



HMD-Based Virtual and Augmented Reality in Medical Education: A Systematic Review

Xuanhui Xu*, Eleni Mangina and Abraham G. Campbell

School of Computer Science, University College Dublin, Dublin, Ireland

Background: Virtual Reality (VR) and Augmented Reality (AR) technologies provide a novel experiential learning environment that can revolutionize medical education. These technologies have limitless potential as they provide in effect an infinite number of anatomical models to aid in foundational medical education. The 3D teaching models used within these environments are generated from medical data such as magnetic resonance imaging (MRI) or computed tomography (CT), which can be dissected and regenerated without limitations.

Methods: A systematic review was carried out for existing articles until February 11, 2020, in EMBASE, PubMed, Scopus, ProQuest, Cochrane Reviews, CNKI, and OneSearch (University College Dublin Library) using the following search terms: (Virtual Reality OR Augmented Reality OR mixed reality) AND [“head-mounted” OR “face-mounted” OR “helmet-mounted” OR “head-worn” OR oculus OR vive OR HTC OR hololens OR “smart glasses” OR headset AND (training OR teaching OR education)] AND (anatomy OR anatomical OR medicine OR medical OR clinic OR clinical OR surgery OR surgeon OR surgical) AND (trial OR experiment OR study OR randomized OR randomised OR controlled OR control) NOT (rehabilitation OR recovery OR treatment) NOT (“systematic review” OR “review of literature” OR “literature review”). PRISMA guidelines were adhered to in reporting the results. All studies that examined people who are or were medical-related (novel or expert users) were included.

Result: The electronic searches generated a total of 1,241 studies. After removing duplicates, 848 remained. Of those, 801 studies were excluded because the studies did not meet the criteria after reviewing the abstract. The full text of the remaining 47 studies was reviewed. After applying inclusion criteria and exclusion criteria, a total of 17 studies (1,050 participants) were identified for inclusion in the review.

Conclusion: The systematic review provides the current state of the art on head-mounted device applications in medical education. Moreover, the study discusses trends toward the future and directions for further research in head-mounted VR and AR for medical education.

Keywords: virtual reality, augmented reality, head-mounted display, surgery, medicine, systematic review, education

OPEN ACCESS

Edited by:

Philip Breedon,
Nottingham Trent University,
United Kingdom

Reviewed by:

Deborah Richards,
Macquarie University, Australia
Hongchuan Yu,
Bournemouth University,
United Kingdom

*Correspondence:

Xuanhui Xu
xuanhui.xu@ucdconnect.ie

Specialty section:

This article was submitted to
Virtual Reality in Medicine,
a section of the journal
Frontiers in Virtual Reality

Received: 07 April 2021

Accepted: 04 June 2021

Published: 06 July 2021

Citation:

Xu X, Mangina E and Campbell AG
(2021) HMD-Based Virtual and
Augmented Reality in Medical
Education: A Systematic Review.
Front. Virtual Real. 2:692103.
doi: 10.3389/frvir.2021.692103

1 INTRODUCTION

Retaining knowledge in education is challenging. Medical students learn complex structures and anatomy mainly based on teaching material such as books or pictures traditionally, and for some educational institutions with more resources, students may have a chance to dissect actual cadavers. However, the paper-based learning material might cause misunderstandings as it is hard to imagine the 3D relationship between components based on 2D materials. Teaching resources such as real-life cadavers are limited and critically have strict storage restrictions based on health and safety rules.

Therefore, Virtual Reality (VR) and Augmented Reality (AR) interventions based on simulations could offer a possible solution or, at the very least, ease this bottleneck in medical education. They would improve spatial awareness when compared to 2D teaching materials and provide infinite teaching materials that can be the foundation for advancing the accessibility of content for medical education. VR and AR technologies provide a close-to-reality experience for users in industry, education, and gaming. Among all VR/AR formats, the head-mounted display (HMD) provides the most immersive environment, tracking a user's motion and maintains the position of spatial information around them. The 3D teaching models can be generated from medical data like magnetic resonance imaging (MRI) or computed tomography (CT), which allow them to be dissected and regenerated without any limits.

