement of mobile curriculum dissemination apps, embedding of proven social interaction strategies into those apps, and development of more and newer user-friendly digital learning opportunities will lead to greater opportunities for learning and peer and mentor support in the interest of improving the quality of health services. As more countries turn to digital modalities of learning, it will be imperative for programs to be adapted for both the technological ecosystem and the local and national health care systems.

Acknowledgments

This study was supported by Jhpiego, a Johns Hopkins University affiliate.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 14 KB - mededu_v8i2e32614_app1.docx]

Multimedia Appendix 2
Design and level of evidence of included studies.
[DOCX File , 27 KB - mededu_v8i2e32614_app2.docx]

Multimedia Appendix 3
Summary of included articles.
[DOCX File , 44 KB - mededu_v8i2e32614_app3.docx]

References

1. Abawi K, Chandra-Mouli V, Toskin I, Festin MP, Gertiser L, Idris R, et al. E-learning for research capacity strengthening in sexual and reproductive health: the experience of the Geneva Foundation for Medical Education and Research and the Department of Reproductive Health and Research, World Health Organization. Hum Resour Health 2016 Dec 07;14(1):76 [FREE Full text] [doi: 10.1186/s12960-016-0173-0] [Medline: 27927220]
2. Global strategy on human resources for health: Workforce 2030. World Health Organization. 2020 Jul 2. URL: https://apps.who.int/iris/bitstream/handle/10665/250368/9789241511131-eng.pdf [accessed 2022-05-17]
3. Nádas M, Bedenbaugh R, Morse M, McMahon GT, Curry CL. A needs and resource assessment of continuing medical education in Haiti. Ann Glob Health 2015;81(2):248-254 [FREE Full text] [doi: 10.1016/j.aogh.2015.03.003] [Medline: 26088090]
4. Stark CM, Garner CD, Garg A, Bégin F. Building capacity of health professionals in low- and middle-income countries through online continuing professional development in nutrition. J Contin Educ Health Prof 2021 Jan 01;41(1):63-69 [FREE Full text] [doi: 10.1097/CEH.0000000000000334] [Medline: 33560042]
5. Willis-Shattuck M, Bidwell P, Thomas S, Wyness L, Blaauw D, Ditlopo P. Motivation and retention of health workers in developing countries: a systematic review. BMC Health Serv Res 2008 Dec 04;8:247 [FREE Full text] [doi: 10.1186/1472-6963-8-247] [Medline: 19055827]
6. Feldacker C, Pintye J, Jacob S, Chung MH, Middleton L, Iliffe J, et al. Continuing professional development for medical, nursing, and midwifery cadres in Malawi, Tanzania and South Africa: a qualitative evaluation. PLoS One 2017 Oct 17;12(10):e0186074 [FREE Full text] [doi: 10.1371/journal.pone.0186074] [Medline: 29040303]
7. Mbemba GI, Gagnon MP, Hamelin-Brabant L. Factors influencing recruitment and retention of healthcare workers in rural and remote areas in developed and developing countries: an overview. J Public Health Afr 2016 Dec 31;7(2):565 [FREE Full text] [doi: 10.4081/jphia.2016.565] [Medline: 28299160]
8. Bertman V, Petracca F, Makunike-Chikwinya B, Jonga A, Dupwa B, Jenami N, et al. Health worker text messaging for blended learning, peer support, and mentoring in pediatric and adolescent HIV/AIDS care: a case study in Zimbabwe. Hum Resour Health 2019 Jun 07;17(1):41 [FREE Full text] [doi: 10.1186/s12960-019-0364-6] [Medline: 31174543]
9. Dunleavy G, Nikolaou CK, Nifakos S, Atun R, Law GC, Tudor Car L. Mobile digital education for health professions: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Feb 12;21(2):e12937 [FREE Full text] [doi: 10.2196/12937] [Medline: 30747711]
10. Labrique A, Vasudevan L, Weiss W, Wilson K. Establishing standards to evaluate the impact of integrating digital health into health systems. Glob Health Sci Pract 2018 Oct 10;6(Suppl 1):S5-17 [FREE Full text] [doi: 10.9745/GHSP-D-18-00230] [Medline: 30305335]
11. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug 6;1(2):160-171 [FREE Full text] [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]
12. Long LA, Pariyo G, Kallander K. Digital technologies for health workforce development in low- and middle-income countries: a scoping review. Glob Health Sci Pract 2018 Oct 10;6(Suppl 1):S41-S48 [FREE Full text] [doi: 10.9745/GHSP-D-18-00167] [Medline: 30305338]

13. Mungo C, Osongo CO, Ambaka J, Randa MA, Samba B, Ochieng CA, et al. Feasibility and acceptability of smartphone-based cervical cancer screening among HIV-positive women in Western Kenya. JCO Glob Oncol 2021 May;7:686-693 [FREE Full text] [doi: 10.1200/GO.21.00013] [Medline: 33999653]
14. Mwaikambo L, Avila M, Mazursky S, Nallathambi K. Utilizing eLearning to strengthen the capacity of global health practitioners and institutions around the world. Knowl Manag E-Learn 2012 Sep;4(3):293-309. [doi: 10.34105/j.kmel.2012.04.024]
15. Pimmer C, Mhango S, Mzumara A, Mbvundula F. Mobile instant messaging for rural community health workers: a case from Malawi. Glob Health Action 2017;10(1):1368236 [FREE Full text] [doi: 10.1080/16549716.2017.1368236] [Medline: 28914165]
16. Yigzaw M, Tebekaw Y, Kim YM, Kols A, Ayalew F, Eyassu G. Comparing the effectiveness of a blended learning approach with a conventional learning approach for basic emergency obstetric and newborn care training in Ethiopia. Midwifery 2019 Nov;78:42-49. [doi: 10.1016/j.midw.2019.07.014] [Medline: 31349183]
17. Ugwa E, Kabue M, Otolorin E, Yenokyan G, Oniyire A, Orji B, et al. Simulation-based low-dose, high-frequency plus mobile mentoring versus traditional group-based trainings among health workers on day of birth care in Nigeria; a cluster randomized controlled trial. BMC Health Serv Res 2020 Jun 26;20(1):586 [FREE Full text] [doi: 10.1186/s12913-020-05450-9] [Medline: 32590979]
18. Yeboah A, Ouedraogo V, Faso JB, Manji A, Tanzania J, Siamwanza N, et al. Designing digital learning for a continuum of learning from pre-service to in-service. Global Digital Health Forum. 2018 Dec 11. URL: https://lib.digitalsquare.io/handle/123456789/77112 [accessed 2022-05-17]
19. Raiman L, Antbring R, Mahmood A. WhatsApp messenger as a tool to supplement medical education for medical students on clinical attachment. BMC Med Educ 2017 Jan 06;17(1):7 [FREE Full text] [doi: 10.1186/s12909-017-0855-x] [Medline: 28061777]
20. Johnston MJ, King D, Arora S, Behar N, Athanasiou T, Sevdalis N, et al. Smartphones let surgeons know WhatsApp: an analysis of communication in emergency surgical teams. Am J Surg 2015 Jan;209(1):45-51. [doi: 10.1016/j.amjsurg.2014.08.030] [Medline: 25454952]
21. Amry AB. The impact of WhatsApp mobile social learning on the achievement and attitudes of female students compared with face to face learning in the classroom. Eur Sci J 2014;22(10) [FREE Full text] [doi: 10.19044/esj.2014.v10n22p%p]
22. Paez A. Gray literature: an important resource in systematic reviews. J Evid Based Med 2017 Aug;10(3):233-240. [doi: 10.1111/jebm.12266] [Medline: 28857505]
23. Johns Hopkins Nursing Evidence-Based Practice Appendix D: Evidence Level and Quality Guide Evidence Levels Quality Guides Level I Experimental study, randomized controlled trial (RCT) Systematic review of RCTs, with or without meta-analysis. The Johns Hopkins Hospital. 2017. URL: https://www.mghpcs.org/EED/EBP/Assets/documents/pdf/2017_Appendix [accessed 2022-05-17]
24. Tashakkori A, Teddlie C. Mixed Methodology: Combining Qualitative and Quantitative Approaches. Thousand Oaks, CA, USA: Sage Publications; 1998.
25. Nzabonimpa JP. Quantitizing and qualitizing (im-)possibilities in mixed methods research. Method Innov 2018 Jul 27;11(2):205979911878902. [doi: 10.1177/2059799118789021]
26. Creamer EG. An Introduction to Fully Integrated Mixed Methods Research. Thousand Oaks, CA, USA: Sage Publications; May 2017.
27. Lizarondo L, Stern C, Carrier J, Godfrey C, Rieger K, Salmond S, et al. 8.2 Concepts and considerations for mixed methods systematic reviews. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Adelaide, Australia: Joanna Briggs Institute; 2020.
28. Tashakkori A, Johnson RB, Teddlie C. Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences. Thousand Oaks, CA, USA: Sage Publications; Oct 2020.
29. Saldana J. An introduction to codes and coding. In: Saldana J, editor. The Coding Manual for Qualitative Researchers. 3rd ed. London, UK: Sage Publications; 2009:3-12.
30. Kabinga-Makukula M, Lyambai K, Wahila R, Mwape L. Use of instant messaging to enhance leadership and management training for rural nurse managers. Nurs Manag (Harrow) 2019 Dec 02;26(6):22-27. [doi: 10.7748/nm.2019.e1834] [Medline: 31612665]
31. Woods J, Moorhouse M, Knight L. A descriptive analysis of the role of a WhatsApp clinical discussion group as a forum for continuing medical education in the management of complicated HIV and TB clinical cases in a group of doctors in the Eastern Cape, South Africa. South Afr J HIV Med 2019 Aug 1;20(1):982 [FREE Full text] [doi: 10.4102/sajhivmed.v20i1.982] [Medline: 31534790]
32. Willemse JJ, Jooste K, Bozalek V. Experiences of undergraduate nursing students on an authentic mobile learning enactment at a higher education institution in South Africa. Nurse Educ Today 2019 Mar;74:69-75. [doi: 10.1016/j.nedt.2018.11.021] [Medline: 30594902]
33. Pimmer C, Brühlmann F, Odetola TD, Oluwasola DO, Dipeolu O, Ajuwon AJ. Facilitating professional mobile learning communities with instant messaging. Comput Educ 2019 Jan;128:102-112. [doi: 10.1016/j.compedu.2018.09.005]

34. Pimmer C, Abiodun R, Daniels F, Chipps J. "I felt a sense of belonging somewhere". Supporting graduates' job transitions with WhatsApp groups. Nurse Educ Today 2019 Oct;81:57-63. [doi: 10.1016/j.nedt.2019.06.010] [Medline: 31330403]
35. Biemba G, Chiluba B, Yeboah-Antwi K, Silavwe V, Lunze K, Mwale RK, et al. Impact of mobile health-enhanced supportive supervision and supply chain management on appropriate integrated community case management of malaria, diarrhoea, and pneumonia in children 2-59 months: a cluster randomised trial in Eastern Province, Zambia. J Glob Health 2020 Jun;10(1):010425 [FREE Full text] [doi: 10.7189/jogh.10.010425] [Medline: 32509293]
36. Makwabe EF, Pereira-Kamath N, Vincent L, Mngumi J, Kelly H, Pichan F, et al. SAT-489 Audit of a WhatsApp based continuing education platform to facilitate on the job formal learning to improve quality of care delivered among haemodialysis nurses. Kid Int Rep 2020 Mar 1;5(3):S204. [doi: 10.1016/j.ekir.2020.02.521]
37. Abiodun R, Daniels F, Pimmer DC, Chipps J. A whatsapp community of practice to support new graduate nurses in South Africa. Nurse Educ Pract 2020 Jul;46:102826. [doi: 10.1016/j.nepr.2020.102826] [Medline: 32778376]
38. Peponis C, Khaliq M, Ismail Ali A, Bose D, Wicks L, Tessema G. An international instant-messaging journal club: a modern, fun and global approach to a traditional teaching tool. Trop Doct 2020 Jan;50(1):49-53. [doi: 10.1177/0049475519876856] [Medline: 31547781]
39. Pimmer C, Chipps J, Brysiewicz P, Walters F, Linxen S, Gröhbiel UG. Supervision on social media: use and perception of Facebook as a research education tool in disadvantaged areas. Int Rev Res Open Distribut Learn 2016 Sep 26;17(5):200-214. [doi: 10.19173/irrodl.v17i5.2547]
40. Asgary R, Cole H, Adongo P, Nwameme A, Maya E, Adu-Amankwah A, et al. Acceptability and implementation challenges of smartphone-based training of community health nurses for visual inspection with acetic acid in Ghana: mHealth and cervical cancer screening. BMJ Open 2019 Jul 16;9(7):e030528 [FREE Full text] [doi: 10.1136/bmjopen-2019-030528] [Medline: 31315879]
41. Asiedu A, Nelson AR, Gomez PP, Tappis H, Effah F, Allen C. "It builds your confidence… you've done well": healthcare workers' experiences of participating in a low-dose, high-frequency training to improve newborn survival on the day of birth in Ghana. Gates Open Res 2019 May 22;3:1470 [FREE Full text] [doi: 10.12688/gatesopenres.12936.1] [Medline: 31410394]
42. Feldacker C, Jacob S, Chung MH, Nartker A, Kim HN. Experiences and perceptions of online continuing professional development among clinicians in sub-Saharan Africa. Hum Resour Health 2017 Dec 29;15(1):89 [FREE Full text] [doi: 10.1186/s12960-017-0266-4] [Medline: 29284492]
43. Hoedebecke K, Mahmoud M, Yakubu K, Kendir C, D'Addosio R, Maria B, et al. Collaborative global health E-learning: a Massive Open Online Course experience of young family doctors. J Family Med Prim Care 2018;7(5):884-887 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_186_18] [Medline: 30598927]
44. Scott KW, Dushime T, Rusanganwa V, Woskie L, Attebery C, Binagwaho A. Leveraging massive open online courses to expand quality of healthcare education to health practitioners in Rwanda. BMJ Open Qual 2019 Nov 6;8(4):e000532 [FREE Full text] [doi: 10.1136/bmjoq-2018-000532] [Medline: 31799443]
45. Shah S, Knoble S, Ross O, Pickering S. A distance blended learning program to upgrade the clinical competence of district non-doctor anesthesia providers in Nepal. World J Surg 2017 Dec;41(12):3006-3011. [doi: 10.1007/s00268-017-4273-3] [Medline: 29038830]
46. Hockenberry M, Mulemba T, Nedege A, Madumetse K, Higgins J. Distance-based education for nurses caring for children with cancer in Sub-Saharan Africa. J Pediatr Oncol Nurs 2020;37(5):321-329. [doi: 10.1177/1043454220938355] [Medline: 32659198]
47. Asgary R, Staderini N, Mthethwa-Hleta S, Lopez Saavedra PA, Garcia Abrego L, Rusch B, et al. Evaluating smartphone strategies for reliability, reproducibility, and quality of VIA for cervical cancer screening in the Shiselweni region of Eswatini: a cohort study. PLoS Med 2020 Nov 19;17(11):e1003378 [FREE Full text] [doi: 10.1371/journal.pmed.1003378] [Medline: 33211691]
48. Allen M. It Takes a Village: Midwives use WhatsApp for peer support. Digital Square. 2020. URL: https://lib.digitalsquare.io/handle/123456789/77068 [accessed 2022-05-17]
49. Muhe LM, Iriya N, Bundala F, Azayo M, Bakari MJ, Hussein A, et al. Evaluation of distance learning IMCI training program: the case of Tanzania. BMC Health Serv Res 2018 Jul 13;18(1):547 [FREE Full text] [doi: 10.1186/s12913-018-3336-y] [Medline: 30001709]
50. Ajuwon A, Pimmer C, Odetola T, Gröhbiel U, Oluwasola O, Olaleye O. Mobile Instant Messaging (MIM) to support teaching practice: insights from a nurse tutor program in Nigeria. Malawi Med J 2018 Jun;30(2):120-126 [FREE Full text] [doi: 10.4314/mmj.v30i2.12] [Medline: 30627340]
51. Pimmer C, Brühlmann F, Odetola TD, Dipeolu O, Gröhbiel U, Ajuwon AJ. Instant messaging and nursing students' clinical learning experience. Nurse Educ Today 2018 May;64:119-124. [doi: 10.1016/j.nedt.2018.01.034] [Medline: 29475195]
52. Pimmer C, Lee A, Mwaikambo L. Mobile instant messaging: new knowledge tools in global health? Knowl Manag E-Learn 2018;10(3):349. [doi: 10.34105/j.kmel.2018.10.019]
53. Kaphle S, Matheke-Fischer M, Lesh N. Effect of performance feedback on community health workers' motivation and performance in Madhya Pradesh, India: a randomized controlled trial. JMIR Public Health Surveill 2016 Dec 07;2(2):e169 [FREE Full text] [doi: 10.2196/publichealth.3381] [Medline: 27927607]

