frontiers | Frontiers in Virtual Reality

TYPE Review
PUBLISHED 04 July 2024
DOI 10.3389/frvir.2024.1402093

OPEN ACCESS

EDITED BY
Tor Finseth,
Honeywell, United States

REVIEWED BY
Philipp Klimant,
Hochschule Mittweida, Germany
Bethany McCurdy,
Iowa State University, United States

*CORRESPONDENCE
Casey J. Clay,
casey.clay@usu.edu

RECEIVED 16 March 2024
ACCEPTED 30 May 2024
PUBLISHED 04 July 2024

CITATION
Clay CJ, Budde JR, Hoang AQ and Gushchyan A (2024), An evaluation of the effectiveness of immersive virtual reality training in non-specialized medical procedures for caregivers and students: a brief literature review.
Front. Virtual Real. 5:1402093.
doi: 10.3389/frvir.2024.1402093

An evaluation of the effectiveness of immersive virtual reality training in non-specialized medical procedures for caregivers and students: a brief literature review

Casey J. Clay[1]*, John R. Budde[2], Andrea Q. Hoang[1] and Armen Gushchyan[1]

[1]Department of Special Education and Rehabilitation Counseling, Utah State University, Logan, UT, United States, [2]Children's Hospital of Orange County, Orange, CA, United States

As technological advancements continue to redefine the landscape of adult education and training, virtual reality (VR) has emerged as a potent tool for enhancing skill acquisition. This literature review synthesizes existing research on the utilization of immersive and discrete VR in training adults in the medical domain. The primary focus was on understanding the effectiveness, challenges, and potential applications of VR-based training programs. Specifically, we reviewed studies related to practical skills critical to safety that target a non-surgical, discrete medical procedure (e.g., diabetes care procedures, how to correctly set a G-tube, CPR, correct personal protective equipment [PPE] usage) using an immersive VR technology as a training modality. Further, the studies reviewed had to include a comparison of immersive VR training to that of a business-as-usual (BAU) method. We conducted a review of the six studies that met the criteria and coded variables related to what technology was used, targeted skills being trained, social validity, effectiveness of the intervention, and whether generalization occurred. Key themes explored in the literature include the role of immersive experiences in enhancing learning outcomes when comparing VR training to BAU and the adaptability of VR platforms to different skill sets. Special attention was given to identifying factors that contribute to the success or limitation of VR-based training initiatives, including individual differences, technology acceptance, and effectiveness.

KEYWORDS

medical training, skills training, training, virtual reality, virtual reality training, immersive virtual reality, virtual training

Introduction

In the ever-evolving landscape of adult education and training, the integration of virtual reality (VR) has emerged as a revolutionary paradigm, offering immersive and interactive experiences that transcend traditional learning methods. In this article, we will review and summarize utilized hardware tools, technical skill acquisition, and technical skill generalization with healthy adult individuals in the context of VR; with a specific focus on medical settings and occupations where these skills are of paramount importance for ensuring safety and optimal performance. The scope of this review is brief and narrow to focus on recent studies that involve fully immersive VR training experiences. As the technology and training applications are rapidly advancing, it is critical to examine the comparison of these new procedures to business-as-usual (BAU) analogue training.

Background

The evolution of virtual reality in adult training in medical settings

The advent of VR has revolutionized adult learning, providing realistic and contextually rich environments for skill development. VR technologies offer a unique platform for adults to engage in experiential learning, fostering the acquisition of both cognitive and psychomotor skills (Xie et al., 2021). The medical field, characterized by the complexity of tasks and the critical nature of decisions, stands out as a promising domain for the application of VR in skill acquisition. Medical professionals are routinely confronted with high-stakes situations that demand a sophisticated set of psychological skills, including decision-making under pressure, effective communication, teamwork, and stress management (Taylor et al., 2017; Schut and Driessen, 2019). The ability to acquire, develop, and maintain these skills is not only crucial for individual wellbeing but is also directly linked to patient safety and the overall quality of healthcare delivery.

Virtual reality offers medical professionals the opportunity to engage in realistic scenarios that mirror the complexities of their daily practice. Simulated environments allow for repeated practice in a controlled setting, fostering the development of psychological skills without compromising patient safety. Research by Cook et al. (2011) and Ziv et al. (2006) demonstrates the efficacy of VR in enhancing clinical skills and decision-making among medical practitioners.

The dynamic nature of VR technology suggests an exciting future for its application in medical training. Continued advancements, including the integration of augmented reality and artificial intelligence, hold promise for creating even more realistic and adaptive learning environments tailored to individual needs. For example, the need may vary from training complex skills to training non-specialized skills for adults in the medical profession.

Theoretical framework

Adult learning theories and virtual reality

Understanding the mechanisms underlying adult learning is essential for designing effective VR-based training programs. A common approach to adult learning is Pedagogy of Practices which involves practice-based learning experiences, which are essential to developing fluency (Grossman et al., 2009a; Grossman et al., 2009b). This approach is heavily used in the area of training individuals on how to teach and train others (i.e., schoolteachers, clinical psychologists, and clergy). Practice-based experiences can extend on-the-job experience through simulation (Dalinger et al., 2020). It may be the case that virtual simulations are most impactful when they recreate endogenous training scenarios in settings with minimal risk (Grassini et al., 2020; Renganayagalu et al., 2021). The structure of the training scenarios should ensure evidence-based pedagogical practice methodologies (Grossman et al., 2019), various representations within a professional context (Danielson and Matson, 2018), and opportunities for learner engagement in real world situations (Oprean and Balakrishnan, 2020). Clear structure reduces complexity, prepares the trainee for success, and ensures specific feedback can be built into training (DeGraff et al., 2015; Schutz et al., 2018; Pastore and Andrde, 2019; Richmond et al., 2019). These training considerations can be viewed as universal and should be taken into account when training any type of human services provider.

Skills, including decision-making, communication, and stress management, are integral components of successful adult learning and in particular are put to use in medical practice. Models such as the Deliberate Practice framework by Ericsson (2008) and the Five Stage Model of Skill Acquisition by Dreyfus and Dreyfus (1980) provide insights into the stages of skill development and the role of immersive experiences in achieving expertise. Both of these models emphasize the mastery of skills through concrete experience.