Simulators in general are widely used in medical education and assessment. To date, several systematic reviews have investigated the efficacy of VR simulation training in laparoscopy (Larsen et al., 2012; Alaker et al., 2016). The results showed that VR laparoscopy simulation provided an effective and ethical way to train residents' surgical skills. VR simulation can play an important role in addressing the issue of low training efficiency. However, the main VR simulators used as interventions were LapSim^{®1} and Simendo^{®2}, which did not utilize HMDs. Based on the current authors' knowledge, another systematic review in the usage of HMD-based VR and AR in medical education does not exist. Previous studies have identified some situations where HMDs are suitable for skill acquisition (Jensen and Konradsen, 2018), including cognitive skills related to remembering and understanding spatial information and knowledge. As learning a kinesthetic-based medical skill highly relies on spatial cognition, the immersion provided by an HMD logically then becomes a natural requirement for this review, to explore if VR and AR may potentially benefit medical skill acquisition.

This study focuses on a systematic review to evaluate the effectiveness of applying HMD for VR or AR applications in medical education that can benefit medical training. To this end, the systematic review will answer the following questions that are proposed in the protocol (Section 1.1):

- Compared with the standard teaching method or other types of simulators, what are the comparative effectiveness of HMD VR or AR usage in medical teaching? (Advantages)
- What are the disadvantages of using HMD VR or AR, and which one has lower side effects? (Disadvantages)
- Is there a definitive advantage of HMD VR and AR when used for increasing the efficiency of teaching in medicine? (Proof)
- Do HMD VR and AR have the potential to be support tools for medical education? (Support)

1.1 Protocol

A systematic review was carried out until February 11, 2020. PRISMA guidelines were adhered to in reporting the results of this study (Moher et al., 2009). Methods of the analysis and inclusion criteria were specified in advance and documented in a protocol. The protocol has been registered in PROSPERO, the international prospective register of systematic reviews, where it can be accessed (Registration number: CRD42020165310)³.

1.2 Search Strategy

The literature search and initial screening were conducted by XX; abstract screening was conducted independently by three authors (XX, EM, and AC), while the disagreement was confirmed by discussion; full article screening and data extraction were conducted by XX. Databases searched were EMBASE, PubMed, Scopus, ProQuest, Cochrane Reviews, CNKI, and OneSearch (University College Dublin Library) on title, abstract, and keywords; searches from Google Scholar are also acceptable. The terms used for searching were as follows: (Virtual Reality OR Augmented Reality OR mixed reality) AND [“head-mounted” OR “face-mounted” OR “helmet-mounted” OR “head-worn” OR oculus OR vive OR HTC OR hololens OR “smart glasses” OR headset AND (training OR teaching OR education)] AND (anatomy OR anatomical OR medicine OR medical OR clinic OR clinical OR surgery OR surgeon OR surgical) AND (trial OR experiment OR study OR randomized OR randomised OR controlled OR control) NOT (rehabilitation OR recovery OR treatment) NOT (“systematic review” OR “review of literature” OR “literature review”).

1.3 Selection Criteria

All studies examining the general adult human population or healthy adult humans and people who are or were medical-related (novel or experts) were included. Studies in which individuals were selected with extreme motion sickness, other diagnosed illness, or disability and studies in which individuals were not medical-related are excluded. No year publication limits were set. English and Chinese text publications were included as

¹<https://surgicalscience.com/systems/lapsim/>

²<https://www.simendo.eu/>

³Xu, X, Abraham, G. C., and Eleni, M. Can the head-mounted device improve medical education quality? Protocol for a systematic review. PROSPERO 2020 CRD42020165310 Available from https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020165310

one author was a native Chinese speaker, which allowed a unique chance to expand the search. The search was last updated on February 11, 2020. The titles and abstracts database searches were screened to identify potentially relevant records for full-text screening. The titles and abstracts of all remaining records were screened for eligibility to identify records for full-text screening. All records identified for full-text screening were screened to identify records for inclusion in the review. All data that were potentially relevant to the review were then extracted from the studies selected for final inclusion and collated in a spreadsheet as follows: details of publication, participant characteristics, sample size, setting, intervention, study design, data type, and result.

A meta-analysis was not undertaken due to the considerable heterogeneity among the studies included in this review. Therefore, a descriptive approach to data synthesis was adapted, whereby summaries of included studies will be presented. Included studies will be presented in line with the outcomes identified from active interventions that involve HMD VR or AR, specifically, changes in surgery or anatomy training or outcomes related to the trainer or trainee's experience (satisfaction and motivation), population characteristics (study design and study outcome measures), and methodological approach (randomized control studies and crossover studies).