54. Pollack TM, Nhung VT, Vinh DT, Hao DT, Trang LT, Duc PA, et al. Building HIV healthcare worker capacity through telehealth in Vietnam. BMJ Glob Health 2020 Apr 8;5(4):e002166 [FREE Full text] [doi: 10.1136/bmjgh-2019-002166] [Medline: 32337087]
55. Gross CM. Is WhatsApp Suitable for Transmitting Learning Content? Global Digital Health Forum. 2017 Dec 6. URL: https://tinyurl.com/4hhyhap4 [accessed 2022-05-17]
56. Dzuwa H. Improving the quality of child health care in Malawi at the primary level through decision-support tools. Integrated Community Case Management (iCCM). 2016. URL: https://lib.digitalsquare.io/handle/123456789/77604 [accessed 2021-01-30]
57. Gross CM. Adapting eLearning content for an interactive voice response course for professional development in Kenya. Global Digital Health Forum. 2017 Dec 5. URL: https://tinyurl.com/2p8k5573 [accessed 2022-05-17]
58. Asgary R, Adongo PB, Nwameme A, Cole HV, Maya E, Liu M, et al. mHealth to train community health nurses in visual inspection with acetic acid for cervical cancer screening in Ghana. J Low Genit Tract Dis 2016 Jul;20(3):239-242 [FREE Full text] [doi: 10.1097/LGT.0000000000000207] [Medline: 27030884]
59. Nieminen T, Prättälä R, Martelin T, Härkänen T, Hyyppä MT, Alanen E, et al. Social capital, health behaviours and health: a population-based associational study. BMC Public Health 2013 Jun 27;13:613 [FREE Full text] [doi: 10.1186/1471-2458-13-613] [Medline: 23805881]
60. Putnam RD. Bowling alone: America's declining social capital. J Democr 1995 Jan;6(1):65-78. [doi: 10.1353/jod.1995.0002]
61. Inglehart R. Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies. Princeton, NJ, USA: Princeton University Press; May 25, 1997.
62. Brusamento S, Kyaw BM, Whiting P, Li L, Tudor Car L. Digital health professions education in the field of pediatrics: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Sep 25;21(9):e14231 [FREE Full text] [doi: 10.2196/14231] [Medline: 31573906]
63. 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] [Medline: 31829208]
64. Car J, Carlstedt-Duke J, Tudor Car L, Posadzki P, Whiting P, Zary N, Digital Health Education Collaboration. Digital education in health professions: the need for overarching evidence synthesis. J Med Internet Res 2019 Feb 14;21(2):e12913 [FREE Full text] [doi: 10.2196/12913] [Medline: 30762583]
65. Fouasson-Chailloux A, Daley P, Menu P, Gross R, Dauty M. Social media in health studies: a systematic review of comparative learning methods. Int J Environ Res Public Health 2022 Feb 15;19(4):2205 [FREE Full text] [doi: 10.3390/ijerph19042205] [Medline: 35206401]
66. Davis WM, Ho K, Last J. Advancing social media in medical education. CMAJ 2015 May 19;187(8):549-550 [FREE Full text] [doi: 10.1503/cmaj.141417] [Medline: 25852033]
67. Latif MZ, Hussain I, Saeed R, Qureshi MA, Maqsood U. Use of smart phones and social media in medical education: trends, advantages, challenges and barriers. Acta Inform Med 2019 Jun;27(2):133-138 [FREE Full text] [doi: 10.5455/aim.2019.27.133-138] [Medline: 31452573]
68. Wang T, Guo X, Wu T. Social capital and digital divide: implications for mobile health policy in developing countries. J Healthc Eng 2021 Jan 26;2021:6651786 [FREE Full text] [doi: 10.1155/2021/6651786] [Medline: 33575019]
69. Oleribe OE, Momoh J, Uzochukwu BS, Mbofana F, Adebiyi A, Barbera T, et al. Identifying key challenges facing healthcare systems in Africa and potential solutions. Int J Gen Med 2019 Nov 6;12:395-403 [FREE Full text] [doi: 10.2147/IJGM.S223882] [Medline: 31819592]
70. De Ponti R, Marazzato J, Maresca AM, Rovera F, Carcano G, Ferrario MM. Pre-graduation medical training including virtual reality during COVID-19 pandemic: a report on students' perception. BMC Med Educ 2020 Sep 25;20(1):332 [FREE Full text] [doi: 10.1186/s12909-020-02245-8] [Medline: 32977781]
71. Figueroa CA, Luo T, Aguilera A, Lyles CR. The need for feminist intersectionality in digital health. Lancet Digit Health 2021 Aug;3(8):e526-e533 [FREE Full text] [doi: 10.1016/S2589-7500(21)00118-7] [Medline: 34325855]
72. Pham Q, Wiljer D, Cafazzo JA. Beyond the randomized controlled trial: a review of alternatives in mHealth clinical trial methods. JMIR Mhealth Uhealth 2016 Sep 09;4(3):e107 [FREE Full text] [doi: 10.2196/mhealth.5720] [Medline: 27613084]
73. Luna D, Almerares A, Mayan JC, González Bernaldo de Quirós F, Otero C. Health informatics in developing countries: going beyond pilot practices to sustainable implementations: a review of the current challenges. Healthc Inform Res 2014 Jan;20(1):3-10 [FREE Full text] [doi: 10.4258/hir.2014.20.1.3] [Medline: 24627813]
74. Chmela-Jones KA. Students' experiences of human-centred interventions in a design education setting. S Afr J High Educ 2019 Dec;33(6):55-69. [doi: 10.20853/33-6-2936]
75. Phiri S, Feldacker C, Chaweza T, Mlundira L, Tweya H, Speight C, Lighthouse Group. Integrating reproductive health services into HIV care: strategies for successful implementation in a low-resource HIV clinic in Lilongwe, Malawi. J Fam Plann Reprod Health Care 2016 Jan;42(1):17-23 [FREE Full text] [doi: 10.1136/jfprhc-2013-100816] [Medline: 25902815]
76. Nadhamuni S, John O, Kulkarni M, Nanda E, Venkatraman S, Varma D, et al. Driving digital transformation of comprehensive primary health services at scale in India: an enterprise architecture framework. BMJ Glob Health 2021 Jul;6(Suppl 5):e005242 [FREE Full text] [doi: 10.1136/bmjgh-2021-005242] [Medline: 34312149]

77. Iyawa GE, Herselman M, Botha A. Digital health innovation ecosystems: from systematic literature review to conceptual framework. Procedia Comput Sci 2016;100:244-252. [doi: 10.1016/j.procs.2016.09.149]
78. Mars M, Morris C, Scott RE. WhatsApp guidelines - what guidelines? A literature review. J Telemed Telecare 2019 Oct;25(9):524-529. [doi: 10.1177/1357633X19873233] [Medline: 31631763]
79. Ndlovu K, Scott RE, Mars M. Interoperability opportunities and challenges in linking mhealth applications and eRecord systems: Botswana as an exemplar. BMC Med Inform Decis Mak 2021 Aug 21;21(1):246 [FREE Full text] [doi: 10.1186/s12911-021-01606-7] [Medline: 34419020]

Abbreviations

CPD: continuing professional development
LMIC: low- and middle-income country
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by T Leung; submitted 03.08.21; peer-reviewed by L Suppan, R Kyabaggu, D Davies, JM Ocampo; comments to author 15.01.22; revised version received 26.01.22; accepted 26.04.22; published 07.06.22.

Please cite as:
Guillaume D, Troncoso E, Duroseau B, Bluestone J, Fullerton J
Mobile-Social Learning for Continuing Professional Development in Low- and Middle-Income Countries: Integrative Review
JMIR Med Educ 2022;8(2):e32614
URL: https://mededu.jmir.org/2022/2/e32614
doi:10.2196/32614
PMID:35671080

Viewpoint

Needs, Challenges, and Applications of Artificial Intelligence in Medical Education Curriculum

Joel Grunhut[1*], BA; Oge Marques[2,3*], PhD; Adam T M Wyatt[4*], PhD

[1]Schmidt College of Medicine, Florida Atlantic University, Boca Raton, FL, United States

[2]Department of Electrical Engineering and Computer Science, College of Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL, United States

[3]Department of Biomedical Sciences, Schmidt College of Medicine, Florida Atlantic University, Boca Raton, FL, United States

[4]Department of Population Health and Social Medicine, Schmidt College of Medicine, Florida Atlantic University, Boca Raton, FL, United States

[*]all authors contributed equally

Corresponding Author:
Joel Grunhut, BA
Schmidt College of Medicine
Florida Atlantic University
777 Glades Road BC-71
Boca Raton, FL, 33431
United States
Phone: 1 561 297 4828
Email: jgrunhut2019@health.fau.edu

Abstract

Artificial intelligence (AI) is on course to become a mainstay in the patient’s room, physician’s office, and the surgical suite. Current advancements in health care technology might put future physicians in an insufficiently equipped position to deal with the advancements and challenges brought about by AI and machine learning solutions. Physicians will be tasked regularly with clinical decision-making with the assistance of AI-driven predictions. Present-day physicians are not trained to incorporate the suggestions of such predictions on a regular basis nor are they knowledgeable in an ethical approach to incorporating AI in their practice and evolving standards of care. Medical schools do not currently incorporate AI in their curriculum due to several factors, including the lack of faculty expertise, the lack of evidence to support the growing desire by students to learn about AI, or the lack of Liaison Committee on Medical Education’s guidance on AI in medical education. Medical schools should incorporate AI in the curriculum as a longitudinal thread in current subjects. Current students should understand the breadth of AI tools, the framework of engineering and designing AI solutions to clinical issues, and the role of data in the development of AI innovations. Study cases in the curriculum should include an AI recommendation that may present critical decision-making challenges. Finally, the ethical implications of AI in medicine must be at the forefront of any comprehensive medical education.

(JMIR Med Educ 2022;8(2):e35587) doi:10.2196/35587

KEYWORDS

artificial intelligence; AI; medical education; medical student

Introduction

Artificial intelligence (AI) and its applications hold great promise for solving many of health care’s global issues including making diagnoses, facilitating diagnostics, decision-making, big-data analytics, and administration [1,2]. AI has the potential to solve the global doctor shortage and bring access to health care to remote areas of the world [2].

Many fields of medicine have already seen benefit from the practical application of AI. Examples include the detection of atrial fibrillation, epilepsy seizures, and hypoglycemia, or the diagnosis of disease based on histopathological examination or medical imaging [3]. The use of AI is not limited to the fields of radiology or pathology; rather, those fields are indicators of the power of AI for image recognition, a singular form of data that transverses many fields spanning from primary care practice to urgent midsurgical decisions. Recent data show that every specialty in medicine is exploring the use of AI in assisting physicians [4]. Deep learning algorithms can make functional sense of increasing amounts of data used by individuals daily through wearables, smartphones, and other mobile monitoring sensors in different areas of medicine [3]. AI will continue to improve its capabilities to perform analysis and provide intelligent actionable recommendations on most forms of data [5]. It is expected that the future advancements in AI will permeate all aspects of medicine [6].

Disruption

As AI continues to make advancements in health care, it is not without challenges. AI is met with resistance from physicians ill-equipped for such an evolution of clinical practice [3]. It is likely that physicians would benefit from the advancements of AI, but an understandable fear of replacement may prevent these opportunities. Furthermore, a lack of knowledge in AI can create skepticism in the trustworthiness of a machine learning prediction. This pushback may be preventing a large part of the health care sector from adapting to AI as other professional sectors continue to use AI solutions to advance their industries [3].

AI will change the dynamics of the traditional clinician-patient dyad to a much more ethically and emotionally complex clinician-AI-patient triad. This will dramatically alter the trust and accountability aspects with ethical, legal, and financial implications [7]. Physicians must be prepared for this great change [8].

Emerging technologies such as AI have the potential to disrupt labor markets maintained through traditional education programs. In order to be resilient to these market disruptions, physician training programs will require change [9]. The current undergraduate medical education (UME) curriculum is increasingly out of sync with the new needs of an evolving technology. Although most medical schools embrace change and strive to regularly update the components of the medical curriculum, a major overhaul is difficult to achieve and may be a hindrance to the implementation of AI in the curriculum. The path to significant curricular reform is difficult because of a variety of factors, including deeply embedded values and the accreditation process [10].