Another approach to adult learning is based on the principles of andragogy, as opposed to pedagogy, proposed by Malcolm Knowles, this approach attempt to depart from studies of learning in children instead emphasize self-directed learning and practical application (Knowles, 2014). Adult learning processes focused on practical application and the need for self-directed learning might, therefore, align seamlessly with the immersive and experiential nature of VR, which allows for practical use cases and control by the individual to learn at their own pace. Additionally, situated learning theories, such as those advanced by Lave and Wenger (1996), support the idea that learning is most effective when situated in authentic contexts, a principle well-aligned with the capabilities of VR technology. Further, rather than asking about cognitive processes and conceptual structures, they ask what kinds of social engagement provided the context for learning to take place (Lave and Wenger, 1996).

VR allows for a methodology for training based on adult learning theories. These situated learning theories support the idea of learning in authentic contexts, which align with VR technology's immersive nature. Virtual simulations can be theory-based and impactful when recreating authentic scenarios with minimal risk.

Transferability to real-world settings

An essential criterion for the success of VR-based training is the transferability of acquired skills to real-world situations. The transferability of trained skills across contexts has been referred to as generalization (Ducharme and Feldman, 1992; Gianoumis and Sturmey, 2012). These measures are critical given the nature of VR training, which necessitates testing of performance maintenance outside of the virtual environment. Research conducted by Levac et al. (2019) and Kim et al. (2019) provides valuable insights into the extent to which skills developed in virtual environments translate to improved performance in clinical practice. Conclusions reached in previous research reveal skills generally transferred from therapeutic practice to the real world.

Generalization is also referred to as knowledge or skill transfer following virtual training (Gasteiger et al., 2022). Multiple reviews provide evidence that skills generalize after VR training (Alaker et al., 2016; Abich et al., 2021; Gastieger et al., 2022). However, most of the reviews included highly specialized skills training (e.g., surgery, aviation) that involve several complex behaviors. A further summary of the research is needed involving discrete non-specialized skills that also have a knowledge component to aid in skill performance in the medical domain. Additionally, reporting on examples of research comparing a control (i.e., business-as-usual) group is needed to establish the validity of VR training procedures for non-specialized skills.

Challenges and considerations

Individual differences and acceptance

While the potential benefits of VR in skill development are evident, individual differences in technology acceptance and proficiency can influence the effectiveness of training programs. The work of Cook et al. (2013) and Cowling and Birt (2018) sheds light on the importance of considering individual characteristics in the design and implementation of virtual and immersive VR-based interventions. Critical to the acceptance of these technologies is user satisfaction or social validity of the utilization of VR for training purposes, which takes into account individual characteristics of the design.

There has been an increase in the number of studies that involve the use of immersive virtual reality (VR) in training methods for trainees across a wide range of skills and industries. However, majority of the studies in the healthcare industry focus on highly specialized skills (e.g., surgery, and disease detection). Examination of research focusing on less specialized skills (e.g., donning/doffing protective equipment exists, resuscitation, etc.) but it is unclear how effective interventions may be given training to different populations of less complex skills. Within the healthcare industry, the use of VR technology in training procedures is becoming more and more common as is presenting data that suggest trainee satisfaction and acquisition of skills. However, recent studies suggest there is a lack of evidence for directly evaluating the extent to which trainees undergoing VR training acquire skills and their ability to generalize those skills to situations and settings outside of the training environment (Levac et al., 2019; Grassini et al., 2020; Clay et al., 2021). Therefore, the purpose of this literature review is to summarize the research on VR-based training to teach non-specialized medical or safety-related skills, identify the effectiveness of the training in achieving target outcomes, and provide directions for future application and research.

Materials and methods

Researchers followed PRISMA checklist guidelines (Page et al., 2021) when conducting this review. Researchers first began by using ProQuest to search for articles on immersive virtual reality training. The search was conducted in November 2022 and included all published articles up until that date. The ProQuest database itself included the following databases in all search attempts conducted: APA PsycINFO, Coronavirus Research Database, Education Collection, Psychology Database, PTSDpubs, and Publicly Available Content Database. Researchers selected ProQuest as it was made available through professional organizations to which researchers belonged, as well as that it was likely to include virtual reality training studies that may have been conducted during the pandemic due to the focus on virtual training during that time. Researchers used the ProQuest search feature and searched the terms "immersive virtual reality," "medical training procedure," and "learning outcome." Researchers limited the results to "peer-reviewed" articles only and the source type to "scholarly journals" only. This search retrieved 1,556 results. From these results, the second author initially screened titles and article abstracts that met all of the following criteria: (a) involved the use of virtual reality (VR)-based training to teach a skill; (b) involved a medical or safety-related training; (c) involved participants who worked or studied in a healthcare-related field; (d) was published in a peer-reviewed journal; (e) was published in English. Any articles that did not meet all the previously mentioned criteria were not included for further review.

This initial screening produced a total of 44 included articles, which were then screened once again to determine which articles met the final inclusion criteria: (a) used immersive virtual reality (IVR) that included a headset and hand controllers; (b) involved the training of non-surgical healthcare-related procedures only; (c) included participants who were medical students, practitioners, emergency responders or life support safety instructors; (d) compared the performance between participants who underwent VR-based training and participants who underwent traditional BAU training. As a result, a total of six articles were included in this review (see Figure 1).

Interobserver agreement (IOA)

Interobserver agreement (IOA) efforts were conducted for the final inclusion criterion screening and the coding of relevant variables. Specifically, two observers apart from the second author screened 22 of the 44 articles (50%) that met the initial criterion to determine which articles met the final inclusion criteria. A second observer also conducted IOA regarding the coding of relevant variables for three out of the six articles (50%) included in this review. IOA was calculated by dividing agreements by agreements plus disagreements and multiplying by 100. For the final eligibility screening, the IOA between observers was 98%. For coding of relevant variables across observers, IOA was 100%.

Results

Participants and setting

Out of the six studies reviewed, all participants were identified to be either students, instructors, or practitioners in some healthcare-related capacity (see Table 1). A cumulative total of 319 participants were identified within the studies selected. In all, participants were identified as 148 medical students within a university-based medical program (46%) (Birrenbach et al., 2021; Han et al., 2021; Kravitz