Data were extracted using a standardized template to capture information relating to PICO: population (grade and sex), intervention (characteristics of the HMD VR and AR tool), comparator (traditional/other teaching methods), and outcomes (assessment score, time, and subjective feeling).

1.4 Quality Assessment

The study's quality information was collected using the RoB 2.0 tool (Sterne et al., 2019) for assessing the risk of bias. The risk of bias plot was created by using the Robvis tool (McGuinness and Higgins, 2021).

2 RESULTS

2.1 Study Selection

The electronic searches generated a total of 1,241 studies. After removing duplicates, 848 remained. Of these, 801 studies were excluded because the studies did not meet the criteria after reviewing the abstract. The full text of the remaining 47 studies was reviewed. Among those, 30 studies were discarded due to the reasons in the flowchart (Figure 1). After applying the inclusion criteria and exclusion criteria, 17 studies were identified for inclusion in the review (Table 1). No unpublished relevant studies were obtained. Figure 2 shows what year the screened and included studies were published, illustrating the increasing interest in the topic over the last number of years.

2.2 Study Characteristics

2.2.1 Methods

Among all studies selected, 15 were randomized controlled trials (RCTs), nine of which were published in English (Coulter et al., 2007; Stepan et al., 2017; Pulijala et al., 2018; Alismail et al., 2019;

Logishetty et al., 2019a; Logishetty et al., 2019b; Rojas-Muñoz et al., 2019; Frederiksen et al., 2020; Zackoff et al., 2020) and seven of which were published in Chinese (Meng et al., 2018; Cai et al., 2019; Chen et al., 2019; Jiang et al., 2019; Wang H. et al., 2019; Wang P. et al., 2019). One was a randomized single-blinded crossover study (Harrington et al., 2018) in English, and one was a six-week pre-post comparison study (Logishetty et al., 2020) in English.

2.2.2 Participants

The included studies involved 1,050 participants and 978 of those participated in RCTs (Figure 3). The main inclusion criteria entailed medical students (first-year to master students), surgical trainees, and nursing interns.