UME must begin to welcome the future and provide students access to a broader scope of health care through AI. Knowledge on data science, assessing algorithmic quality, and differentiating among different AI products are necessary components of medical education, which faculty must begin to incorporate. Medical schools must begin to teach and nurture unique human abilities that give physicians a competitive advantage over computers to establish an irreplaceable role in the future [9].

What skills Do Physicians Need?

The practice of medicine is entering the age of AI in which the use of data to improve clinical decision-making will grow, bringing forth the need for skillful medicine-machine interaction [11]. Educating the next generation of physicians with the right techniques and adaptations to AI will enable them to become part of this emerging data science revolution [8]. Currently, there are different approaches for physicians to become accompanied to AI. There are physicians taking courses in data science, fellowship opportunities, and data scientists entering medical education programs. These represent a small fraction of the total physician population, and therefore a more integrative and forward approach in UME is necessary.

Medical professionals need to be sufficiently trained in AI, its advantages, and its potential to lower cost, improve quality, and expand access to health care. Of equal importance, physicians should be knowledgeable in its shortfalls such as transparency and liability. AI needs to be seamlessly integrated across the different aspects of the curriculum to achieve these goals [11].

When incorporating recommendations from AI solutions to a patient's course of treatment, physicians should be capable to answer any concerns that patients may have. Perhaps even more importantly, physicians are responsible for ensuring that AI becomes a technology beneficial for patient care and not possibly a cause of harm. The technological revolution raises many challenges with regards to ethical considerations of AI-based implementation in health care. Minority exclusions in databases, issues with legal protections, and a decrease in humanistic touch, among other ethical issues, raise concern for an adaptation of AI in health care. These reasons bring forth the importance of acquiring sufficient knowledge and experience about AI, an obligation of high importance for future physicians [12].

Medical schools should take necessary steps to educate students with widespread knowledge of basic and clinical medicines along with data science, biostatistics, bioethical implications of AI, and evidence-based medicine. Part of a medical student's training should include developing the abilities to distinguish correct information from rhetoric and to understand how to create and disseminate thoroughly validated, trustworthy information for patients and the public [12].

Suggested Steps

Currently, the state of AI in medical education is in its infancy and speculative stages [13-16]. Previously, we have shown that the majority of published literature on the topic call for change in undergraduate medical education and that research is necessary to support curricular changes [17]. Even so, few have given thought to the steps that must be taken to create this change. This is expected because of the difficulty in implementing major curricular changes. Here, we provide an outline of the perceived difficulties and offer solutions to resolve these challenges (Table 1).

Medical school curricular changes are difficult to implement due to resistance to change. This resistance is justified through a lack of consensus on how to incorporate change and an already busy curriculum. For this reason, proposing additional courses or workload will likely be met with criticism from medical education faculty in the best interest of medical students. On a level of national infrastructure, these issues can be supported by leaders of medical education and organizations. These perceptions can be clarified easily through the addition of 1-3 questions on the annual Association of American Medical Colleges Graduation Questionnaire to gauge interest and ability over time from students. For example, the questionnaire can ask for agreement on whether AI should be taught during UME, what year of training it should be taught in, and how it should be incorporated. These can translate into accreditation requirements and drive change forward.

Table 1. Multitiered solutions to include AI[a] in the medical education curriculum.

Levels and target areas of improvement	Examples
National infrastructure	
AAMC[b]	Questionnaire, materials, and guidance
LCME[c]	Minimal requirements and expert panels
Multi-institutional research	Longitudinal research on attitudes and quality improvements
Individual school	
Faculty expansion	Bioethics, bioinformatics, and medical AI experience
Basic knowledge lessons	Introductory courses, benefits and pitfalls of AI, and ethics of AI
Case-based learning	Multispecialty implications in previous cases and biostatistical implications
Student specific	
Journal clubs or reading groups	Specialty-specific journals, health care systems journals, and AI in health care journals
Use of AI in clinical setting	Tumor board, radiology rotation, and point of care ultrasound

[a]AI: artificial intelligence.

[b]AAMC: Association of American Medical Colleges.

[c]LCME: Liaison Committee on Medical Education.

A more overarching question is how undergraduate medical educators can unite to perform high-quality research on the incorporation of AI in the curriculum. Would individual school reports with differing standards in research and protocol do justice to a necessary change across all undergraduate medical education? A joint and united research effort from multiple medical schools would provide a multifaceted and diverse input on the issue and is necessary.

On an individual school level, investments of resources will be necessary to create improvements in the curriculum. A longitudinal AI thread throughout the UME should be advocated to solve time and content constraint issues. Courses teaching evidence-based medicine should incorporate an additional perspective of evaluating the input of AI. Medical school faculty may not be equipped to answer questions or discuss the role of AI in evidence-based medicine. Therefore, it is imperative to add clinician data scientists or computer science and engineering faculty from other schools to medical school teaching faculty.

Students learning public health sciences must be introduced to a background in AI in order to know what AI can and cannot do for their future research and practices. It is far too difficult to teach entire courses of AI in a medical school curriculum. An introductory lecture to AI in medicine is a necessity and should be advocated for. In most schools, the instructor will likely be a data scientist, but it is important that the instructor has teaching experience, is familiar with medical students, and is conversant with the role of AI in medicine. Preferably, the instructor will have already taught about the role of AI in medicine previously in a computer science course. Most importantly, instructors should have levels of competency in bioethics to address the expansive ethical issues AI has brought to health care.

Case-based learning and simulation learning can also incorporate AI-based recommendations in clinical scenarios. By integrating AI into cases and simulations, students can have exposure and familiarity with AI-based solutions. Collaborating and managing AI applications will require a deep understanding of probabilistic reasoning and ethical consideration from medical students [10]. These lessons should be taught from faculty with knowledge about the accuracy and interpretations of AI recommendations. It is likely that medical schools will need to hire additional faculty in this field to ensure quality delivery of AI in medicine content.

Another suggestion is for institutions to offer access to AI in web-based health care programs created at other institutions (eg, Stanford University's Artificial Intelligence in Healthcare professional program). These programs are taught by faculty who are world-renowned experts at the interface of health care and AI, and the programs are available to medical students with a cost. Providing access to a program will enable medical schools to infuse new knowledge in a curriculum that could not be provided otherwise.

On an individual student level, students can help drive change in their education with a proactive engagement with AI. Previously, we have shown that most of the current literature agrees that medical students do not need to learn how to code and create AI tools, but they should understand how AI works and its limitations [17]. Students can read about AI advancements in health care in medical journals. AI research has a strong appearance in many leading journals, including medical AI-themed journals. Reading the current trends in AI in health care will inform and prepare students for the future of health care. When combined with evidence-based medicine learning at medical school, students will be able to assess the integrity of AI research.

Students can expose themselves to AI in the clinical setting as well. There are already reports of students receiving individual portable ultrasounds with AI-driven software to help advance their education [18,19]. Students can see AI applications in their radiology rotations and discuss its role in clinical decision-making with radiologists and clinicians across different specialties involved in patient care impacted by AI applications. AI integration will be specialty specific, but AI as a whole will likely be present across health care. For example, a tumor board, consisting of radiologists, pathologists, oncologists, and surgeons, is impacted by AI applications. Students can tailor their exposure toward fields of interests. The suggestions above apply to medical education as a whole and highlight the importance and room for improvement in general medical education. Specialty-specific AI topics would not be necessary for UME but perhaps in residency and beyond. Nevertheless, through a longitudinal curricular thread in AI, medical schools can expose their students to a wide variety of specialties using AI.

Lessons in decision-making with AI will apply to individuals across the spectrum of health care, both in city and rural settings. It is expected that AI will permeate various health care settings because of its potential to expand access to health care. Therefore, individual students should be prepared by learning about AI no matter what area of medicine they choose.

A Methodological Approach

New medical curricular changes can have tremendous positive and negative impacts on medical students. Additionally, a change in curriculum such as the introduction of an entire novel topic is a difficult task. Therefore, it is important that medical education specialists across different regions work in unison to create and assess the implemented changes. Educational research is vital to assessing the effectiveness of different curricular reforms [20]. One suggestion to achieving such success is to begin a long-term study to measure the outcomes of different implementations of AI in the curriculum.

Research measuring student and faculty attitudes, skill level, and specific needs of AI in UME is crucial and urgent at this point in time. Further efforts to incorporate these suggestions should be measurable and have interpretable data to advance the implementation of AI in UME. A concerted multi-institutional study is a logical approach in order to achieve these goals.

Medical education deans need to gather to discuss and plan AI curricular reform. At the organizational level, medical education governing bodies must enact and promote these changes. The Liaison Committee on Medical Education should provide suggestions and guidelines of how to best incorporate AI to medical education through special committees.

In 2018, the American Medical Association called for “Research regarding the effectiveness of AI instruction in medical education on learning and clinical outcomes” [21]. Three years later, the available literature suggests UME has been slow to address this call [17]. Thus, further efforts should be made to advance this original call into practice.

Conclusion

The current and future advancements of AI in medicine oblige undergraduate medical educators to act and implement AI in the curriculum. Longitudinal research plans are necessary to effectively study how to best achieve these curricular changes. Medical education governing bodies, medical education deans, and medical education researchers should begin to implement AI in the undergraduate medical education curriculum. Moving forward with collective agreement from these entities will ensure current students—our future physicians—receive adequate AI exposure.

Authors' Contributions

All authors contributed equally to this work.

Conflicts of Interest

None declared.

References

1. Chen J. Playing to our human strengths to prepare medical students for the future. Korean J Med Educ 2017 Oct;29(3):193-197 [FREE Full text] [doi: 10.3946/kjme.2017.65] [Medline: 28870022]
2. Meskó B, Hetényi G, Győrffy Z. Will artificial intelligence solve the human resource crisis in healthcare? BMC Health Serv Res 2018 Jul 13;18(1):545 [FREE Full text] [doi: 10.1186/s12913-018-3359-4] [Medline: 30001717]
3. Briganti G, Le Moine O. Artificial Intelligence in Medicine: Today and Tomorrow. Front Med (Lausanne) 2020 Feb 5;7:27 [FREE Full text] [doi: 10.3389/fmed.2020.00027] [Medline: 32118012]
4. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. NPJ Digit Med 2020 Sep 24;3(1):126 [FREE Full text] [doi: 10.1038/s41746-020-00333-z] [Medline: 33043150]
5. Gallix B, Chong J. Artificial intelligence in radiology: who's afraid of the big bad wolf? Eur Radiol 2019 Apr 8;29(4):1637-1639. [doi: 10.1007/s00330-018-5995-9] [Medline: 30734848]
6. Wiljer D, Hakim Z. Developing an Artificial Intelligence-Enabled Health Care Practice: Rewiring Health Care Professions for Better Care. J Med Imaging Radiat Sci 2019 Dec;50(4 Suppl 2):S8-S14. [doi: 10.1016/j.jmir.2019.09.010] [Medline: 31791914]
7. Chang A. From the Moon and Formula One to Beethoven's Ninth Symphony: Musings of a Clinician-Data Scientist. Stanford University Center for Artificial Intelligence in Medicine & Imaging. URL: https://aimi.stanford.edu/events/aimi-invited-lecture-moon-and-formula-one-beethovens-ninth-symphony-musings-clinician-data [accessed 2022-05-31]

8. Kolachalama VB, Garg PS. Machine learning and medical education. NPJ Digit Med 2018 Sep 27;1(1):54 [FREE Full text] [doi: 10.1038/s41746-018-0061-1] [Medline: 31304333]
9. Li D, Kulasegaram K, Hodges BD. Why We Needn't Fear the Machines. Academic Medicine 2019;94(5):623-625. [doi: 10.1097/acm.0000000000002661]
10. Wartman S, Combs C. Reimagining Medical Education in the Age of AI. AMA J Ethics 2019 Feb 01;21(2):E146-E152 [FREE Full text] [doi: 10.1001/amajethics.2019.146] [Medline: 30794124]
11. Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P. Introducing Artificial Intelligence Training in Medical Education. JMIR Med Educ 2019 Dec 03;5(2):e16048 [FREE Full text] [doi: 10.2196/16048] [Medline: 31793895]
12. Park SH, Do K, Kim S, Park JH, Lim Y. What should medical students know about artificial intelligence in medicine? J Educ Eval Health Prof 2019 Jul 03;16:18 [FREE Full text] [doi: 10.3352/jeehp.2019.16.18] [Medline: 31319450]
13. Brouillette M. AI added to the curriculum for doctors-to-be. Nat Med 2019 Dec 05;25(12):1808-1809. [doi: 10.1038/s41591-019-0648-3] [Medline: 31806886]
14. Chan KS, Zary N. Applications and Challenges of Implementing Artificial Intelligence in Medical Education: Integrative Review. JMIR Med Educ 2019 Jul 15;5(1):e13930 [FREE Full text] [doi: 10.2196/13930] [Medline: 31199295]
15. Masters K. Artificial Intelligence developments in medical education: a conceptual and practical framework. MedEdPublish 2020;9(1):06. [doi: 10.15694/mep.2020.000239.1]
16. Srivastava TK, Waghmare L. Implications of Artificial Intelligence (AI) on Dynamics of Medical Education and Care: A Perspective. JCDR 2020:JI01-JI02. [doi: 10.7860/jcdr/2020/43293.13565]
17. Grunhut J, Wyatt AT, Marques O. Educating Future Physicians in Artificial Intelligence (AI): An Integrative Review and Proposed Changes. J Med Educ Curric Dev 2021 Sep 06;8:23821205211036836 [FREE Full text] [doi: 10.1177/23821205211036836] [Medline: 34778562]
18. Ireson M, Warring S, Medina-Inojosa J, O'Malley MT, Pawlina W, Lachman N, et al. First Year Medical Students, Personal Handheld Ultrasound Devices, and Introduction of Insonation in Medical Education. Ann Glob Health 2019 Oct 15;85(1):123 [FREE Full text] [doi: 10.5334/aogh.2565] [Medline: 31673510]
19. Lewis Katz School of Medicine at Temple University is the First Medical School in Philadelphia and the East Coast to Provide All First-Year Students with Butterfly iQ+ Ultrasound Devices. Temple University. URL: https://tinyurl.com/b6jdrcuz [accessed 2022-05-31]
20. Guan C, Mou J, Jiang Z. Artificial intelligence innovation in education: A twenty-year data-driven historical analysis. International Journal of Innovation Studies 2020 Dec;4(4):134-147. [doi: 10.1016/j.ijis.2020.09.001]
21. Augmented Intelligence in Health Care. American Medical Association. 2018. URL: https://www.ama-assn.org/system/files/2019-01/augmented-intelligence-policy-report.pdf [accessed 2022-05-31]

Abbreviations

AI: artificial intelligence
UME: undergraduate medical education

Edited by N Zary; submitted 09.12.21; peer-reviewed by R Russell, SH Mousavi; comments to author 03.02.22; revised version received 06.02.22; accepted 27.05.22; published 07.06.22.