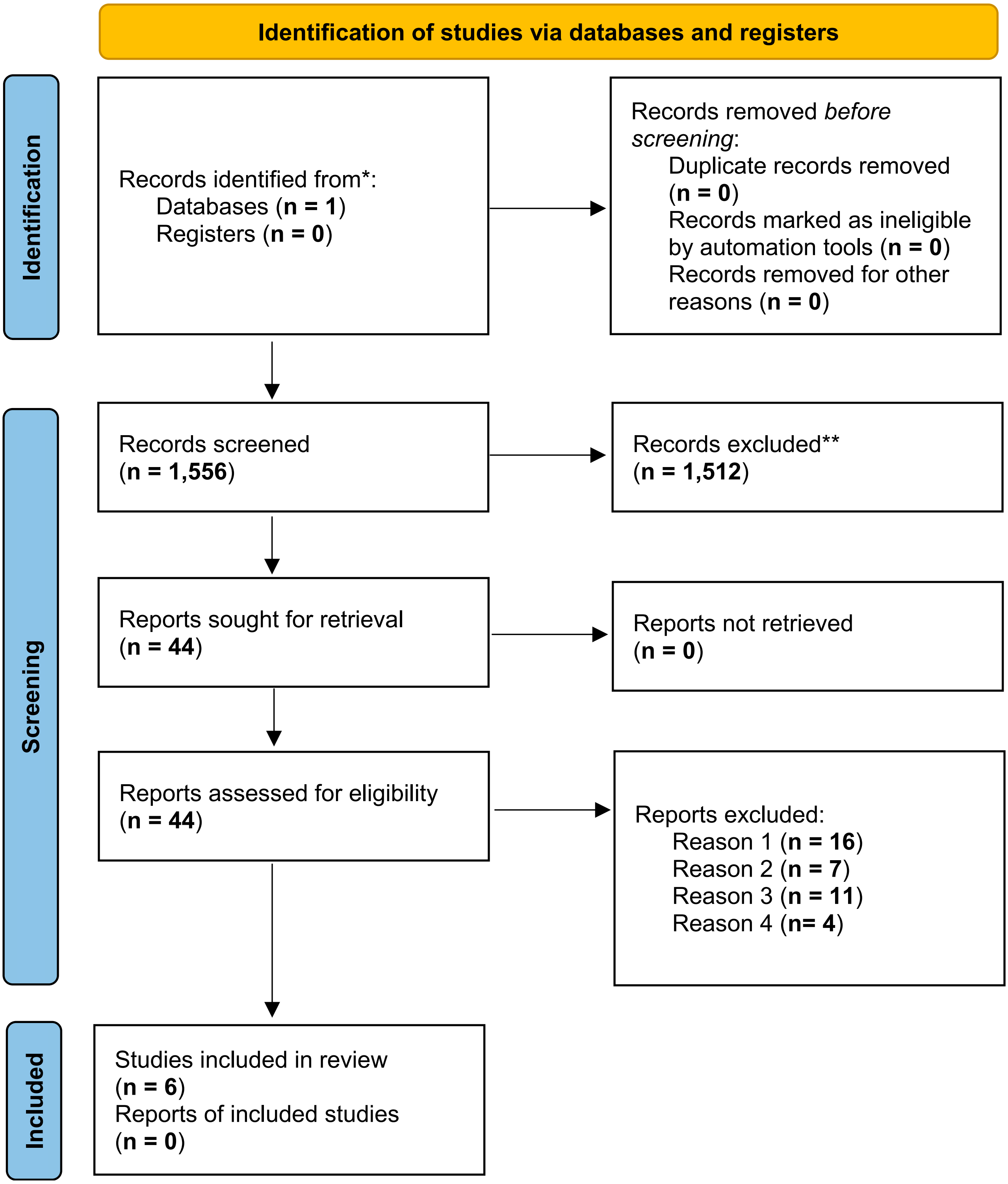

FIGURE 1
PRISMA identification process flowchart. Note: Reason 1: Did not use immersive virtual reality training (trained with a headset); Reason 2: Participants were not students or caregivers of patients; Reason 3: Did not compare VR training group with a control group/traditionally trained group; Reason 4: Did not train for non-surgical medical procedures. **These records were excluded from further review due to not meeting all of the following criteria: (a) involved the use of VR-based training to teach a skill; (b) involved a medical or safety-related training; (c) involved participants who worked or studied in a healthcare-related field; (d) was published in a peer-reviewed journal; (e) was published in English.

10.3389/frvir.2024.1402093

TABLE 1 Summary of demographics and setting.

Citation	Participants	Setting
Birrenbach et al. (2021)	Medical students (years 3–6 out of a 6-year curriculum) at the University of Bern	Emergency department of the Inselspital University Hospital, Bern, Switzerland
	(N = 29; intervention VR training, n = 15, vs. control video-based instruction, n = 14)	
Han et al. (2021)	N = 95. Senior year medical students from Yonsei University College of Medicine in Seoul, South Korea. CG: 39, EG: 56)	Virtual reality training classrooms within the teaching hospital in 2019
Kiyozumi et al. (2022)	A total of 13 instructor candidates (students) who participated in the VR course and 10 students in face-to-face groups (N = 23, CG = 10, EG = 13)	3 laboratories on the campus of the National Defense Medical College in Tokorozawa, Japan, other meeting rooms between December 2021 and February 2022
Kravitz et al. (2022)	Fifty-four participants were randomized, mostly consisting of medical students (n = 24 {44%}) or emergency medicine and otolaryngology residents (n = 19 {35%}). CG = 27, EG = 27	Study training classrooms at the Montefiore Medical Center, Wakefield Campus, Bronx, United States.
Sapkaroski et al. (2022)	A split-cohort study was performed with trainee practitioners (n = 70) and qualified practitioners (n = 9). Participants were randomly assigned to four groups: clinician VR (CVR), clinician role-play (CRP), trainee VR (TVR), and trainee RP (TRP)	Study training classrooms at the School of Medicine, Nursing and health Sciences, at Monash University in Clayton Victoria, Australia between 2020–2021
Yu et al. (2021)	Senior nursing students were divided into an experimental group (n = 25) experiencing virtual reality simulation and routine neonatal intensive care unit practice and a control group (n = 25) having routine neonatal intensive care unit practice	At training classrooms at the College of Nursing, Institute of Health Sciences, Gyeongsang National University, Jinju, Republic of Korea between 2019 and 2020

et al., 2022), 23 advanced resuscitation training instructor candidates (7%) (Kiyozumi et al., 2022), 19 emergency medicine and otolaryngology residents (6%) (Kravitz et al., 2022), 79 radiology student practitioners (25%) (Sapkaroski et al., 2022), and 50 nursing students (16%) (Yu et al., 2021).

There were multiple training settings found within the articles. One study utilized the emergency department of an inpatient hospital unit (17%) (Birrenbach et al., 2021); one other study conducted training within a research laboratory facility (17%) (Kiyozumi et al., 2022), and four of the studies utilized open-space classrooms located on a university campus for training purposes (66%) (Han et al., 2021; Yu et al., 2021; Kravitz et al., 2022; Sapkaroski et al., 2022).