2.2.3 Intervention

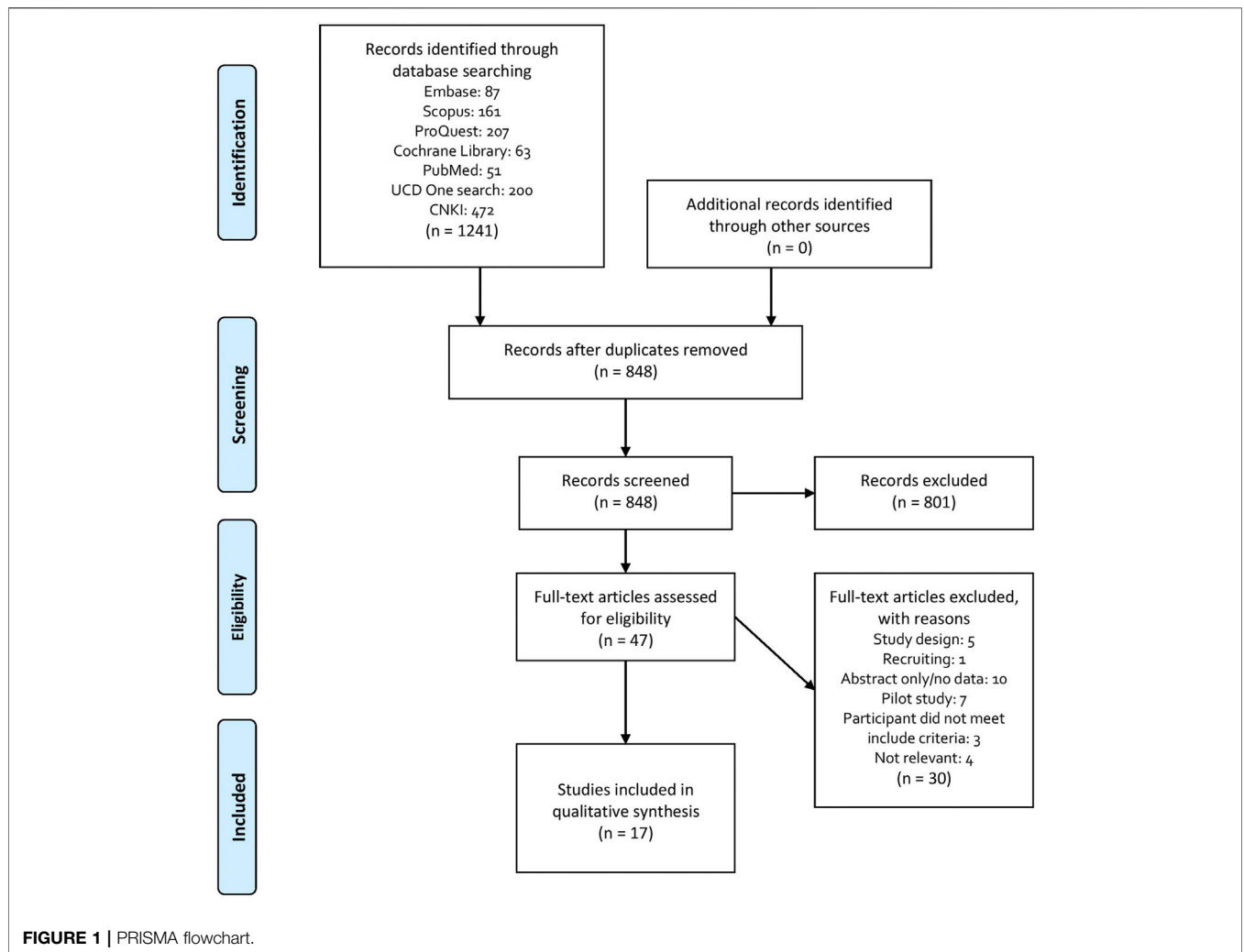
The interventions applied in the studies were VR headsets or AR headsets. A total of 11 studies used VR HMD as interventions, including Oculus Rift ($n = 4$), HTC VIVE ($n = 4$), Gear VR ($n = 2$), and customized HMD ($n = 1$). The rest of the six studies used AR HMD as interventions, including HoloLens ($n = 5$) and AiRScouter glasses ($n = 1$).

2.2.4 Outcomes

Out of 15 RCTs studies, four studies compared the VR/AR HMD method with the traditional teaching method, such as paper materials, 2D videos, and slides (Pulijala et al., 2018; Cai et al., 2019; Jiang et al., 2019; Wang H. et al., 2019). Five studies compared VR/AR HMD with other teaching methods such as desktop and real person guide (Coulter et al., 2007; Logishetty et al., 2019a; Logishetty et al., 2019b; Rojas-Muñoz et al., 2019; Wang H. et al., 2019). Five studies used VR/AR HMD as additional training to support education (Stepan et al., 2017; Meng et al., 2018; Alismail et al., 2019; Chen et al., 2019; Zackoff et al., 2020). One study used HMD to increase the immersive feeling (Frederiksen et al., 2020).

2.2.4.1 Compared with Traditional Method

Four studies compared VR/AR intervention with traditional teaching methods (Table 2). Pulijala et al. (2018) designed an RCT ($n = 91$) to be able to compare immersive VR training with traditional teaching. The study group used Oculus Rift with Leap Motion tracker to interact with the anatomy data and viewed 360° videos of an operating room, while the control used a standard PowerPoint presentation and viewed 2D video of similar content. The knowledge gained was significantly increased in scores for both the VR group ($p = 0.024$) and the control group ($p = 0.025$); however, the participants who used the VR headset performed better overall, especially for the early stage (first-year and second-year) residents. This is common in AR/VR training, where it has been found that the nonexperts appear to benefit the most from the experience (Pringle et al. 2018). Another example comes from Cai et al. (2019) who conducted a similar controlled study ($n = 50$). The study and control groups were given theoretical training using the virtual 3D model generated from CT and MRI scans simultaneously. The intervention was then applied to the study group, where they used the HTC VIVE VR headset to watch real



operation 360° video to learn the anatomy and operation process, while the control group learned from presentation slides, anatomy pictures, models, and 2D videos explained by the lecturers. The control group also entered the operation room to observe the operation process. The results showed a significant difference in test score between the intervention and the control groups ($p = 0.023$).