Please cite as:
Grunhut J, Marques O, Wyatt ATM
Needs, Challenges, and Applications of Artificial Intelligence in Medical Education Curriculum
JMIR Med Educ 2022;8(2):e35587
URL: https://mededu.jmir.org/2022/2/e35587
doi:10.2196/35587
PMID:35671077

Original Paper

Artificial Intelligence Education for the Health Workforce: Expert Survey of Approaches and Needs

Kathleen Gray[1], PhD; John Slavotinek[2,3], MD; Gerardo Luis Dimaguila[4], PhD; Dawn Choo[1], PhD

[1]Centre for Digital Transformation of Health, The University of Melbourne, Parkville, Australia
[2]South Australia Medical Imaging, Flinders Medical Centre, Bedford Park, Australia
[3]College of Medicine and Public Health, Flinders University, Adelaide, Australia
[4]Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Australia

Corresponding Author:
Kathleen Gray, PhD
Centre for Digital Transformation of Health
The University of Melbourne
Level 13, VCCC Building
305 Grattan St
Parkville, 3010
Australia
Phone: 61 3 8344 8936
Email: kgray@unimelb.edu.au

Abstract

Background: The preparation of the current and future health workforce for the possibility of using artificial intelligence (AI) in health care is a growing concern as AI applications emerge in various care settings and specializations. At present, there is no obvious consensus among educators about what needs to be learned or how this learning may be supported or assessed.

Objective: Our study aims to explore health care education experts' ideas and plans for preparing the health workforce to work with AI and identify critical gaps in curriculum and educational resources across a national health care system.

Methods: A survey canvassed expert views on AI education for the health workforce in terms of educational strategies, subject matter priorities, meaningful learning activities, desired attitudes, and skills. A total of 39 senior people from different health workforce subgroups across Australia provided ratings and free-text responses in late 2020.

Results: The responses highlighted the importance of education on ethical implications, suitability of large data sets for use in AI clinical applications, principles of machine learning, and specific diagnosis and treatment applications of AI as well as alterations to cognitive load during clinical work and the interaction between humans and machines in clinical settings. Respondents also outlined barriers to implementation, such as lack of governance structures and processes, resource constraints, and cultural adjustment.

Conclusions: Further work around the world of the kind reported in this survey can assist educators and education authorities who are responsible for preparing the health workforce to minimize the risks and realize the benefits of implementing AI in health care.

(JMIR Med Educ 2022;8(2):e35223) doi:10.2196/35223

KEYWORDS

artificial intelligence; curriculum; ethics; human-computer interaction; interprofessional education; machine learning; natural language processing; professional development; robotics

Introduction

Background

Artificial intelligence (AI) is widely expected to have broad and deep impacts on health care. In the past few years, several books have appeared on this topic whose covers connect it variously with business success, creative destruction, robotic assistance, care that is more human, and treatment that is more precise. Dozens of review papers have synthesized the growing body of scientific literature focusing on applications in an array of sociotechnical factors and care specializations: aged care, decision support, efficiencies, emergency medicine, ethics, nursing, pathology, psychiatry, and workflows.

Across the board, a key consideration is how to prepare the current and future health workforce for the possibility of using AI in health care. The World Health Organization and many national health systems around the world have flagged the importance of a health workforce that understands how to work properly with AI [1]. However, there is no straightforward answer to the question of how to provide education and professional development that prepares the health workforce adequately to do so (this topic is not to be confused with the use of AI-supported teaching techniques and learning analytics tools in health professions education, as reviewed by Hasan Sapci and Aylin Sapci [2]).

Disproportionately few peer-reviewed papers have been written on this aspect of AI in health care. Those extant studies focus chiefly on medical professionals [3-10]; there is scant reference to other health professions (eg, nursing [11]) or the health workforce broadly [12,13]. Almost all of them are commentaries by individuals advocating for appropriate education and development. A few papers are based on surveying the attitudes or knowledge of current medical students about AI [14], and one paper even offers a way to measure medical students' readiness to learn about AI [15]. At present, there is no obvious consensus among educators about what needs to be learned or how this learning may be supported or assessed.

In Australia, although medical specialist colleges such as radiology, dermatology, and ophthalmology are developing AI competencies and training packages for their members, most people in the health sector whose work will be affected by the increased use of AI have minimal access to relevant education or professional development. The Australian Alliance for Artificial Intelligence in Healthcare—a nonprofit network of >100 partners and stakeholders in academia, government, consumer, clinical, industry organizations, and peak bodies seeking to translate AI technologies into health services [16]—offered a way to consult widely about how to approach AI education and development for a national health workforce. The Alliance's Workforce Working Group, convened by KG and JS, gave rise to a project to gather and share information on how educational authorities are preparing Australia's health workforce to work with AI and what gaps in curriculum and educational resources may need attention.

Objective

The aim of this paper is to provide an overview and summary of educators' ideas and plans for educating the health workforce on the use of AI in health systems and services as a basis for strategic planning, investment, and further research in this area.

Methods

Overview

We used an expert survey method to gather information about AI education for the Australian health workforce. For the purpose of this work, we followed the Australian Institute of Health and Welfare (2019) definition to scope the health workforce [17], covering the 15 kinds of health care practitioners registered by the Australian Health Practitioner Regulation Agency (AHPRA) Boards, as well as other health care professionals (eg, audiologists and speech pathologists). Other professionals such as health administration and support workers (eg, health information managers, health technology suppliers, and health researchers) were also considered. We considered education that might be delivered in many contexts, such as a formal study program for entry into the workforce, continuing professional development endorsed as appropriate to maintain currency in the workforce, postbasic training leading to recognition as a specialist in the workforce, and an examination to certify competence as required to practice legally in the workforce.

Participants

The intended participants were individuals who self-identified as having high-level expertise and experience in health workforce education and development; importantly, they were not required or expected to be experts in AI. They must indicate a relevant role that they held currently in a related organization for their responses to be included in this study; eligible roles included but were not limited to manager, coordinator, director, or committee chair (paid or unpaid) of an education or development portfolio. Respondents might hold such a role in more than one organization; they were not required to identify themselves or their organizational affiliations. The responses were expected to express the informed perspective of that individual and were not expected to be the official view of any organization. Although the exact number of potential respondents was unknown, the Alliance's Workforce Working Group considered it feasible to reach at least 100 people.

Survey Design

A 6-part survey was developed based on a scan of scholarly and gray literature about AI in health care. Our scan used search terms that occurred in combination in items published during the 2018-2020 period retrieved from Google Scholar and Google, representing the three intersecting fields of interest: AI (including expert systems, machine learning, and robotics), health (including medicine, nursing, allied health, and digital health), and education (including curriculum, teaching, and professional development). The sources of general relevance to our research are referenced in the *Introduction* section.

However, we found no existing question set suitable for our purpose, so we selected definitions and terms from recent authoritative sources. For example, we used a broad definition of AI that included machine learning, natural language processing, computer vision, and chatbots following the Academy of Medical Royal Colleges in the United Kingdom [18]. We derived a list of specific topics [19] and, for each topic, we provided a brief scope note for non-AI experts [19-29]. We derived a set of attitudes and beliefs from work by an internationally recognized advocate for professional development in the field of AI in medicine [30]. The survey was worded so that it could capture perspectives on AI education across professions and jurisdictions and allow for the expression of ideas about educating for organizational and technological change and social and global responsibility following the recommendations by Frenk et al [31]. The survey sections were

(1) roles held in relevant types of organizations, (2) educational strategies and approaches in use or intended, (3) specific AI topics that are important in education and current content available, (4) learning activities and experiences that are important to support education, (5) attitudes and beliefs about AI that are important for education to address, and (6) additional comments.

Part 1 provided a list of relevant types of organizations and was the only section that was compulsory to complete; the others could be skipped over. Parts 1, 2, and 6 provided free-text response options. Parts 3, 4, and 5 provided 5-point Likert scale response options for 12-15 statements each plus free-text response options. The details of the survey items are presented in Multimedia Appendix 1.

The survey items were piloted with members of the Alliance's Workforce Working Group. They found that the survey required approximately half an hour to complete; this was considered a possible deterrent but nevertheless an efficient way to elicit initial input on a complex topic from a range of educational experts.

Data Collection and Analysis

A survey website and web-based form were created and tested using the Qualtrics (Qualtrics International, Inc) account of the University of Melbourne, and responses were monitored and summarized progressively during the period from October 2020 to December 2020 by GD. Recruitment occurred mainly through the Alliance's electronic communication channels with members and partners, with periodic reminders and targeted follow-up messages to publicly listed contacts of major organizations such as the AHPRA Boards and professional colleges. Raw data were stored on a secure Qualtrics server in Australia.

Data were deidentified and aggregated, and distinctive written expressions were paraphrased so that no individual or organization would be readily identifiable. Descriptive statistical analysis of Likert scale data was performed using standard software in Qualtrics; the low number of responses did not warrant inferential analysis. Free-text data were thematically analyzed using grounded theory. Data were coded by an experienced qualitative analyst working independently (DC) and then reviewed jointly with another analyst (KG) until they reached an agreement on data interpretation and representation. A detailed initial report on quantitative and qualitative data was reviewed and critiqued by a meeting of the Alliance's Workforce Working Group before the data were summarized further for publication.

Ethics Approval

This study received Human Research Ethics approval (2056392.1) from the University of Melbourne.

Results

A total of 103 people accessed the survey website as recorded by their responses to a verification question, of which 81 (78.6%) proceeded through the participant information home page and a consent form webpage and clicked on the *start survey* button. Of those 81 participants, 39 (48%) completed part 1, 29 (36%) completed part 2, 25 (31%) completed part 3, 23 (28%) completed part 4, 23 (28%) completed part 5, and 15 (19%) completed part 6.

Part 1: Educational Experts' Focus Areas

Most of the 39 survey respondents who completed part 1 held senior education-related roles in one or more education and training organizations: 46% (18/39) had roles in universities or other government-registered training organizations, and 17% (7/39) had roles in unregistered professional or industry training providers. The next largest group of respondents (5/39, 13%) held senior education-related roles in government-registered health care provider organizations. Of the remaining respondents who specified an organization type, 5% (2/39) each were from a national accrediting body, an independent medical research institute, a professional association, or an industry association and 3% (1/39) each were from an unregulated accrediting body or health care provider. The details are summarized in Table 1.

Table 1. The organizations of the educational experts.

Type of organizations where the experts were based	Respondents, n (%)
Education and training provider within the scope of the Australian National Training System	18 (46)
Other education and training provider	7 (17)
Organization registered as a health care provider with the Australian Department of Health	5 (13)
National board that registers practitioners and students and accredits education programs within the scope of the Australian Health Practitioner Regulation Agency	2 (5)
Independent medical research institute	2 (5)
Professional association	2 (5)
Industry association	2 (5)
Organization type not otherwise specified	2 (5)
Other organization that certifies or accredits individuals or programs	1 (3)
Other organization that provides health care services	1 (3)

The health workforce of immediate focus for most survey respondents (31/39, 80%) was working health care practitioners, mainly those registered with AHPRA (respondents were involved in education of enrolled nurses, medical professionals, midwives, nurse practitioners, paramedics, podiatrists, physicians, and dental practitioners) and speech pathologists, a nonregulated group. Four other workforce subgroups were the main concern among 5% (2/39) and 10% (4/39) of respondents: nonprofessional aged care, disability care, and community care workers; health information communication technology and health informatics workers; health data scientists and biomedical researchers; and university students across the range of the health workforce.

Part 2: Organizational Approaches to AI Education

There were 36% (29/81) of survey respondents who offered insights into strategic thinking about how the current and future health workforce will acquire knowledge and skills to work with AI in health care. They identified distinct enablers and barriers to implementing strategic actions. They also provided examples of activities being planned or underway in their organizations to support learning and development.

Organizational strategizing ranged across four stages: stage 0 (meaning not yet under consideration), stage 1 (meaning there was consideration, exploration, and planning), stage 2 (meaning implementation was being designed), and stage 3 (meaning implementation was occurring). In some quarters, no strategic thought was being given to workforce learning and development for AI in health care; an example of stage 0 was "This has not been a focus area at this point in time for this organisation." An example of stage 1 was "We use accreditation reports from universities on how they are adapting their courses for the future and meetings with health care and government stakeholders." Some consideration, exploration, and formal planning approaches were underway, formulating initial ideas about what AI education was relevant for the current and future workforce. Respondents described overarching strategic considerations such as the clinical currency of the workforce and the upskilling of the workforce to display excellence in care. Among the approaches in which they were involved, they reported horizon scanning, identifying future roles for the profession, aggregating feedback from staff about professional knowledge gaps with impacts on service delivery and outcomes, consulting with senior leadership, and reviewing government policies and sector literature. More formal planning that they reported included conducting a review of professional performance frameworks and accreditation reports from universities, establishing advisory groups, instigating future needs planning within committees, and meeting with government stakeholders, as well as referencing AI education in strategy and planning documents. An example of stage 2 was "The inclusion of AI/ML [artificial intelligence and machine learning]...in all courses to prepare graduates for the workforce of the future is of great strategic importance and strategies are in place to commence this work." This stage was marked by educational design strategies such as planning to integrate AI topics in graduate coursework, taking AI in health care into account when conducting a curriculum review, and planning specific content for continuing professional development. An example of stage 3 was the implementation of "specialized courses, micro-credits, for medical school students." Respondents at this stage gave examples of specific activities that organizations were undertaking. These included developing specialized training courses, publishing articles in member newsletters to create awareness, promoting AI and ethics education in health care, using emerging and new technology in the delivery of education, and using simulated training in education.

Seven themes emerged to describe key opportunities and enablers of these organizational strategies: mobilization of expertise, influential leadership, leveraging collaborations, expanding continuing professional development, higher education planning and programming, government drivers, and health service improvements. Respondents saw possibilities to access academic expertise on AI and its applications, establish links with and participate in networks of experts, and become involved in consultative forums. They perceived that the championing of AI education for the health workforce was facilitated by dynamic thinkers who hold senior roles within organizations and by instrumental health care and academic stakeholders who are rallying to have greater influence. They also reported that there were strategic opportunities available within their organizations; for instance, to forge cross-sector collaborations, leverage support from medical colleges for AI education, and capitalize on investments in AI hubs. A further suggestion was that AI education for the health workforce could be facilitated by advocating for health service improvements; in other words, for infrastructure and quality improvements that would give rise to better patient experiences and outcomes and increased productivity and economic benefits. Already existing requirements for continuing professional development were recognized as enablers; continuing professional development initiatives around AI could upskill people in the workforce, including those in clinical and supervisory roles. Similarly, survey respondents looked to find ways through higher education planning and programming to use university resources, invest in academic teaching in the area, integrate AI applications in the design of instructional delivery, and comprehensively review courses so as to implement AI education within them. Respondents thought that government policy development, endorsement, and support were important drivers of AI education for the health workforce; for example, "a strong desire from the Health Department as a policy maker, funder and implementer."