Technology Used

All studies included articles that utilized immersive VR equipment and software programming for training purposes (see Table 2). In terms of the specific VR headset and hand controller equipment, five studies utilized an Oculus Rift headset and hand controller (83%) (Birrenbach et al., 2021; Han et al., 2021; Kiyozumi et al., 2022; Sapkaroski et al., 2022) while one study utilized an HTC Vive Pro Full-Kit Head Mounted Display (HMD) with an Ultraleap Leap Motion hand controller (17%) (Yu et al., 2021). There was a high degree of variation in software programming methods used for training purposes, in part due to the variation of the training curriculum among included articles.

Dependent and independent variables

Across all included studies, the dependent variable consisted of an evaluation of the extent to which trainees correctly implemented a discrete, non-surgical, medical, and/or safety-related procedure. The independent variable consisted of a training program (e.g., VR training or BAU training) designed to teach participants to correctly implement the procedure(s) of focus. There were a range of procedures that participants were trained on within the studies reviewed. Specific skills taught to participants included hygiene-safety-related procedures such as hand disinfection, nasopharyngeal swab taking, and the donning and doffing of personal protective equipment (PPE) (Birrenbach et al., 2021; Kravitz et al., 2022). Skills taught to participants also included patient safety and evaluation procedures such as Immediate Cardiac Life Support (ICLS) training (Kiyozumi et al., 2022), provider-patient MRI-related communication skills (Sapkaroski et al., 2022), and conducting the Neurologic Physical Exam (NPE) (Han et al., 2021).

Effectiveness of intervention

In each of the articles reviewed, there were measures taken to evaluate participant performance across multiple phases of training. All six studies involved dividing participants into either experimental groups (EG) or control groups (CG). The experimental group consisted of participants undergoing VR-based training while the control group involved traditional or BAU training for the skill of focus. Four of the studies reviewed required participants to undergo both pre-and post-test evaluations of procedural implementation (67%) (Birrenbach et al., 2021; Yu et al., 2021; Kravtiz et al., 2022; Sapkaroski et al., 2022) while two of the studies did not conduct a pre-training evaluation of participant performance prior to receiving training (33%) (Han et al., 2021; Kiyozumi et al., 2022). Across the four studies in which pre-and post-test evaluations of participant performance were conducted, it was

TABLE 2 Summary of coded variables of study details.

Citation	Technology used	Targeted skills to train	Effectiveness of intervention	Control/Test group comparison
Birrenbach et al. (2021)	COVID-19 VR Strikes Back (CVRSB) module, version 1.1.6), a software platform developed by ORamaVR SA, and the Oculus Rift S head mounted device and hand controllers	1) hand disinfection 2) nasopharyngeal swab taking 3) donning/doffing of PPE	Both groups perform significantly better after training, with the effect sustained over 1 month	Hand Disinfection No significant difference in performance between groups across BL, post-test 1/2 Nasopharyngeal Swab Acquisition: the VR group performed better; scored a median of 14 out of 17 points (IQR 13–15) versus 12 out of 17 points (IQR 11–14) in the control group, $p = 0.03$ Contamination During Doffing No significant difference between the number of contaminated areas during doffing was found between the groups at both time points. Increase in performance from post-test 1 to post-test 2 across both groups
Han et al. (2021)	VRNET, a VR program that performs neurological tests using Oculus Rift (Oculus VR, SF, United States)	Conducting the Neurologic Physical Exam (NPE) with a confederate patient	Reported effective in teaching students how to conduct and correctly implement the NPE (no pre-test)	There were no statistical differences in VRNET's realness and student satisfaction between the EG and CG groups. However, a statistically significant difference was found in the Neurologic Physical Exam (NPE) score ($p = 0.043$); the EG group had higher NPE scores (3.81 ± 0.92) than the CG group (3.40 ± 1.01)
Kiyozumi et al. (2022)	NEUTRANS (Synamon Inc.) virtual workshop allows users to interact as avatars in a virtual space by using a head-mounted display. Instructor trainers and instructor candidates (students) entered the virtual space using the Oculus Quest 2 (Meta Platforms Inc.) or Oculus Lift (Meta Platforms Inc.)	Advanced resuscitation training. Specifically, immediate cardiac life support (ICLS) course—an ALS training course that was approved by the Japanese Association of Acute Medicine (JAAM)	Reported effective in teaching students how to conduct and correctly implement the ICLS procedure (no pre-test though)	The overall evaluation scores for the VR and face-to-face groups did not differ at the level of statistical significance (median 3.8, IQR 3.8–4.0 and median 4.2, IQR 3.9–4.2, respectively; $p = 0.41$
Kravitz et al. (2022)	PPE training on the Oculus Quest (Menlo Park, CA: Facebook, Inc.) using a PPE training program created by Axonpark, Inc. (Fort Lauderdale, FL)	Donning and doffing performance (each measured separately) after VR and e-module PPE training among medical staff and medical students at a single institution	Both groups performed significantly better after training	The VR group ($n = 27$ {50%}) performed better than the control in the overall PPE scores but this was not statistically significant (mean {SD}, VR: 55.4 {4.4} vs. emodule: 53.3 {8.1}; $p = 0.40$). VR participants also reported higher levels of preparedness and confidence after training
Sapkaroski et al. (2022)	VR Simulation Learning Environment (CETSOL). Using the Oculus Rift	The use (selection) of empathic language for a specific task to coach anxious and claustrophobic patients in MRI. Self-report questionnaire and a guided communication scenario tool with embedded scoring developed by a panel of experts	All groups except the CRP group reported a significant improvement in self-assessed communication scores following training: 11% for TVR ($p < 0.05$), 4.3% for TRP ($p < 0.05$), and 7.2% for CVR ($p < 0.05$)	Both VR training groups (TVR and CVR) performed better on average than their role-play counterparts (5% and 11%); however, the results were only statistically significant for the trainees: $p < 0.05$. Results of the SE-12 communication questionnaire showed that the intragroup pre-training scores were not significantly different amongst either the trainees or clinicians. However, post-training, the TVR group's ability was perceived to be 6.7% greater than that of the TRP group, and the CVR group's ability was perceived to be 7.2% greater than CRP group's ability
Yu et al. (2021)	High-risk neonatal infection control (HirNIC) VR simulation program employed in this study was developed	The HirNICCS_K consists of five subdomains: basic care, skin care, feeding management, medication	Both groups perform significantly better after training	No significant difference between the experimental and control groups (U = 272.00, $p = .213$) in

(Continued on following page)

TABLE 2 (*Continued*) Summary of coded variables of study details.