Jiang et al. (2019) designed an RCT ($n = 52$) to evaluate the effect of the application of mixed reality technology in teaching spine and spinal cord injury. They developed a mixed reality teaching model with real patient case’s MRI 3D reconstruction. The lecturer equipping HoloLens demonstrated the operation process to the study group, and the teaching content was delivered through a monitor. Students in the study group used HoloLens. They learned the process in a simulated environment with all virtual content being synchronized on all headsets, while the control group was taught through the traditional method, including slides and paper teaching materials. The posttest result did not show a significant difference in score between the two groups ($p > 0.05$), but the participants in the study group had a better understanding of the 3D structure ($p < 0.01$). By utilizing

the same approach, Wang P. et al. (2019) conducted an RCT experiment ($n = 120$) to explore the effect of this technology in hepatobiliary surgery. Theoretical and surgical operation assessment showed a significant difference in score between the study and control groups, which was different from the previous study ($p < 0.05$). The study group’s error rate was significantly lower than the control group ($p < 0.05$).

2.2.4.2 Compared with Nontraditional Method

Five studies compared VR/AR HMD with other teaching methods, such as customized simulators or an expert one-to-one guide (Table 3). The earliest randomized controlled study ($n = 25$) explored HMD’s effect on medical education learning performance was conducted back in 2007 (Coulter et al., 2007). The study group wore a stereoscopic HMD as a fully immersed system, while the control group used a simulation via a PC monitor as a partially immersed system. The result showed significant difference in the pre/posttest in overall ($p < 0.001$), study group ($p < 0.001$), and control group ($p = 0.024$). The study group showed a higher gain than the control group. Logishetty et al. (2019a) conducted an RCT ($n = 24$) to determine that the

TABLE 1 | General data of 17 included studies.