There were three major kinds of challenges and barriers to these organizational strategies: the lack of governance structures and processes, resource constraints, and cultural unreadiness. Respondents described ambiguity about the roles of workforce organizations and government departments in AI education for the health workforce and no clarity around processes for further investment in this area; a respondent said that "whilst we track and identify and amplify emerging issues...it still needs government regulation and funding to make the changes called for." Resource constraints consisted of interrelated human, time, and funding issues. Human resource challenges were identified as "lack of a dedicated workforce" and limitations in expertise in the form of "not having a strong background in technology in healthcare" and "access to skills and consultants." Time issues commented on were "competing priorities" and "curriculum already over-burdened." Funding constraints were described as "limited university resources, poor external support from government," "funding to do meaningful research and ongoing education," and "funding to access and deploy technological solutions." In terms of cultural unreadiness for AI education, respondents described "a challenge as we bridge the gap between the early and late adopters"; "a combination of resisting everything new, lack of knowledge in the area and complexity of implementing in an already over-stuffed curriculum"; and "compartmentalisation of the educational program."

Activities that respondents reported preparing or implementing in their organizations to advance workforce knowledge and skill building regarding AI included auditing their current resources and skills, engaging stakeholders and developing research programs to improve engagement, conducting expert and stakeholder focus groups to identify growth opportunities, mobilizing capability and potential collaborations, providing relevant work placements, scoping current courses and qualifications available, and doing a gap analysis. Specific examples included seeking "high level advice on workforce future needs, and ways in which new course structures can enable that"; developing "surveys, prompt cards, reference videos, face-to-face training sessions, reference manuals, escalation plans"; proposing "short, non-compulsory, advanced courses for medical under- and post-graduates"; and using AI-enabled web-based learning platforms.

Part 3: AI Educational Content and Provision

Approximately 31% (25/81) of survey respondents ranked a list of specific topics that could potentially be considered essential for health workforce competence in the next decade and judged whether the related education available to the health workforce at the time of the survey was sufficient to meet the need. Figure 1 shows the topics deemed essential, juxtaposed with the adjudged adequacy of the education being provided. Discrepancies between the importance of a topic and the adequacy of current education provision were consistently observed across all topics. Most respondents agreed or strongly agreed on the importance of three essential topics: criteria for judging whether large data sets are suitable for use in high-value clinical AI applications (20/25, 81%), general ethical implications (20/25, 81%), and machine learning (19/81, 76%). In contrast, only 15% (4/25), 22% (6/25), and 14% (4/25) of respondents agreed that these 3 topics were being taught adequately, respectively. This contrast was apparent even in the two topics that were ranked lowest although still considered essential by over half of respondents—natural language processing (15/25, 59%) and robotic process automation (14/25, 55%)—and only 14% (4/25) and 10% (3/25) of respondents thought that current education provision was adequate on each topic, respectively.

Figure 1. Artificial intelligence topics essential to teach and now taught. AI: artificial intelligence.

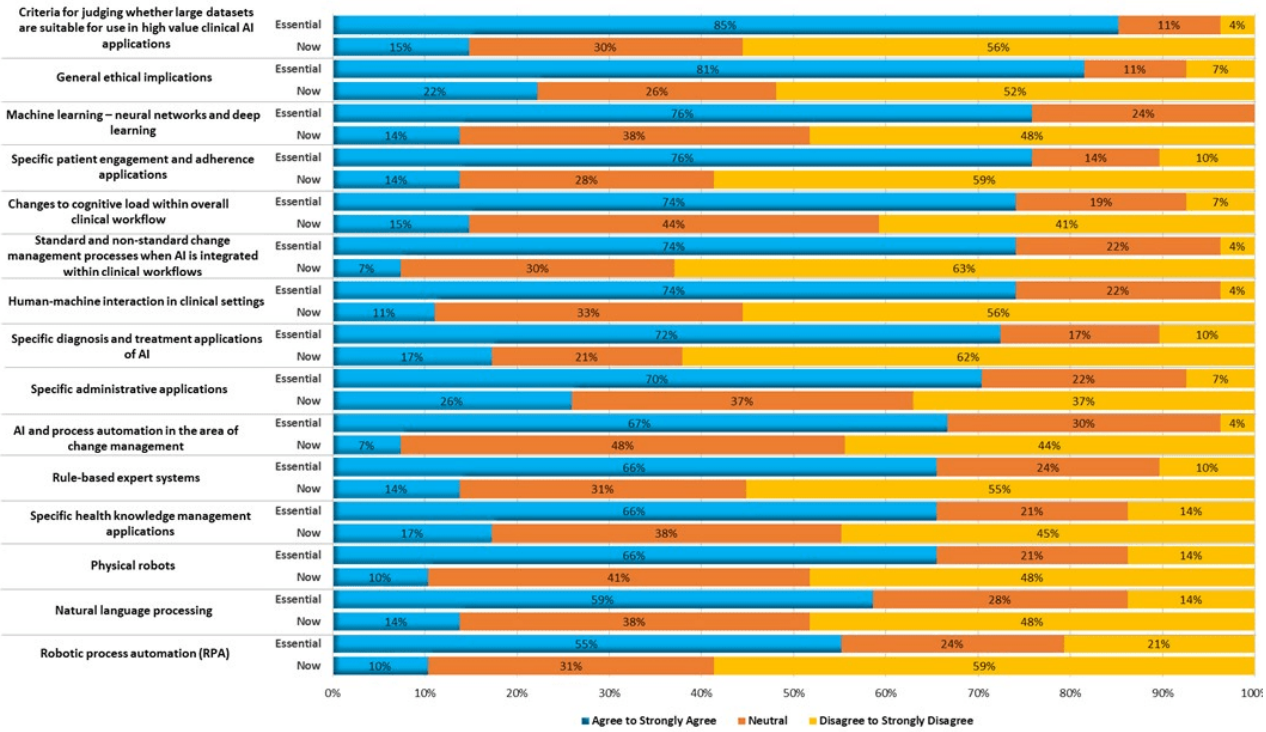


Table 2 synthesizes and paraphrases the respondents' free-text comments elaborating on this part of the survey, with a few particularly pointed remarks quoted in full. The respondents also outlined additional AI education topics that they considered essential related to privacy, data security, product evaluation (including knowledge of adverse events), data accuracy, and representativeness of data sets for the Australian context. Additional importance was placed on "explainability, ... education in AI/ML programming using open source tools, AI governance skills," and preparing the workforce for "what to do when AI stops working."

Table 2. Comments on artificial intelligence (AI) educational content.

Topics	Key themes and remarks[a]
Criteria for judging whether large data sets are suitable for use in high-value clinical AI applications	• Need for understanding the use of algorithms in machine learning • Limited access to informational resources for current students enrolled in clinical degrees • Useful and relevant topic for research training and participation
General ethical implications	• Complex area • Tailoring breadth and depth of training and informational tools would be warranted for different roles and contexts • "This is a minefield area!"
Machine learning—neural networks and deep learning	• Specific concepts to include the applications, implications, consequences, and limitations of using machine learning • Focus on future needs • Targeted and focused development of a group of individuals rather than the whole workforce
Specific patient engagement and adherence applications	• Specialized training needs will be required depending on the care pathway, specific apps, and progress of these technologies • Having established governance structures around value-based health care to complement education • Changes to cognitive load within overall clinical workflow • Tailoring levels of educational uptake for different disciplines • "Models of care using these tools need to be built and clinically governed"
Changes to cognitive load within overall clinical workflow	• Tailoring levels of educational uptake for different disciplines
Change management processes when AI is integrated within clinical workflows	• Not currently part of education for health care practitioners • "All of our training is still delivered face-to-face"
Human-machine interaction in clinical settings	• Lack of informational access to this topic for current students in clinical degrees • This topic could be reframed as part of ethical issue training
Specific diagnosis and treatment applications of AI	• Might not be relevant to certain workforce roles • End users of specific diagnosis and treatment applications of AI might not require in-depth specialist training and education • This area will need to evolve to meet future needs (ie, development of standards and clinical governance regarding skill competencies)
Specific administrative applications	• Area of interest given that a workforce competent in specific administrative applications would bring about productivity and clinical quality benefits • Digital and ICT[b] specialist workforce will require knowledge of specific administrative applications; the health care workforce could contribute by providing clinical input in this area • AI and process automation in the area of change management • Inclusion of risk management strategies in education and training • "Knowing about the very many near misses is more important for the purposes of refining AIML[c] than critical incidents alone"
Rule-based expert systems	• Area of great potential and benefit (ie, reduction in cognitive load errors in emergency and intensive care settings) • Further analysis required to understand the health workforce's receptivity toward using rule-based expert systems and the implications for clinical practice in the next decade • Specific health knowledge management applications
Specific health knowledge management applications	• Limited access to resources (ie, databases) • Participation in research projects was a way to promote learning

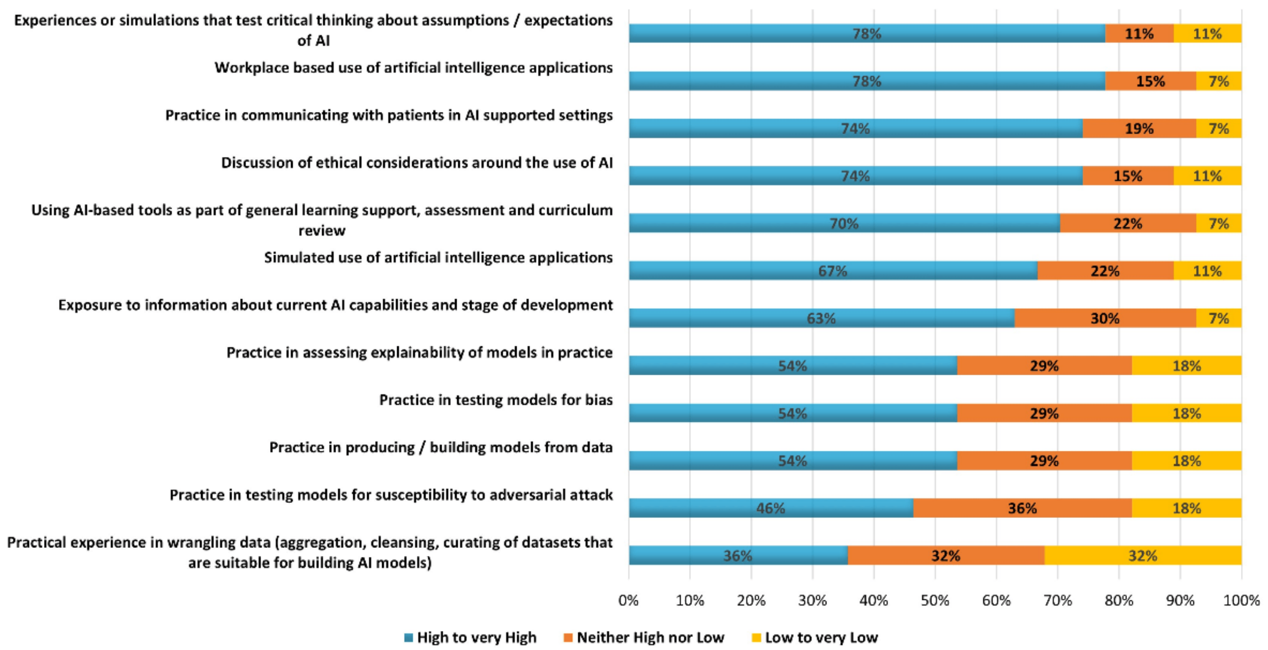
Topics	Key themes and remarks[a]
Physical robots	• Highly relevant to medical professionals, nursing, aged care, and allied health • A lack of clarity around the implications of using physical robots in clinical practice (ie, concrete examples would be required to understand how health workforce job roles might interact with physical robots) • The need for training to be value-adding to ensure that physical robots improve and do not hinder health care workflow • "Doctors probably learn more about robots from their kids' toys than from their training."
NLP[d]	• Lack of access to real-world health data to teach learners about using algorithms; limited number of education opportunities and digital health literacy resources to support learning • The clinical workforce might only require a general understanding of how NLP tools work, its applications, limitations, and consequences of use in health care • Expertise of NLP specialists could be leveraged
RPA[e]	• Important for the digital and ICT workforce to acquire skills in this area to support the health workforce in automating processes • Health workforce could benefit from greater knowledge of ways to identify opportunities to apply RPA.

[a]Pointed remarks are in quotes.
[b]ICT: information and communication technology.
[c]AIML: artificial intelligence and machine learning.
[d]NLP: natural language processing.
[e]RPA: robotic process automation.

Part 4: AI Educational Methods

Approximately 19% (23/81) of survey respondents addressed a set of 12 statements describing educational learning activities that could be used to build knowledge and skills for working with AI in health care and rated them according to the value they perceived each method to have (Figure 2). In total, 6 methods of learning were thought to be highly or very highly valuable by two-third (15/23, 67%) to three-quarter (17/23, 74%) of respondents, whereas only two methods were thought valuable by fewer than half of all respondents; namely, practice in testing models for vulnerability (11/23, 46%) and practice in wrangling data (8/23, 36%).