Citation	Technology used	Targeted skills to train	Effectiveness of intervention	Control/Test group comparison
	by Yu and Mann. A Vive Pro Full-Kit Head Mounted Display and sensor (HTC VIVETM, United States), a Leap Motion Controller™ (Ultraleap, United States) hand-tracking device with a VR Developer Mount, and a VR kit containing an EliteDesk 800 G4 laptop computer	and invasive procedure management, and environmental management according to the scenario topics		terms of HirNIC knowledge. Based on the pretest and posttest results, HirNIC performance self-efficacy significantly increased in both the experimental (t = 10.03, $p < .001$) and control (t = 7.48, $p < 0.001$) groups. The experimental group showed a greater self-efficacy increase than the control group, indicating that the VR program was effective in improving self-efficacy (t = 2.16, $p = 0.018$). In the domains of basic care (t = 2.73, $p = 0.005$) and skin care and environmental management (t = 2.28, $p = 0.013$), the experimental group had significantly higher self-efficacy scores than the control group. The experimental group also showed a higher score in the feeding management domain, but the difference from the control group was not significant (t = 1.28, p = 0.103)

TABLE 3 Summary social validity and generalization.

Citation	Social validity	Generalization
Birrenbach et al. (2021)	Yes, on user satisfaction with their respective training method (USEQ survey). The satisfaction of participants in the VR group measured by the USEQ was significantly higher than that of the participants in the control group	Yes, All PTs underwent a post-test outside of training program
Han et al. (2021)	Yes, participant surveys of VRNET's realness and student satisfaction with training overall. Here were no statistical differences found in realness (SP group 4.27 ± 0.75, SP with VRNET group 4.28 ± 0.56, $p = 0.92$) and satisfaction (SP group 4.23 ± 0.71, SP with VRNET group 4.21 ± 0.66, p = 0.839) of students	Yes, All PTs underwent a post-test outside of training program
Kiyozumi et al. (2022)	Yes, all respondents required to complete a satisfaction survey, needing to provide a score of 1 or higher on the 5-point Likert scale survey. All respondents (14/14, 100%) were satisfied with the VR course, providing a score of 4 or higher on the 5-point Likert scale. However, many respondents indicated that the teaching of specific discrete skills was not suitable for VR training, whereas the domains of performance evaluation and feedback were deemed suitable for VR training	No, no generalization probes were conducted (i.e., no evaluation of skills learned took place outside of the virtual setting for participants within the EG)
Kravitz et al. (2022)	Yes, all respondents required to complete a satisfaction survey, needing to provide a score of 1 or higher on the 5-point Likert scale survey. Perceived preparedness, log odds of reporting a lower score of 3 compared to 4 or 5 is 1.08 points lower in VR than e-module ($p = 0.05$). Perceived confidence of retention, the log odds of reporting a lower score of 3 compared to 4 or 5 is 1.55 points lower in VR than for e-module ($p = 0.007$). Distraction, E-module group reported being distracted compared to VR (59% vs. 48%, $p = 0.58$, not statistically significant)	No, no generalization on probes were conducted (i.e., no evaluation of skills learned took place outside of the virtual setting for participants within the EG)
Sapkaroski et al. (2022)	Yes, participants completed a pre/post self-efficacy questionnaire (SE-12). All groups except the CRP group reported a significant improvement in self-assessed communication scores following training	No, no generalization on probes were conducted (i.e., no evaluation of skills learned took place outside of the virtual setting for participants within the EG)
Yu et al. (2021)	Yes, three item questionnaire on learner satisfaction. Each item was rated on a five-point Likert scale ranging from "very unsatisfied" (1 point) to "very satisfied" (5 point). The experimental group had significantly higher scores than the control group	No, no generalization on probes were conducted (i.e., no evaluation of skills learned took place outside of the virtual setting for participants within the EG)

Note. CRP, clinician role-play; EG, experimental group; PT, participant; SE, self-efficacy; SP, standardized patient; USEQ, user satisfaction evaluation questionnaire.

reported that participants in both the CG and the EG performed significantly better in post-training assessment in comparison to the pre-training assessments conducted. However, it should be noted that for one of these four studies that reported overall improvement among participants after training completion, there was not a statistically significant improvement from pre- to post-test performance for the clinician BAU group (Sapkaroski et al., 2022). Although these four studies generally recorded statistically significant improvements in performance following training, only three of these studies reported no statistically significant difference in performance levels between both the CG and EG, although the EG scores were identified to have been higher than that of the CG, respectively (Birrenbach et al., 2021; Yu et al., 2021; Kiyozumi et al., 2022; Sapkaroski et al., 2022). One out of the four studies demonstrated that participants in the VR training group performed better on post-test evaluations than that of the BAU group with a statistically significant difference indicated (Sapkaroski et al., 2022). However, Sapkaroski et al. (2022) did identify a statistically significant difference between VR and BAU trainee sub-groups regarding communication questionnaire scores.

Out of the two studies that did not include a pre-test assessment of participant performance, one of these studies found that participants in the EG group recorded statistically higher written assessment scores than the CG (50%) (Han et al., 2021), while the other study demonstrated no statistically significant difference in level of performance between the CG and EG (50%) but did report higher self-efficacy scores for participants within the EG in comparison participants within the CG (Kiyozumi et al., 2022).

Generalization

Some of the articles reviewed included the purpose of evaluating the degree to which the generalization of skills learned by participants within their respective training programs could be demonstrated or not (Table 3). Out of all six articles reviewed, two of them documented an evaluation of skills learned by participants outside of the training setting (33%) which also acted as a post-test evaluation of performance (Birrenbach et al., 2021; Han et al., 2021). Four studies reported there not being any measures or attempts to evaluate the generalization of skills learned by participants (67%) (Yu et al., 2021; Kiyozumi et al., 2022; Kravitz et al., 2022; Sapkaroski et al., 2022). Across all studies included, only one study conducted a maintenance probe to assess the extent to which skills learned in the training program by participants maintained over time (17%) (Birrenbach et al., 2021), while all other studies did not include any attempts of evaluate participant performance after an extended period had elapsed following training completion (83%).

Social validity

All six studies included recorded measures related to social validity and user satisfaction among participants (100%) (Table 3). A range of user satisfaction assessments was documented, involving the evaluation of trainee satisfaction via the User Satisfaction Evaluation Questionnaire (USEQ) and other close-ended, five-point Likert scale surveys (Yu et al., 2021; Kiyozumi et al., 2022; Kravtiz et al., 2022). In general, all studies reported high satisfaction with the VR training program and in some cases higher satisfaction in the experimental group than in the control. For example, Yu et al. (2021) found the experimental group's satisfaction with the VR program was much higher than the learner satisfaction exhibited by the control group. The learners in the experimental group were more likely to score highly "I want to recommend to others," and described the program as "fun."