Author	Sample size	Intervention	Study design	Result
Frederiksen et al. (2020)	First-year student $n = 31$, intervention $n = 16$, control $n = 15$	Oculus Rift, 360° video	Randomized control trial	Control: not capable Both groups' total time ($p < 0.001$) and blood loss ($p < 0.001$); cognitive load increase (study group: 43.1%, control group: 23%, $p < 0.001$); motion sickness ($p = 0.62$)
Logishetty et al. (2019a)	Postgraduate student $n = 24$, intervention $n = 12$, control $n = 12$	HTC VIVE	Randomized control trial	Control: conventional preparation materials Procedure-based assessment (PBA) level (study group level 3b, control group level 2a, $p < 0.001$); PBA satisfactory ($p < 0.001$); task-specific checklist ($p < 0.001$); inclination and anteversion error from target ($p < 0.001$); operative time ($p = 0.005$)
Logishetty et al. (2019b)	Final-year medical student $n = 24$, intervention $n = 12$, control $n = 12$	MicronTracker HoloLens	Randomized simulation trial	Control: surgeon guide Study group improvement $p < 0.001$; control group improvement $p < 0.001$; improvement between groups $p = 0.281$
Pulijala et al. (2018)	Master student $n = 91$, intervention $n = 51$, control $n = 40$	Oculus Rift with Leap Motion	Multicentre parallel single-blind randomized controlled trial	Control: traditional teaching method Self-confidence level $p = 0.034$; knowledge improvement (study group performed better, $p > 0.05$)
Stepan et al. (2017)	Medical students $n = 66$, intervention $n = 33$, control $n = 33$	Oculus Rift	Randomized controlled study	Control: no VR as additional method post intervention, or retention quizzes assessments ($p > 0.05$); engaging, enjoyable, and useful (all $p < 0.01$); motivation assessment ($p < 0.01$)
Zackoff et al. (2020)	Third-year medical student $n = 168$, intervention $n = 78$, control $n = 90$	Oculus Rift	Randomized controlled prospective study	Control: no VR as additional method Consideration/interpretation of mental status ($p < 0.01$); assignment of the appropriate respiratory status assessment ($p < 0.01$); recognition of a need for escalation of care ($p = 0.0004$)
Rojas-Muñoz et al. (2019)	Medical students $n = 20$	HoloLens, Kinect	User study STAR AR HMD vs. conventional telestrator	Control: conventional telestrator Place error (task 1, $p < 0.001$; task 2, $p = 0.01$); time (study group longer, task 1, $p < 0.001$; study group longer, task 2, $p = 0.013$); focus shifts (task 1, $p < 0.001$; task 2, $p = 0.0038$)
Alismail et al. (2019)	Healthcare medical center $n = 32$, intervention $n = 15$, control $n = 17$	AiRScouter WD-200B glasses	Randomized study	Control: no AR as additional method time for ventilation (study group longer, $p = 0.005$); per cent adherence to the intubation checklist ($p < 0.001$)
Coulter et al. (2007)	$n = 25$, fully immersive $n = 13$, partially immersive $n = 12$	Customized HMD	Randomized study	Control: PC Time ($p = 0.004$); both groups' score pre-post ($p < 0.001$, study group: $p < 0.001$, control group: $p = 0.024$)
Jiang et al. (2019)	Sophomore undergraduate $n = 52$, intervention $n = 26$, control $n = 26$	HoloLens	Randomized study	Control: traditional teaching method Three-dimensional construction and enhancement of class atmosphere ($p < 0.01$); class satisfaction and initiative of learning ($p < 0.05$)
Wang H. et al. (2019)	Nursing student $n = 125$, intervention $n = 62$, control $n = 63$	HTC VIVE	Randomized study	Control: intravenous injection simulator Both groups' posttest scores ($p < 0.05$); knowledge improvement (study group performed better, $p > 0.05$); critical thinking and clinical reasoning and clinical learning ($p < 0.05$)
Wang P. et al. (2019)	Surgery postgraduate $n = 120$, intervention $n = 60$, control $n = 60$	HoloLens	Randomized study	Control: traditional teaching method Theoretical examination and surgical operation assessment ($p < 0.05$); error rate ($p < 0.05$); satisfaction ($p < 0.05$)
Cai et al. (2019)	Clinical medicine student $n = 50$, intervention $n = 25$, control $n = 25$	HTC VIVE	Randomized study	Control: traditional teaching method Average score ($p < 0.05$); satisfaction ($p < 0.05$)
Chen et al. (2019)	Neurosurgery interns $n = 80$, intervention $n = 40$, control $n = 40$	HoloLens	Randomized study	Control: no AR as additional method Mastery degree of lateral ventricle anatomy ($p < 0.05$); proficiency in puncture procedure ($p < 0.05$); confidence ($p < 0.05$); first-pass success rate of puncture ($p < 0.05$)
Meng et al. (2019)	Undergraduate intern $n = 70$, intervention $n = 35$, control $n = 35$	Gear VR	Randomized study	Control: no AR as additional method Number of patients ($p > 0.05$); theoretical exam ($p < 0.05$); satisfaction ($p > 0.05$)

(Continued on following page)

TABLE 1 | (Continued) General data of 17 included studies.

Author	Sample size	Intervention	Study design	Result
Harrington et al. (2018)	Undergraduate $n = 40$, group A $n = 20$, group B $n = 20$	Gear VR	Randomized single-blinded crossover study	Higher engagement level ($p < 0.0001$); across time periods ($p < 0.0007$); lower TUIT ($p < 0.0001$); across time periods ($p < 0.0005$)
Logishetty et al. (2020)	Orthopedic residents $n = 32$	HTC VIVE	6-week pre-post comparison with expert performance	Reached expert levels 9 of 10 metrics; procedural errors reduced by 79%; assistive prompts reduced by 70%; procedural duration reduced by 28%