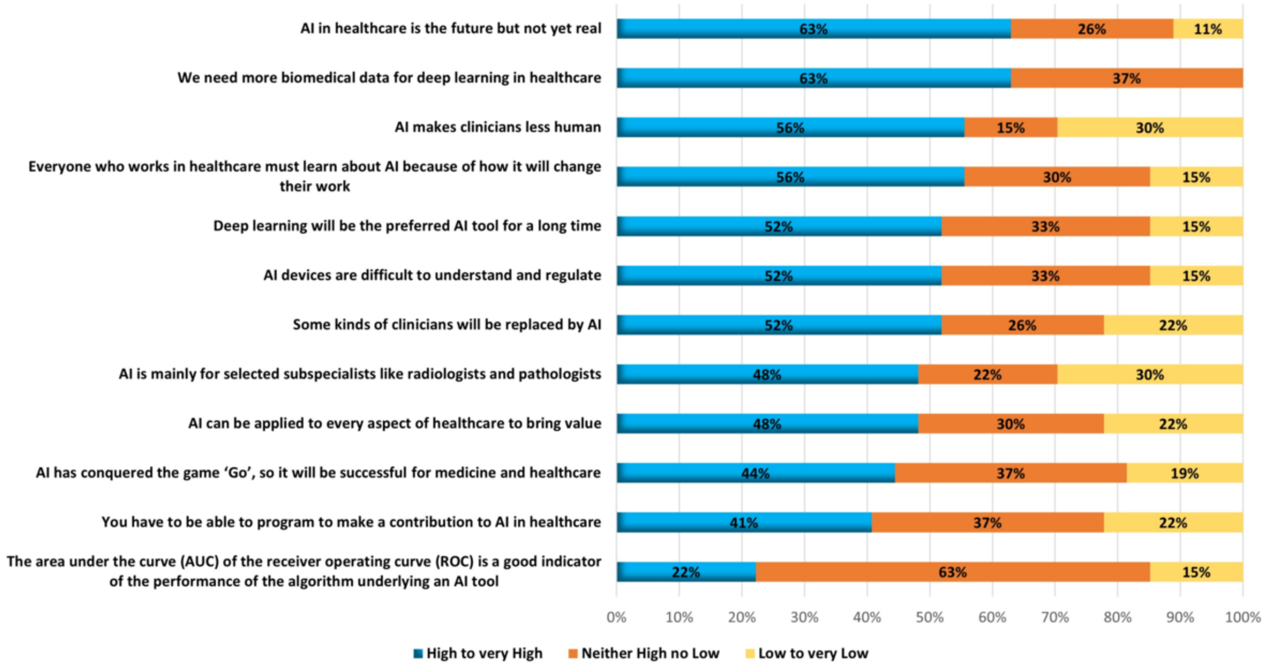
Figure 2. Value of different educational methods to teach artificial intelligence (AI).



Part 5: Education Regarding AI Attitudes and Beliefs

Approximately 28% (23/81) of survey respondents considered a set of 12 statements describing poorly informed but widely held attitudes and beliefs about AI and rated them according to the importance of addressing them in AI education (Figure 3). A total of 7 attitudes were considered highly or very highly necessary to address in education by most respondents (between 12/23, 52% and 14/23, 63%). In total, 63% (14/23) of respondents were undecided about the following statement—*The area under the curve (AUC) of the receiver operating curve (ROC) is a good indicator of the performance of the algorithm underlying an AI tool*—with fewer than one-quarter (5/23, 22%) rating it important for education to address this belief.

Figure 3. Importance of education to address attitudes and beliefs about artificial intelligence (AI).



Discussion

Principal Findings

This survey gathered detailed observations about AI education from a range of experts in senior roles associated with health workforce training and professional development. Their thought leadership reflects an array of mainstream health care professions as well as important groups who are ancillary to frontline health care workers. Responses to this survey showed that many different health workforce subgroups and interest groups have a stake in education and professional development on AI in health care; this is not only of concern in selected medical specializations but also in fields ranging from aged care to speech pathology. As this study reached throughout a national health system and elicited responses from a wide range of senior stakeholders and influencers in health workforce education and professional development, it makes a distinctive contribution compared with much of the previously published work on this topic, which reflects expert perspectives of individuals or narrowly defined workforce subgroups.

The development of the survey instrument was informed by a multifaceted view of the curriculum, so it elicited ideas not only about what should be learned (topics) but also how (learning activities), why (attitudes and beliefs), and the mechanisms helping and hindering educational change (strategic actions). Basing the survey around accepted concepts and definitions of AI in health care from published literature aided in clarifying priorities and avoided overlooking major aspects of AI education for the health workforce. Furthermore, the content of the survey itself may have served an informational, informal learning purpose for some of the people who started but did not complete it. By increasing their awareness of AI in health care, the survey project may have prompted some respondents to give further attention to this immature area of health workforce education and professional development. The design of the survey is such that it may be adapted and reused in other settings and could support longitudinal research over time as this aspect of workforce education and development matures.

Responses showed that the health sector is broadly in agreement but, on the whole, has not progressed far in plans to address this workforce need. In some quarters, it is possibly misdirected. For example, some responses conflated AI with health data analytics and digital health generally, which would dilute the deeper understanding of the implications of AI; some of the strategies proposed using AI-supported teaching techniques and learning analytics tools, which per se would not lead to a deeper understanding of AI-supported health care. Many varied opportunities and enablers of action were identified, suggesting optimism about the ability to make progress on this area of workforce education; the systemic challenges and barriers mentioned were fewer, although they presented substantial roadblocks.

The results also highlighted the priorities for education to address social and technical facets of AI, including ethical implications, suitability of large data sets for use in AI clinical applications, principles of machine learning, and specific diagnosis and treatment applications of AI as well as alterations to cognitive load during clinical work and the interaction between humans and machines in clinical settings. Although identification of priorities represents a first step, there are many activities such as capability building among educators and competency-based content development that are needed before implementation can occur at scale in health workforce education. The respondents also outlined barriers to implementation, which included a lack of governance structures and processes, resource constraints, and cultural adjustment. These are ubiquitous and represent major challenges; they are not specific to education in AI but are known to affect other areas, too, such as education to prepare the workforce for digital health generally [32].

Among the responses were almost no examples of educational resources or approaches that would build or benchmark competencies to work in interprofessional health care settings or in international contexts. Having noted this, resource constraints and development costs may be ameliorated if educators are aware of a number of web-based resources already in existence. The Australian eHealth Research Centre has compiled a series of real-world use cases of AI in health care for public information and education [33]. Around the world, a selective list of useful foundations for workforce education and development includes the Coursera AI course catalogue [34], the collection of readings on machine learning and AI on Medium [35], the AI Adventures playlist of Google Cloud Tech [36], a UK National Health Service and University of Manchester interprofessional course on AI for health care [37], and a certification examination and supporting learning and professional development overseen by a multidisciplinary advisory group of domain experts in the American Board of Artificial Intelligence in Medicine [38].

Limitations

Although the results reflected a cross-section of professions, organizations, and jurisdictions, they do not have statistical power. It was not possible to calculate a participation rate; overall, web-based access numbers met the minimum expectations of the potential reach of the survey, but completion numbers fell short. The survey required complex responses from a group of respondents known to be time-poor; therefore, survey fatigue likely accounted for some of the difference between the number who started and the number who completed all sections. Nevertheless, this was the most efficient method available to the researchers to begin a national interprofessional investigation, and it provides material that can be used in follow-up workshops for subgroups of those targeted in the survey.

Furthermore, the survey yielded valuable qualitative data; the free-text responses were thoughtful and extensive, and they provide a consolidated view of Australian educational experts' observations and aspirations regarding AI education for the health workforce. This study was the first of its kind and not only in Australia; however, it is only a first step in work toward education and professional development on AI that is delivered efficiently to the whole health workforce as well as tailored carefully for different roles and responsibilities within it.

Conclusions

This survey provides a baseline for further work by those responsible for enabling the health workforce as the optimization and ramifications of AI in health care unfold. These are early days in supporting the current and future health workforce to be able to work safely and effectively with AI, but the situation is evolving rapidly. There are calls for this work to proceed in partnership between education providers and AI technology providers to ensure that uniform training is available across health care subgroups and jurisdictions [39]. The methodology used to design and conduct this survey can be adapted for use in other health systems beyond Australia. From other areas of education on health informatics and digital health, we know that some topics and questions will be of global interest and concern, whereas other topics and questions will need to be customized to the distinctive social, political, and technical contexts of particular regional and national health care systems. Wider administration of surveys such as this one and detailed work to address the priority learning needs will assist educators and education authorities around the world who are responsible for preparing the health workforce to minimize the risks and realize the benefits of implementing AI in health care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey items.
[DOCX File , 24 KB - mededu_v8i2e35223_app1.docx]

References

1. Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Geneva: World Health Organization; 2021.
2. Sapci AH, Sapci HA. Artificial intelligence education and tools for medical and health informatics students: systematic review. JMIR Med Educ 2020 Jun 30;6(1):e19285 [FREE Full text] [doi: 10.2196/19285] [Medline: 32602844]
3. Carin L. On Artificial Intelligence and Deep Learning Within Medical Education. Acad Med 2020 Nov;95:10-11. [doi: 10.1097/ACM.0000000000003630] [Medline: 32769462]

4. Chan KS, Zary N. Applications and challenges of implementing artificial intelligence in medical education: integrative review. JMIR Med Educ 2019 Jun 15;5(1):e13930 [FREE Full text] [doi: 10.2196/13930] [Medline: 31199295]
5. Mazurowski MA. Artificial intelligence may cause a significant disruption to the radiology workforce. J Am Coll Radiol 2019 Aug;16(8):1077-1082. [doi: 10.1016/j.jacr.2019.01.026] [Medline: 30975611]
6. Miller DD. The medical AI insurgency: what physicians must know about data to practice with intelligent machines. NPJ Digit Med 2019;2:62 [FREE Full text] [doi: 10.1038/s41746-019-0138-5] [Medline: 31388566]
7. Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P. Introducing artificial intelligence training in medical education. JMIR Med Educ 2019 Dec 03;5(2):e16048 [FREE Full text] [doi: 10.2196/16048] [Medline: 31793895]
8. Park SH, Do K, Kim S, Park JH, Lim Y. What should medical students know about artificial intelligence in medicine? J Educ Eval Health Prof 2019 Jul 03;16:18 [FREE Full text] [doi: 10.3352/jeehp.2019.16.18] [Medline: 31319450]
9. Pucchio A, Eisenhauer EA, Moraes FY. Medical students need artificial intelligence and machine learning training. Nat Biotechnol 2021 Mar 10;39(3):388-389. [doi: 10.1038/s41587-021-00846-2] [Medline: 33692507]
10. Wartman SA, Combs CD. Medical education must move from the information age to the age of artificial intelligence. Acad Med 2018;93(8):1107-1109. [doi: 10.1097/acm.0000000000002044]
11. Skiba DJ. The invisible health care professional: exploring the intersection of data, devices, and artificial intelligence. Nurs Educ Perspect 2018;39(4):264-265. [doi: 10.1097/01.nep.0000000000000371]
12. Bhardwaj D. Artificial intelligence: patient care and health professional's education. J Clin Diagnostic Res 2019:3-5. [doi: 10.7860/jcdr/2019/38035.12453]
13. Hodges BD. Learning from Dorothy Vaughan: artificial intelligence and the health professions. Med Educ 2018 Jan 19;52(1):11-13. [doi: 10.1111/medu.13350] [Medline: 28722148]
14. Santos DP, Giese D, Brodehl S, Chon SH, Staab W, Kleinert R, et al. Medical students' attitude towards artificial intelligence: a multicentre survey. Eur Radiol 2019 Apr 6;29(4):1640-1646. [doi: 10.1007/s00330-018-5601-1] [Medline: 29980928]
15. Karaca O, Çalışkan SA, Demir K. Medical artificial intelligence readiness scale for medical students (MAIRS-MS) - development, validity and reliability study. BMC Med Educ 2021 Feb 18;21(1):112 [FREE Full text] [doi: 10.1186/s12909-021-02546-6] [Medline: 33602196]
16. Australian Alliance for Artificial Intelligence in Healthcare. URL: https://aihealthalliance.org [accessed 2022-03-15]
17. Australia's Health 2018. Australian Institute of Health and Welfare. 2018. URL: https://www.aihw.gov.au/reports/australias-health/australias-health-2018/contents/table-of-contents [accessed 2022-03-15]
18. Artificial intelligence in healthcare. Academy of Medical Royal Colleges. 2019. URL: https://www.aomrc.org.uk/reports-guidance/artificial-intelligence-in-healthcare/ [accessed 2022-03-15]
19. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun 13;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
20. Kapoor A. Deep learning vs machine learning: a simple explanation. Hackernoon. 2020. URL: https://hackernoon.com/deep-learning-vs-machine-learning-a-simple-explanation-47405b3eef08 [accessed 2022-03-15]
21. Natural language processing. Wikipedia. 2022. URL: https://en.wikipedia.org/wiki/Natural_language_processing [accessed 2022-03-15]
22. Abu-Nasser B. Medical expert systems survey. Int J Eng Inform Syst 2017;1(7):218-224 [FREE Full text]
23. Robot. Wikipedia. 2022. URL: https://en.wikipedia.org/wiki/Robot [accessed 2022-03-15]
24. Lawton G. Robotic process automation (RPA). TechTarget. 2021. URL: https://internetofthingsagenda.techtarget.com/definition/robotic-process-automation [accessed 2022-03-15]
25. Rowe J. Provider taps AI-powered app to improve medication adherence. AI Powered Healthcare. 2019. URL: https://www.healthcareitnews.com/ai-powered-healthcare/provider-taps-ai-powered-app-improve-medication-adherence [accessed 2022-03-15]
26. Beyond a biomedical search engine: AI-based discovery that matters. Quertle. URL: https://quertle.com/products/qinsight/ [accessed 2022-03-15]
27. Bresnick J. 5 steps for planning a healthcare artificial intelligence project. Health IT Analytics. 2018. URL: https://healthitanalytics.com/features/5-steps-for-planning-a-healthcare-artificial-intelligence-project [accessed 2022-03-15]
28. Chao G, Hurst E, Shockley R. The evolution of process automation: moving beyond basic robotics to intelligent interactions. IBM Institute for Business Value. 2018. URL: https://www.ibm.com/downloads/cas/QAQMRGVN [accessed 2022-03-15]
29. Quiroz JC, Laranjo L, Kocaballi AB, Berkovsky S, Rezazadegan D, Coiera E. Challenges of developing a digital scribe to reduce clinical documentation burden. NPJ Digit Med 2019;2:114 [FREE Full text] [doi: 10.1038/s41746-019-0190-1] [Medline: 31799422]
30. Chang A. Common misconceptions and future directions for ai in medicine: a physician-data scientist perspective. In: Artificial Intelligence in Medicine. Cham: Springer; 2019:3-6.
31. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. Lancet 2010 Dec;376(9756):1923-1958. [doi: 10.1016/s0140-6736(10)61854-5]

32. Nazeha N, Pavagadhi D, Kyaw BM, Car J, Jimenez G, Tudor Car L. A digitally competent health workforce: scoping review of educational frameworks. J Med Internet Res 2020 Nov 05;22(11):e22706 [FREE Full text] [doi: 10.2196/22706] [Medline: 33151152]
33. Australian eHealth Research Centre. URL: https://aehrc.com/ai/ [accessed 2022-03-15]
34. Coursera Artificial Intelligence Catalogue. URL: https://www.coursera.org/courses?query=artificial [accessed 2022-03-15]
35. Maini V. Machine learning for humans. Medium. 2017. URL: https://medium.com/machine-learning-for-humans/why-machine-learning-matters-6164faf1df12 [accessed 2022-03-15]
36. Google Cloud Tech's AI adventures. Youtube. URL: https://www.youtube.com/playlist?list=PLIivdWyY5sqJxnwJhe3etaK7utrBiPBQ2 [accessed 2022-03-15]
37. AI for healthcare: equipping the workforce for digital transformation. Futurelearn. URL: https://www.futurelearn.com/courses/artificial-intelligence-in-healthcare [accessed 2022-03-15]
38. American Board of Artificial Intelligence in Medicine. URL: https://abaim.org [accessed 2022-03-15]
39. A roadmap for artificial intelligence in healthcare for Australia. Australian Alliance for Artificial Intelligence in Healthcare. 2021. URL: https://aihealthalliance.org/wp-content/uploads/2021/12/AAAiH_Roadmap_1Dec2021_FINAL.pdf [accessed 2022-03-15]

Abbreviations

AHPRA: Australian Health Practitioner Regulation Agency
AI: artificial intelligence

Edited by N Zary; submitted 25.11.21; peer-reviewed by Y Li, A Arbabisarjou; comments to author 16.12.21; revised version received 04.02.22; accepted 05.03.22; published 04.04.22.