Discussion

As the field of VR-based training matures, the development of evidence-based guidelines becomes imperative. Drawing on insights from research, this paper reviews general practices in the design, implementation, and assessment of VR programs aimed at enhancing non-specialized skills in medical settings where the skills are critical to safety. The summary of this topic is useful as it provides a baseline of understanding of training outcomes for non-specialized skills, speaking to the generality and effectiveness of VR training when considering the participants engaging in this type of training.

Interestingly, we found several studies reported no statistically significant differences between trainees who underwent VR training versus BAU. While this may appear to be a result of failed effects, in essence, it speaks to the interchangeability and substitution of BAU for VR in terms of effectiveness. It may then follow that other variables (e.g., cost savings, trainee satisfaction) may impact the choice of one method over the other (Young and Greenberg, 2013; Bumbach et al., 2022). As a case in point, a study by Farra et al. (2019) found overall cost savings for a hospital when they compared VR versus BAU training on safety drills for hospital workers. The researchers found initially that VR is more expensive for the live drill (BAU) versus for VR training. When development costs are extrapolated to repeated training over 3 years, however, the virtual exercise becomes less expensive, while the cost of live exercises remains fixed. The authors concluded the larger initial investment in virtual reality can be spread across many trainees and a longer time with little additional cost.

Of note, pedagogy and specific learning theories were seldom mentioned, if at all, in the articles reviewed. Though not recognized in all fields, human service fields frequently incorporate the use of hands-on simulation training with opportunities for the trainee to rehearse and practice. As an evidence-based example, behavioral skills training (BST) is frequently used to build skills for individuals delivering behavior assessment and intervention strategies/techniques for children with disabilities (Shea et al., 2020; Schaefer et al., 2021; Smith et al., 2022). BST entails four separate phases of training: instructions, modeling, rehearsal, and feedback. The first phase involves didactic instruction which seeks to explain the how and why of the processes that underlie the strategy being taught.

Though not as important as the modeling or the rehearsal and feedback phases, an explanation of procedures satisfies an element of social validity within teaching a new skill. Next, the expert models the skill for the trainee. This can be done *in-vivo*, through a confederate, by viewing pre-recorded video, or in a virtual space. The third step involves rehearsal with the trainee in a role-play format. During the process the final step is executed, feedback. The expert gives direct feedback to the trainee on what the trainee did correctly and how their performance could be improved. This process is typically repeated until the trainee demonstrates skill mastery.

Similar to BST, training in the medical field involving simulation involves the technique of rapid cycle deliberate practice (RCDP). In RCDP learners rapidly cycle between deliberate practice and directed feedback within the simulation scenario until mastery is achieved (Taras and Everett, 2017). RCDP has been shown to be effective in teaching a spectrum of resuscitation skills, such as airway management skills (e.g., Gross et al., 2016), pediatric resuscitation (Hunt et al., 2014), and resuscitation for cardiac arrest (e.g., Kutzin and Janicke, 2015). Future research should further investigate VR models of BST (e.g., Clay et al., 2021) and RCDP.

Ethical considerations surrounding the use of VR in medical training warrant careful examination. Issues such as informed consent, psychological wellbeing, and potential desensitization to critical situations necessitate a thoughtful and ethical approach to the development and implementation of VR programs. Recent studies by Zechner et al. (2023), Gasteiger et al. (2022), and Slater et al. (2020) contribute valuable perspectives on ethical considerations in the use of VR for psychological skill development. As Zechner et al. (2023) mention, for police officers to be able to properly make decisions in high-risk scenarios, they need to feel a certain amount of stress. Stressful situations could be visually simulated within AR/VR, and as technology improves these simulations will only become more immersive as resolution increases in each eye of the headset (Zhan et al., 2020) and haptic feedback (van Wegen et al., 2023) becomes easier to program into the software through ancillary hardware. This stress is necessary to elicit since most closely resembles the state of the participant in the natural environment and would, theoretically, greatly help with the generalization component of some training. Clearly, this also comes at a potential price as practitioners must delicately balance the introduction of stressful situations as they correspond to the expected application of the strategies by the trainees. If not carefully controlled, undue stress would be presented to a trainee. This evaluation of modulated stressful situations is not present within some trainings in VR/AR literature, and it deserves a closer look.

Related to the potentially aversive experiences in VR, positive experiences reported via social validity measures were found in all studies we reviewed. Of note, all studies reported high satisfaction with the VR training program. Part of the experience of VR training is how immersive the experience is and may impact how a user reports positive experiences in the simulation. Data on the degree of immersion as it relates to positive or negative social experiences would be valuable for future researchers to collect.

Some limitations of this current review include a narrow range in the scope of the variables identified, reviewed, and discussed. We specifically attempted to provide a review of VR programs for discrete skills that may be used in medical settings but excluded specialized skills (e.g., surgery), dynamic skills (e.g., teamwork communication; cf. Bracq et al., 2019), and disease diagnosis as multiple previous reviews of VR skills training exist for these skills. Another limitation related to scope of the review and recent research that has since been published. We conducted the review in late 2022 and it's likely more studies have been published since then, which creates an excellent opportunity for future research. Another limitation is that we did not combine and calculate the overall effect of the treatments for comparisons between studies analyzed as might be the case in larger meta-analyses. Future researchers may attempt effect size calculations when summarizing VR training research for a more complete analysis of the extent to which VR training contributes to the acquisition of technical skills. A third limitation is that we only used one database (i.e., ProQuest). For a more thorough and systematic review, researchers should include comparison with at least one other database.

Conclusion

In summary, our review provides information on VR non-specialized training to practitioners and researchers to inform evidence-based practice and future research directions. Importantly, we provided measures of efficacy that involved comparison to business-as-usual or control training, while also including descriptions of technology that was used as well as identification of social validity and generalization. Despite efficacy being significant for all studies reviewed we would suggest taking caution in interpretation as only one-third of the studies reviewed reported generalization measures. Future directions include assessing both IVR and augmented reality as training tools as well as comparisons of these formats in skills training. We would encourage future researchers to include some measure of generalization outside the virtual environment to increase the validity of training. Future researchers might also investigate which variables lead to the generalization of the skill outside the virtual environment to increase the validity of the training.

Author's note

The authors would like to recognize and thank Jaylene Canales and Kayli Savage for their contributions of assisting in terms of IOA data collection on this project.

Author contributions

CC: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft, Writing–review and editing.

JB: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Writing–original draft, Writing–review and editing. AH: Formal Analysis, Validation, Visualization, Writing–original draft, Writing–review and editing. AG: Data curation, Formal Analysis, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Funding for publication of this article was contributed in part by the Utah State University Library Open Access Fund.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abich, J., Parker, J., Murphy, J. S., and Eudy, M. (2021). A review of the evidence for training effectiveness with virtual reality technology. *Virtual Real.* 25 (4), 919–933. doi:10.1007/s10055-020-00498-8

Alaker, M., Wynn, G. R., and Arulampalam, T. (2016). Virtual reality training in laparoscopic surgery: a systematic review & meta-analysis. *Int. J. Surg.* 29, 85–94. doi:10.1016/j.ijsu.2016.03.034

Birrenbach, T., Zbinden, J., Papagiannakis, G., Exadaktylos, A. K., Muller, M., Hautz, W. E., et al. (2021). Effectiveness and utility of virtual reality simulation as an educational tool for safe performance of COVID-19 diagnostics: prospective, randomized pilot trial. *JMIR Serious Games* 9, e29586. doi:10.2196/29586

Bracq, M. S., Michinov, E., and Jannin, P. (2019). Virtual reality simulation in nontechnical skills training for healthcare professionals: a systematic review. *Simul. Healthc.* 14 (3), 188–194. doi:10.1097/sih.0000000000000347

Bumbach, M. D., Culross, B. A., and Datta, S. K. (2022). Assessing the financial sustainability of high-fidelity and virtual reality simulation for nursing education: a retrospective case analysis sustainability of high-fidelity and virtual reality simulation for nursing education: a retrospective case analysis. *Cin. Comput. Inf. Nurs.* 40(9), 615–623. doi:10.1097/CIN.0000000000000916

Clay, C. J., Schmitz, B. A., Balakrishnan, B., Hopfenblatt, J. P., Evans, A., and Kahng, S. (2021). Feasibility of virtual reality behavior skills training for preservice clinicians. *J. Appl. Behav. Analysis* 54 (2), 547–565. doi:10.1002/jaba.809

Cook, D. A., Hamstra, S. J., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., et al. (2013). Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. *Med. Teach.* 35 (1), e867–e898. doi:10.3109/0142159x.2012.714886

Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., et al. (2011). Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *J. Am. Med. Assoc.* 306 (9), 978–988. doi:10.1001/jama.2011.1234

Cowling, M., and Birt, J. (2018). Pedagogy before technology: a design-based research approach to enhancing skills development in paramedic science using mixed reality. *Multidiscip. Digit. Publ. Inst.* 9 (2), 29. doi:10.3390/info9020029

Dalinger, T., Thomas, K. B., Stansberry, S., and Xiu, Y. (2020). A mixed reality simulation offers strategic practice for pre-service teachers. *Comput. Educ.* 144, 103696. doi:10.1016/j.compedu.2019.103696

Danielson, K., and Matson, C. (2018). Teaching teachers: designing an *NGSS* learning pathway. *Sci. Child.* 55 (7), 69–74. doi:10.2505/4/sc18_055_07_69

DeGraff, T. L., Schmidt, C. M., and Waddell, J. H. (2015). Field-based teacher education in literacy: preparing teachers in real classroom contexts. *Teach. Educ.* 26 (4), 366–382. doi:10.1080/10476210.2015.1034677

Dreyfus, S. E., and Dreyfus, H. L. (1980). A five-stage model of the mental activities involved in directed skill acquisition. *Distribution* 22. doi:10.21236/ada084551

Ducharme, J. M., and Feldman, M. A. (1992). Comparison of staff training strategies to promote generalized teaching skills. *J. Appl. Behav. Analysis* 25 (1), 165–179. doi:10.1901/jaba.1992.25-165

Ericsson, A. K. (2008). Deliberate practice and acquisition of expert performance: a general overview. *Acad. Emerg. Med.* 15 (11), 988–994. doi:10.1111/j.1553-2712.2008.00227.x

Farra, S. L., Gneuhs, M., Hodgson, E., Kawosa, B., Miller, E. T., Simon, A., et al. (2019). Comparative cost of virtual reality training and live exercises for training hospital workers for evacuation. *Comput. Inf. Nurs. Cin.* 37 (9), 446–454. doi:10.1097/CIN.0000000000000540

Gasteiger, N., van der Veer, S. N., Wilson, P., and Dowding, D. (2022). How, for whom, and in which contexts or conditions augmented and virtual reality training works in upskilling health care workers: realist synthesis. *JMIR Serious Games* 10 (1), e31644. doi:10.2196/31644

Gianoumis, S., and Sturmey, P. (2012). Generalization procedures in training interventionists for individuals with developmental disabilities. *Behav. Modif.* 36 (5), 619–629. doi:10.1177/0145445511432920

Grassini, S., Laumann, K., and Rasmussen Skogstad, M. (2020). The use of virtual reality alone does not promote training performance (but sense of presence does). *Front. Psychol.* 11, 1743. doi:10.3389/fpsyg.2020.01743

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., and Williamson, P. W. (2009a). Teaching practice: a cross-professional perspective. *Teach. Coll. Rec.* 111 (9), 2055–2100. doi:10.1177/016146810911100905

Grossman, P., Dean, C. G. P., Kavanagh, S. S., and Herrmann, Z. (2019). Preparing teachers for project-based teaching. *Phi Delta Kappan* 100 (7), 43–48. doi:10.1177/0031721719841338

Grossman, P., Hammerness, K., and McDonald, M. (2009b). Redefining teaching, re-imagining teacher education. *Teach. Teach. Theory Pract.* 15 (2), 273–289. doi:10.1080/13540600902875340

Han, S. G., Kim, Y. D., Kong, T. Y., and Cho, J. (2021). Virtual reality-based neurological examination teaching tool (VRNET) versus standardized patient in teaching neurological examinations for the medical students: a randomized, single-blind study. *BMC Med. Educ.* 21 (1), 493. doi:10.1186/s12909-021-02920-4

Kim, A., Schweighofer, N., and Finley, J. M. (2019). Locomotor skill acquisition in virtual reality shows sustained transfer to the real world. *J. Neuroengineering Rehabilitation* 16 (1), 1–10. doi:10.1186/s12984-019-0584-y

Kiyozumi, T., Ishigami, N., Tatsushima, D., Araki, Y., Yoshimura, Y., and Saitoh, D. (2022). Instructor development workshops for advanced life support training courses held in a fully virtual space: observational study. *JMIR Serious Games* 10 (2), e38952. doi:10.2196/38952

Knowles, M. (2014). "Andragogy: an emerging technology for adult learning," in *Education for adults* (New York, NY: Routledge), 53–70.