Please cite as:
Gray K, Slavotinek J, Dimaguila GL, Choo D
Artificial Intelligence Education for the Health Workforce: Expert Survey of Approaches and Needs
JMIR Med Educ 2022;8(2):e35223
URL: https://mededu.jmir.org/2022/2/e35223
doi:10.2196/35223
PMID:35249885

Original Paper

Current and Future Needs for Human Resources for Ethiopia's National Health Information System: Survey and Forecasting Study

Binyam Tilahun[1], MPH, MSc, PhD; Berhanu F Endehabtu[1], MPH; Kassahun D Gashu[1], MPH; Zeleke A Mekonnen[1,2], MPH; Netsanet Animut[3], MSc; Hiwot Belay[3], MSc; Wubshet Denboba[3], MPH; Hibret Alemu[3], PhD; Mesoud Mohammed[4], MSc; Biruk Abate[4], MPH

[1]Department of Health Informatics, Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia
[2]Health Systems Directorate, Ministry of Health, Addis Ababa, Ethiopia
[3]Data Use Partnership Project, John Snow, Inc (JSI), Addis Ababa, Ethiopia
[4]Policy, Planning, Monitoring and Evaluation Directorate, Ministry of Health, Addis Ababa, Ethiopia

Corresponding Author:
Berhanu F Endehabtu, MPH
Department of Health Informatics
Institute of Public Health, College of Medicine and Health Sciences
University of Gondar
PO Box 196
Gondar
Ethiopia
Phone: 251 921013129
Email: berhanufikadie@gmail.com

Abstract

Background: Strengthening the national health information system is one of Ethiopia's priority transformation agendas. A well-trained and competent workforce is the essential ingredient to a strong health information system. However, this workforce has neither been quantified nor characterized well, and there is no roadmap of required human resources to enhance the national health information system.

Objective: We aimed to determine the current state of the health information system workforce and to forecast the human resources needed for the health information system by 2030.

Methods: We conducted a survey to estimate the current number of individuals employed in the health information system unit and the turnover rate. Document review and key-informant interviews were used to collect current human resources and available health information system position data from 110 institutions, including the Ministry of Health, federal agencies, regional health bureaus, zonal health departments, district health offices, and health facilities. The Delphi technique was used to forecast human resources required for the health information system in the next ten years: 3 rounds of workshops with experts from the Ministry of Health, universities, agencies, and regional health bureaus were held. In the first expert meeting, we set criteria, which was followed by expert suggestions and feedback.

Results: As of April 2020, there were 10,344 health information system professionals working in the governmental health system. Nearly 95% (20/21) of district health offices and 86.7% (26/30) of health centers reported that the current number of health information system positions was inadequate. In the period from June 2015 to June 2019, health information technicians had high turnover (48/244, 19.7%) at all levels of the health system. In the next ten years, we estimate that 50,656 health information system professionals will be needed to effectively implement the Ethiopia's national health information system.

Conclusions: Current health information system–related staffing levels were found to be inadequate. To meet the estimated need of 50,656 multidisciplinary health information system professionals by 2030, the Ministry of Health and regional health bureaus, in collaboration with partners and academic institutions, need to work on retaining existing and training additional health information system professionals.

(JMIR Med Educ 2022;8(2):e28965) doi:10.2196/28965

KEYWORDS

forecasting; human resources; health information system; workforce; Ethiopia; health informatics; healthcare professionals

Introduction

A health information system integrates data generation, compilation, analysis, and synthesis, as well as the reporting and use of the information in evidence-based decision-making. The system can be a hybrid of both electronic and paper-based data collection from health institutions and other related sectors and ensures overall data quality. A strong health information system is central to achieving better health outcomes and is the backbone of a strong health system [1].

Strengthening the Ethiopia's health information system, so that it can produce quality and actionable health data, is critical for planning, mobilizing resources, establishing policies, monitoring health health-related activities and performance [2]. The Ministry of Health of Ethiopia has set strategies to establish a well-functioning health information system with the aim to improve evidence-based decision-making through enhanced partnership, harmonization, and alignment, including the integration of projects and programs at the point of health service delivery [3].

An effective health system is a function of multiple factors. Among the 6 building blocks of its framework for health systems, the World Health Organization includes *human resources* [4,5]. Human resources consume the most financial resources in the health sector; human resources also influence the efficiency and effectiveness of other health system building blocks. Although health sector–related human resources have been persistently neglected for years, currently, more attention is being paid to human resources which is considered, both by the global health community and national governments, to be a critical component in improving health outcome [4,6,7]. An adequate amount and mix of skilled human resources are essential at all levels of the health system. The level and mix of human resources required depend on the goals and objectives of a specific health system [6,8-10].

Globally, countries have been challenged by health care workforce shortages and uneven distribution of skills and professions [11]. The problem is worse in low- and middle-income countries. Ethiopia is one of the countries listed by the World Health Organization as having a health workforce crisis [12]. The implementation of a health information system in Ethiopia has been challenged with shortages of skilled human resources, a poor distribution of professions, and high attrition rates at all levels of the health system for activities that include the collection of reliable and complete information, data analysis and interpretation, and effective use of data for planning and decision-making [11,13].

Even though the country has set strategies to strengthen the health information system, the need for health information system human resources has not been quantified or characterized. Shortages, low skill, and uneven distributions in the workforce will be aggravated with time due to factors such as an expansion of health facilities, population growth, new technology in the health system, individual levels of performance, turnover rates, and health policy [14,15]. Higher education institutions in Ethiopia have shown interest in health information system workforce development; however, there is no national roadmap. Currently, some health science colleges and public universities have opened Diploma and Bachelor of Science Health Information System programs. In addition, some universities have started postgraduate-level Health Informatics programs.

Determining the performance level of the current workforce will help to identify the skills and knowledge required by those jobs and the human resources needed [16]—forecasting considers existing and future human resources–related demand and plans at all levels of the health system. Estimating future health information system human resources is also expected to help in addressing the issues of employing, educating, allocating, retaining, motivating, and managing the workforce, to ultimately improve the health information system.

Delphi is one technique used [17-19] for forecasting requirements. It is particularly helpful when the area of investigation does not allow for the use of analytical techniques but is suitable for expert opinion. The technique helps to facilitate a structured and systematic group communication process [19]. The Delphi technique allows efficient use of experts' diverse knowledge and experiences for the prediction of future human resources demand, despite limitations such as the inability to take into account complex factors that cannot always be predicted accurately by experts [18,20-22].

We aimed to determine the current status of the health information system workforce and forecast the human resources required for the health information system by 2030.

Methods

Ethics

The study protocol was reviewed and approved by ethical review board of the University of Gondar (number RCS/768/19). Informed consent was obtained from each study participant. A letter of permission was also obtained from each regional health bureau. The names of participants and other personal identifiers were not collected in the study.

Study Design

From June 2019 to April 2020, we conducted a national-level assessment of the health information system in Ethiopia in all regions and 2 city administrations to estimate the current amount of human resources and the turnover rate. We collected data from a total of 110 organizations: the Ministry of Health, 4 federal agencies, 11 regional health bureaus, 18 zonal health departments, 21 district health offices, 5 referral hospitals, 12 general hospitals, 8 primary hospitals, and 30 health centers. A total of 278 institution heads, planning department heads, and health management information system heads were interviewed.

We asked each facility and organization to collect and send information on required human resources to the ministry of health officials, and then, these data were used to forecast national human resources requirements. We use the Delphi technique—3 rounds of workshops were held with experts from the Ministry of Health, universities, agencies, and regional health bureaus.

Operational Definition

We defined *health information system professionals* as professionals in the fields *health informatics*, *health system*, *health management*, *computer science and information technology*, *monitoring and evaluation*, and *biostatistics* and other professionals who are directly involved in technical health information system activities such as planning, monitoring and evaluation, system analysis and design, and system administration.

Data Collection

A team (2 to 5 data collectors) was assigned in each region. Two days of training were provided to data collectors. Data collectors were health informatics, health service management, monitoring and evaluation, biostatistics, and health economics professionals. Health system experts from the Policy Plan Monitoring and Evaluation Directorate of the Ministry of Health participated as supervisors. Interview and document review methods were used to collect data.

Delphi Process

Experts and policy makers from the Ministry of Health and experts from universities, federal agencies, regional health bureaus, and other stakeholders were involved in each round. A total of 13 experts were involved in the forecasting process.

Data collected from the facilities and organizations were used to identify and prioritize major factors that could affect estimates for human resources for health information system. A total of 14 criteria were listed for rating, of which future expansion plans, the existing amount of human resources for health information system, the number of organizations, the number of standard health information system positions, and the turnover (or attrition) rate were selected based on ratings to be used for forecasting human resources requirements by 2030.

For the validation process, variables such as current number of health information system positions, currently available health information system workforce, and human resources required for health information system had no variation across facilities. These data did not differ across facilities and were found to be normally distributed (through visual assessment of the histograms).

Results

Current Human Resources

Overview

A total of 10,344 health information system workers were found to be actively working at different levels of the health system as of April 2020. Approximately half (5685/10,344, 55.0%) were diploma-level health information technicians. Staff had an average of 3 years of professional work experience. At the Ministry of Health, 24 employees worked in health information system–related positions (health science: 6/24, 25%; computer science: 6/24 (25%); health informatics: 4/24, 16.7%).

In regional health bureaus, there were 160 health information system–related professionals (health science: 48/160, 30%; computer science: 48/160, 30%; health informatics: 20/160, 12.5%). In all district health offices, a total of 45 workers were assigned to health information system–related activities, of which 19 (42%) were diploma-level health information technicians. Between April 2015 and April 2020, a total of 218 employees left their jobs; diploma-level health information technicians had the highest turnover rate (48/244, 19.7%) at the national level.

We found that health informatics, information technology, health information technician, environmental health, monitoring and evaluation, public health officer, health economics, statistics, Master of Public Health, business management, management, nursing, health officer and computer science graduates were currently working on health information system–related activities at each level of the health system.

Health Information System–Related Positions

The Ministry of Health had 13 health information system positions; however, it was reported that the current number of positions was not adequate, and it was suggested that the number be increased to 40. Federal agencies had an average of 14 positions. Regional health bureaus had an average of 6 positions, and 10 of 11 regional health bureaus reported not having an adequate number of positions; it was suggested that the number of positions be increased to 8 per region. The average number of health information system positions for zonal health department was 2. Fifteen zonal health offices and departments reported having an inadequate number of health information system positions; it was suggested that the number of positions be increased to an average of 4. District health offices had an average of 2 positions. Of those assessed, only 68.2% (30/44) had budgeted for health information system position; 95% (20/21) reported that current number of positions was not adequate and suggested that that the number of positions be increased to 5 per district. Health centers had an average of 1 health information system position; the majority (26/30, 86.7%) suggested that the number of positions be increased to 3 per health center. Three-quarters (6/8, 75%) of the primary hospitals reported not having an adequate number of positions and suggested having 5 positions per primary hospital. General hospitals had an average of 4 health information system positions, and the majority (9/12, 75%) reported that the number of positions was inadequate. Specialized or referral comprehensive hospitals had an average of 6 health information system positions, of which 80% (4/5) were budgeted. The majority (4/5, 80%) of comprehensive specialized or referral hospitals reported that the existing number of health information system positions was inadequate and suggested having 19 positions per specialized or referral hospital (Table S1 in Multimedia Appendix 1).

Estimated Human Resources Needed by 2030

Overview

The Delphi technique was implemented in 2 weeks, and 8 criteria were used to forecast human resources requirements by 2030: future expansion plan, current amount of human resources, current number of organizations, number of standard health information system positions, eHealth initiatives, strategic plan, budget for health, and turnover (or attrition) rate (Table S4 in Multimedia Appendix 1).

At the Federal Level (the Ministry of Health and its Agencies)

Based on the criteria prioritized above, it was estimated that 90 positions will be needed by the Ministry of Health (Master of Health Informatics: 27/90, 30%; Master of Monitoring and Evaluation: 22/90, 24.4%; Master of Information Technology or Computer Science: 21/90, 23.3%). Moreover, it was estimated that 105 positions will be needed for all federal agencies after considering all potential factors affecting human resources development—on average, each federal agency will require 15 employees to work on health information system–related activities—and the majority would need to have a Master of Health Informatics degree (28/105, 26.7%) (Table S7 and S8 in Multimedia Appendix 1).

At the Regional Level (Regional Health Bureaus and Regional Agencies)

Approximately 276 health information system positions will be needed at regional health bureaus, which translates to 25 per region. Most regional health bureaus would require professionals with a Master of Health Informatics degree (55/276, 20%), followed by those with Bachelor of Science degrees in Computer Science or Information Technology professionals (53/276, 19.2%). In all regional agencies, 63 positions will be required—regional agencies will require an average of 9 employees to work on health information system–related activities (Tables S9 and S10 in Multimedia Appendix 1).

Zonal Health Departments and District Health Office

During the survey, 172 health workers were working on health information system–related activities at the zonal health department level. The forecast estimate of human resources was 624 (Table S11 in Multimedia Appendix 1). During the survey in all district health offices, 1428 health workers were working on health information system related activities. A total of 3060 professionals will be needed in district health offices by 2030 (Master of Health Informatics: 1020/3060, 33%; Bachelor of Science in Health Informatics: 1020/3060, 33%; Bachelor of Science in Information Technology or Computer Science: 1020/3060, 33%) (Table S12 in Multimedia Appendix 1).