Kravitz, M. B., Dadario, N. B., Arif, A., Bellido, S., Arif, A., Ahmed, O., et al. (2022). The comparative effectiveness of virtual reality versus e-module on the training of donning and doffing personal protective equipment: a randomized, simulation-based educational study. *Cureus* 14 (3), e23655. doi:10.7759/cureus.23655

Kutzin, J. M., and Janicke, P. (2015). Incorporating rapid cycle deliberate practice into nursing staff continuing professional development. *J. Continuing Educ. Nurs. Thorofare* 46 (7), 299–301. doi:10.3928/00220124-20150619-14

Lave, J., and Wenger, E. (1996). "Communities of practice," in *The encyclopedia of informal education*. Available at: www.infed.org/biblio/communities_of_practice.htm.

Levac, D. E., Huber, M. E., and Sternad, D. (2019). Learning and transfer of complex motor skills in virtual reality: a perspective review. *J. Neuroengineering Rehabilitation* 16, 1–15. doi:10.1186/s12984-019-0587-8

Oprean, D., and Balakrishnan, B. (2020). From engagement to user experience: a theoretical perspective towards immersive learning. *Learn. User Exp. Res.* Available at: https://edtechbooks.org/ux/10_from_engagement_.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ Clin. Res. Ed.* 372, n71. doi:10.1136/bmj.n71

Pastore, S., and Andrade, H. L. (2019). Teacher assessment literacy: a three-dimensional model. *Teach. Teach. Educ.* 84, 128–138. doi:10.1016/j.tate.2019.05.003

Renganayagalu, S. k., Mallam, S. C., and Nazir, S. (2021). Effectiveness of VR head mounted displays in professional training: a systematic review. *Tech. Know Learn* 26, 999–1041. doi:10.1007/s10758-020-09489-9

Richmond, G., Salazar, M. D. C., and Jones, N. (2019). Assessment and the future of teacher education. *J. Teach. Educ.* 70 (2), 86–89. doi:10.1177/0022487118824331

Sapkaroski, D., Mundy, M., and Dimmock, M. R. (2022). Immersive virtual reality simulated learning environment versus role-play for empathic clinical communication training. *J. Med. Radiat. Sci.* 69 (1), 56–65. doi:10.1002/jmrs.555

Schaefer, J. M., and Andzik, N. R. (2021). Evaluating behavioral skills training as an evidence-based practice when training parents to intervene with their children. *Behav. Modif.* 45 (6), 887–910. doi:10.1177/0145445520923996

Schut, S., and Driessen, E. (2019). Setting decision-making criteria: is medical education ready for shared decision making? *Med. Educ.* 53 (4), 324–326. doi:10.1111/medu.13826

Schutz, P. A., Nichols, S. L., and Schwenke, S. (2018). “Critical events, emotional episodes, and teacher attributions in the development of teacher identities,” in *Research on teacher identity: mapping challenges and innovations*, 49–60. doi:10.1007/978-3-319-93836-3_5

Shea, K. A., Sellers, T. P., Smith, S. G., and Bullock, A. J. (2020). Self-guided behavioral skills training: a public health approach to promoting nurturing care environments. *J. Appl. Behav. analysis* 53 (4), 1889–1903. doi:10.1002/jaba.769

Slater, M., Gonzalez-Liencres, C., Haggard, P., Vinkers, C., Gregory-Clarke, R., Jelley, S., et al. (2020). The ethics of realism in virtual and augmented reality. *Front. Virtual Real.* 1, 1. doi:10.3389/frvir.2020.00001

Smith, S. G., Mattson, S. L., Aguilar, J., Pyle, N., and Higbee, T. S. (2022). Behavioral skills training with adult interventionists: a systematic review. *Rev. J. Autism Dev. Disord.* 11, 296–319. doi:10.1007/s40489-022-00339-1

Taras, J., and Everett, T. (2017). Rapid cycle deliberate practice in medical education—a systematic review. *Cureus* 9 (4), e1180. doi:10.7759/cureus.1180

Taylor, L. J., Nabozny, M. J., Steffens, N. M., Tucholka, J. L., Brasel, K. J., Johnson, S. K., et al. (2017). A framework to improve surgeon communication in high-stakes surgical decisions: best case/worst case. *JAMA Surg.* 152 (6), 531–538. doi:10.1001/jamasurg.2016.5674

van Wegen, M., Herder, J. L., Adelsberger, R., Pastore-Wapp, M., van Wegen, E. E. H., Bohlhalter, S., et al. (2023). An overview of wearable haptic technologies and their performance in virtual object exploration. *Sensors Basel, Switz.* 23 (3), 1563. doi:10.3390/s23031563

Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F. D., et al. (2021). A review on virtual reality skill training applications. *Front. Virtual Real.* 2, 645153. doi:10.3389/frvir.2021.645153

Young, B. K., and Greenberg, P. B. (2013). Is virtual reality training for resident cataract surgeons cost effective? *Graefe’s Archive Clin. Exp. Ophthalmol.* 251, 2295–2296. doi:10.1007/s00417-013-2317-4

Yu, M., Yang, M., Ku, B., and Mann, J. S. (2021). Effects of virtual reality simulation program regarding high-risk neonatal infection control on nursing students. *Asian Nurs. Res.* 15 (3), 189–196. doi:10.1016/j.anr.2021.03.002

Zechner, O., Kleygrewe, L., Jaspaert, E., Schrom-Feiertag, H., Hutter, R. V., and Tscheligi, M. (2023). Enhancing operational police training in high stress situations with virtual reality: experiences, tools and guidelines. *Multimodal Technol. Interact.* 7 (2), 14. doi:10.3390/mti7020014

Zhan, T., Yin, K., Xiong, J., He, Z., and Wu, S. T. (2020). Augmented reality and virtual reality displays: perspectives and challenges. *iScience* 23 (8), 101397. doi:10.1016/j.isci.2020.101397

Ziv, A., Erez, D., Munz, Y., Vardi, A., Barsuk, D., Levine, I., et al. (2006). The Israel Center for Medical Simulation: a paradigm for cultural change in medical education. *Acad. Med.* 81 (12), 1091–1097. doi:10.1097/01.ACM.0000246756.55626.1b