Primary, General, Comprehensive, and Specialized Hospitals

A total of 1419 professionals will be needed in comprehensive specialized hospitals by 2030. More than half (817/1419, 57.6%) will be diploma-level health information technicians (Table S13 in Multimedia Appendix 1). There are currently 79 general hospitals, and given the expansion plan, the number of general hospitals will increase to 150. For general hospitals, 2100 professionals will be needed, given that medical record units will be staffed with 7 health information technicians per hospital (Table S14 in Multimedia Appendix 1). There will be a total of 505 primary hospitals by 2030 (211 currently existing and 294 planned), which will require 4545 health information system professionals, with an average of 9 per hospital—5 diploma-level health information technicians e for the medical record unit, and the rest will be for data management—to implement telemedicine and harness information technology development (Table S15 in Multimedia Appendix 1).

Health Centers

Based on the assessment, there will be a total of 4586 health centers by the year 2030. There are currently 3591 health centers, and 995 will be constructed according to the expansion plan. A total of 27,516 professionals (6 per health center) will be required to support the implementation of electronic community health information systems at the health post level and eHealth initiatives in the health system (Table S16 in Multimedia Appendix 1).

Private Health Facilities

The health information system in Ethiopia has been engaging the private health facilities to follow the national health information system guidelines. Private health facilities could demand human resources for health information system in the future. At least 1 diploma-level health information technician for health information system and 1 diploma-level information technology employee for information communication technology support will be required at each private health facility (Table S17 in Multimedia Appendix 1).

Universities and Health Science Colleges

Universities educate and deploy the health professionals that work for the health system, and accordingly, need to have competent staff to train the workforce that will work on health information system–related activities. The experts suggested that 10 of the 55 universities will have health informatics departments by 2030, each requiring 48 professionals (Table S18 in Multimedia Appendix 1). Health science colleges regularly train health information technicians at the diploma level. It was estimated that 27 health science colleges, with 12 professionals each, will be required (Table S19 in Multimedia Appendix 1).

Total Estimated Workforce

A total of 50,656 of health information system employees will be needed by 2030 at the national level. More than half (27,516/50,656, 54.3%) of the workforce will be needed by health centers (Table S5 in Multimedia Appendix 1).

Discussion

The study has shown that the current number of health information system–related positions and the existing workforce were found to be inadequate. There was high turnover in the health information system–related workforce. It was forecast that 50,656 multidisciplinary health information system professionals will be needed nationally by 2030.

From the assessment, we found that approximately 10,344 employees were currently working on health information system–related activities and many district health offices (20/21, 95%) and health centers (26/30, 86.7%) reported that the number of health information system positions was inadequate. This finding indicates that current human resources are inadequate to achieve the national agenda. The level of skill mix was unbalanced. The majority of the health information system workforce currently working on health information system–related activities were diploma-level health information technicians. The correct amount and mix of the workforce is needed to establish well-organized health system [23,24].

We found that there were professionals from fields with no or little relation to health information system activities, such as environmental health, monitoring and evaluation, public health officer, health economics, statistics, Master of Public Health, business management, management, nurse, health officer, and computer science, employed in the health information system. Misplacement of employees in the workforce affects the effectiveness and efficiency of the organization [23]. Hence, every organization must have the right profession with the right knowledge and skill at the right place.

Our findings showed that there was high turnover in the health information system workforce; however, the reason for the turnover and its impact were not assessed in this study. Higher turnover is costly and associated with poor health service outcomes [25]. Health information system tasks, such as managing health care data and infrastructure, are complex and require experienced and well-trained professionals. Well-trained and experienced staff turnover is costly, mainly due to the need to train new employees, and compromises health system performance and the quality of services [8,25,26].

Educating and recruiting alone will not alleviate the workforce shortage. Assessment of the factors related to job satisfaction and retention is critical [27]. We found that diploma holders were the majority (48/244, 19.7%) of those who left their positions. This may be as a result of a lack of professional development opportunities. Until 2018, there was no academic program to upgrade their qualifications, and individuals with diplomas would have to upgrade by do a Bachelor of Science degree in nursing or other health science fields. The University of Gondar now has a post basic degree program in health informatics. There is evidence suggesting that improving the literacy of health workforce increases the retention rate [28]. Therefore, the Ministry of Health should design a strategy to provide scholarship opportunities and other job satisfaction schemes to minimize turnover.

The need for long-term forecasting of human resources for a health system in uncertainty and at the national level is a complex process—different variables must be considered in estimating the amount, skills, and professional mix of health personnel for meeting health system needs, which is information that policy and decision-makers need [29].

Different methods can be used to forecast human resources needs [29-31]. We used the Delphi method. The health information system workforce needed by the end of 2030 will be greater than 50,000. This indicates that the Ethiopia's shortage in health information system human resources is critical. Therefore, the Ministry of Health and the Ministry of Education should work together to ensure that there will be an adequate number of health information system graduates. There are some limitations to this study. First, in collecting data, we used the bottom-up managerial judgment approach and did not consider the final opinion of the bottom managers. Second, we did not collect all facility and administrative office managers' opinions; therefore, forecasts could be higher or lower than the actual need. Third, we assessed the magnitude of health information system staff turnover but we did not address the reasons.

This study has shown that the current health information system workforce is currently inadequate and will be inadequate for future needs. We estimated that approximately 50,656 human resources for health information system will be needed by 2030. Hence, the Ministry of Health and regional health bureaus, in collaboration with health information system and health system strengthening partners, need to train more health information system professionals and develop mechanisms to retain experienced professionals in the health system. Strong collaborations are also needed with universities to ensure the right skills are included within health information system curricula.

Acknowledgments

We acknowledge all local Capacity Building and Mentorship Program universities who supported the data collection in their catchment areas. We are also very grateful for those who participated in this national-level assessment. This work received funding from the Data Use Partnership Project.

Data Availability

The data sets are available upon request from the corresponding author.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary material.
[DOCX File , 89 KB - mededu_v8i2e28965_app1.docx]

References

1. Framework and standards for country health information systems. World Health Organization. 2008. URL: https://apps.who.int/iris/handle/10665/43872 [accessed 2021-02-12]
2. Health sector transformation plan II. Ministry of Health. 2020. URL: https://fp2030.org/sites/default/files/HSTP-II.pdf [accessed 2022-03-31]
3. Health sector transformation plan. Global Financing Facility. 2015. URL: https://www.globalfinancingfacility.org/sites/gff_new/files/Ethiopia-health-system-transformation-plan.pdf [accessed 2020-06-12]
4. Manyazewal T. Using the World Health Organization health system building blocks through survey of healthcare professionals to determine the performance of public healthcare facilities. Arch Public Health 2017;75:50 [FREE Full text] [doi: 10.1186/s13690-017-0221-9] [Medline: 29075485]
5. Monitoring the building blocks of health systems: a handbook of indicators and their measurement strategies. World Health Organization. 2010. URL: https://caribe.observatoriorh.org/sites/caribe.observatoriorh.org/files/webfiles/fulltext/WHO_MBHSS_2010.pdf [accessed 2021-02-16]
6. Workforce planning. Chartered Institute of Personnel and Development. 2018. URL: https://www.cipd.co.uk/Images/workforce-planning-guide_tcm18-42735.pdf [accessed 2021-02-15]
7. Girma S, Yohannes AG, Kitaw Y, Ye-Ebiyo Y, Seyoum A, Desta H, et al. Human resource development for health in Ethiopia: challenges of achieving the millennium development goals. Ethiopia J Health Dev 2007;21(3):2016-2231 [FREE Full text] [doi: 10.4314/ejhd.v21i3.10052]
8. Dussault G, Buchan J, Sermeus W, Padaiga Z. Assessing future health workforce needs? World Health Organization. 2010. URL: https://apps.who.int/iris/handle/10665/332024 [accessed 2021-02-15]
9. Global strategy on human resources for health: workforce 2030. World Health Organization. URL: https://apps.who.int/iris/bitstream/handle/10665/250368/9789241511131-eng.pdf [accessed 2022-03-31]
10. Zulu JM, Kinsman J, Michelo C, Hurtig A. Integrating national community-based health worker programmes into health systems: a systematic review identifying lessons learned from low-and middle-income countries. BMC Public Health 2014 Oct 22;14(1):987 [FREE Full text] [doi: 10.1186/1471-2458-14-987] [Medline: 25245825]
11. Ethiopia Ministry of Health. Developing multiple, integrated strategies to address Ethiopia's human resources for health needs. 2011 Presented at: Second Global Forum on Human Resources for Health; January 25 - 29, 2011; Bangkok, Thailand URL: https://www.who.int/workforcealliance/forum/2011/24_Ethiopia.pdf?ua=1
12. Campbell J, Settle D. Ethiopia: taking forward action on human resources for health (HRH) with DFID/OGAC and other partners. World Health Organization. 2009. URL: http://www.who.int/workforcealliance/knowledge/publications/partner/tfa_ethiopia.pdf [accessed 2020-06-15]
13. Assefa T, Haile Mariam D, Mekonnen W, Derbew M. Health system's response for physician workforce shortages and the upcoming crisis in Ethiopia: a grounded theory research. Hum Resour Health 2017 Dec 28;15(1):86 [FREE Full text] [doi: 10.1186/s12960-017-0257-5] [Medline: 29282069]
14. Health human resources planning and the production of health: development of an extended analytical framework for needs-based health human resources planning. HRH Global Resource Center. 2009. URL: https://www.hrhresourcecenter.org/node/1254.html [accessed 2020-06-15]
15. Szabo S, Nove A, Matthews Z, Bajracharya A, Dhillon I, Singh DR, et al. Health workforce demography: a framework to improve understanding of the health workforce and support achievement of the Sustainable Development Goals. Hum Resour Health 2020 Jan 29;18(1):7 [FREE Full text] [doi: 10.1186/s12960-020-0445-6] [Medline: 31996212]
16. Forecasting the demand for human resources. Jiwaji University. URL: http://www.jiwaji.edu/pdf/ecourse/management/forecasting [accessed 2021-02-09]
17. Abbasi M, Tabatabaeib SAN, Labbaf H. Identify future changes of ICT in human resources management: a Delphi study. Int Acad J Org Behav Human Res Manag 2016;3(1):36-44 [FREE Full text]
18. de Villiers MR, de Villiers PJ, Kent AP. The Delphi technique in health sciences education research. Med Teach 2005 Dec 03;27(7):639-643 [FREE Full text] [doi: 10.1080/13611260500069947] [Medline: 16332558]
19. Milkovich GT, Annoni AJ, Mahoney TA. The use of the Delphi procedures in manpower forecasting. Management Science 1972 Dec;19(4-part-1):381-388 [FREE Full text] [doi: 10.1287/mnsc.19.4.381]
20. Green RA. The Delphi technique in educational research. SAGE Open 2014 Apr 16;4(2):215824401452977 [FREE Full text] [doi: 10.1177/2158244014529773]
21. Linstone HA, Turoff M. The Delphi Method. Reading: Addison-Wesley Publishing Company; 1975.
22. Warner LA. Using the Delphi technique to achieve consensus: a tool for guiding extension programs. University of Florida. 2017. URL: https://edis.ifas.ufl.edu/pdf/WC/WC18300.pdf [accessed 2022-03-31]
23. Dubois C, Singh D. From staff-mix to skill-mix and beyond: towards a systemic approach to health workforce management. Hum Resour Health 2009 Dec 19;7:87 [FREE Full text] [doi: 10.1186/1478-4491-7-87] [Medline: 20021682]

24. Shiferaw AM, Zegeye DT, Assefa S, Yenit MK. Routine health information system utilization and factors associated thereof among health workers at government health institutions in East Gojjam Zone, Northwest Ethiopia. BMC Med Inform Decis Mak 2017 Aug 07;17(1):116 [FREE Full text] [doi: 10.1186/s12911-017-0509-2] [Medline: 28784115]
25. Zurn P, Vujicic M, Lemière C, Juquois M, Stormont L, Campbell J, et al. A technical framework for costing health workforce retention schemes in remote and rural areas. Hum Resour Health 2011 May 06;9(1):8 [FREE Full text] [doi: 10.1186/1478-4491-9-8] [Medline: 21470420]
26. Ishikawa T, Nakao Y, Fujiwara K, Suzuki T, Tsuji S, Ogasawara K. Forecasting maldistribution of human resources for healthcare and patients in Japan: a utilization-based approach. BMC Health Serv Res 2019 Sep 09;19(1):653 [FREE Full text] [doi: 10.1186/s12913-019-4470-x] [Medline: 31500619]
27. Marinucci F, Majigo M, Wattleworth M, Paterniti AD, Hossain MB, Redfield R. Factors affecting job satisfaction and retention of medical laboratory professionals in seven countries of sub-Saharan Africa. Hum Resour Health 2013 Aug 17;11(1):38 [FREE Full text] [doi: 10.1186/1478-4491-11-38] [Medline: 23958152]
28. Martiniuk A, Colbran R, Ramsden R, Karlson D, O'Callaghan E, Lowe E, et al. Hypothesis: improving literacy about health workforce will improve rural health workforce recruitment, retention and capability. Hum Resour Health 2019 Dec 30;17(1):105 [FREE Full text] [doi: 10.1186/s12960-019-0442-9] [Medline: 31888671]
29. Dreesch N, Dolea C, Dal Poz MR, Goubarev A, Adams O, Aregawi M, et al. An approach to estimating human resource requirements to achieve the millennium development goals. Health Policy Plan 2005 Oct;20(5):267-276 [FREE Full text] [doi: 10.1093/heapol/czi036] [Medline: 16076934]
30. Anderson GF, Han KC, Miller RH, Johns ME. A comparison of three methods for estimating the requirements for medical specialists: the case of otolaryngologists. Health Serv Res 1997 Jul;32(2):139-153 [FREE Full text] [Medline: 9180613]
31. Barber P, López-Valcárcel BG. Forecasting the need for medical specialists in Spain: application of a system dynamics model. Hum Resour Health 2010 Oct 29;8:24-4491 [FREE Full text] [doi: 10.1186/1478-4491-8-24] [Medline: 21034458]

Edited by N Zary; submitted 20.03.21; peer-reviewed by I Mircheva, C Gibson, S Hewko, A Rovetta, S Nagavally; comments to author 28.06.21; revised version received 17.08.21; accepted 16.01.22; published 12.04.22.

Please cite as:
Tilahun B, Endehabtu BF, Gashu KD, Mekonnen ZA, Animut N, Belay H, Denboba W, Alemu H, Mohammed M, Abate B
Current and Future Needs for Human Resources for Ethiopia's National Health Information System: Survey and Forecasting Study
JMIR Med Educ 2022;8(2):e28965
URL: https://mededu.jmir.org/2022/2/e28965
doi:10.2196/28965
PMID:35412469

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org