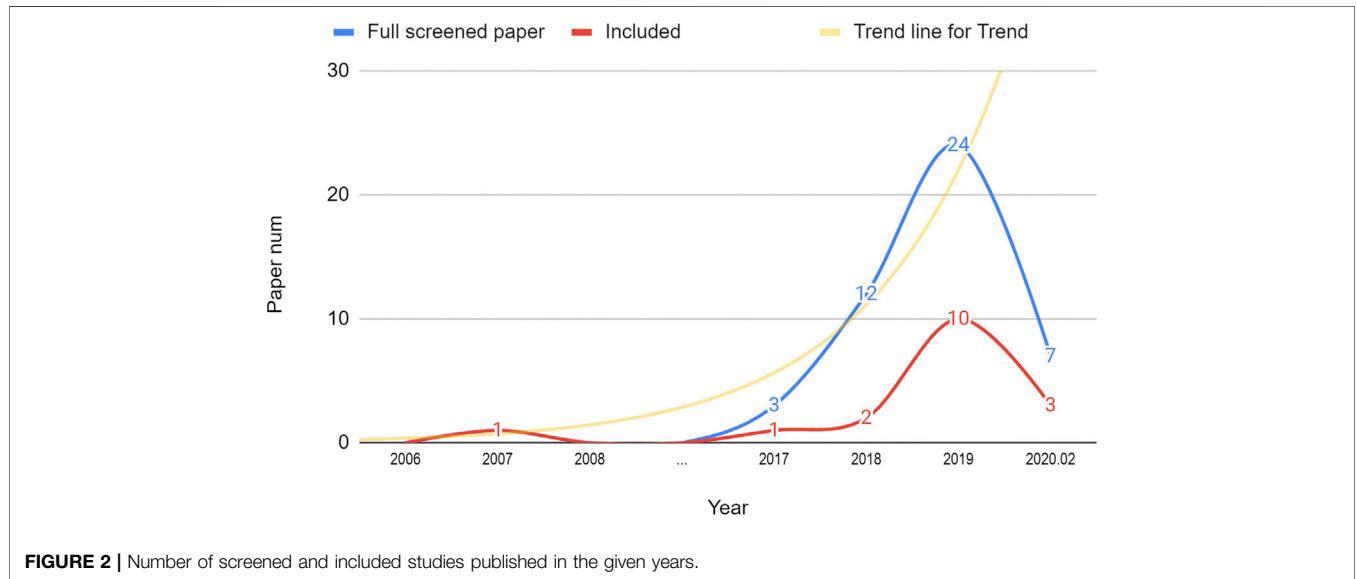


FIGURE 2 | Number of screened and included studies published in the given years.

training effectiveness of using a VR headset was higher than conventional preparation for performing total hip arthroplasty (THA). All participants received standard guidelines and materials to ensure they had similar basic knowledge before the experiment. The study group was enrolled in a six-week VR training program equipped with the HTC VIVE VR headset, while the conventional group received only given preparatory material. The PBA component score and the task-specific checklist score were significantly higher in the VR group than in the control group ($p < 0.001$), which indicated that VR-trained surgeons performed at a higher level than controls. Moreover, the VR group performed faster to complete the procedure ($p = 0.03$) and was more accurate in component orientation (mean error 4° vs. 16°). Another randomized controlled trial was conducted ($n = 125$) by Wang P. et al. (2019). All participants were trained under the teaching mode of “real person training + model assistance + virtual reality.” The study group used the HTC VIVE VR headset as an immersive VR training method, while the control group used an intravenous injection simulator as a nonimmersive VR training method. In the theory test, both study and control groups significantly improved scores ($p < 0.001$). The study group had a higher mean score, but there was no significant difference between groups ($p = 0.136$). However, in the injection test, the study group had significantly higher scores ($p = 0.027$),

demonstrating that the immersive VR training method has similar teaching effectiveness to the customized training tool.

Logishetty et al. (2019b) developed an enhanced AR headset capable of tracking bony anatomy in relation to an implant and designed a randomized trial ($n = 24$) to assess the suitability of it as a training tool for implant orientation. Both groups had standard lectures before the experiment started. During the experiment, both groups had four training sessions, between which there was a 5- to 9-day interval. In each session, the study group used the HoloLens AR headset, while the control group had an expert surgeon who guided the training. The participants in the study group had a significantly lower error of target orientation ($1^\circ \pm 1^\circ$) than those in the control group ($6^\circ \pm 4^\circ$) as they confirmed the final orientation when the headset light turned from red to green ($p < 0.001$). The result showed significant improvements in both groups when compared the final assessment score with the pretest score correspondingly. There was no significant difference in accuracy between the two groups in the assessment ($p = 0.281$) and concealed the pelvic tilt test ($p = 0.301$). 11 of 12 participants stated that they would use the AR platform as a training tool for developing visuospatial skills and 10 of 12 for procedure-specific rehearsals. Most participants (11 of 12) stated that a combination of an expert trainer for learning and AR for unsupervised training would be

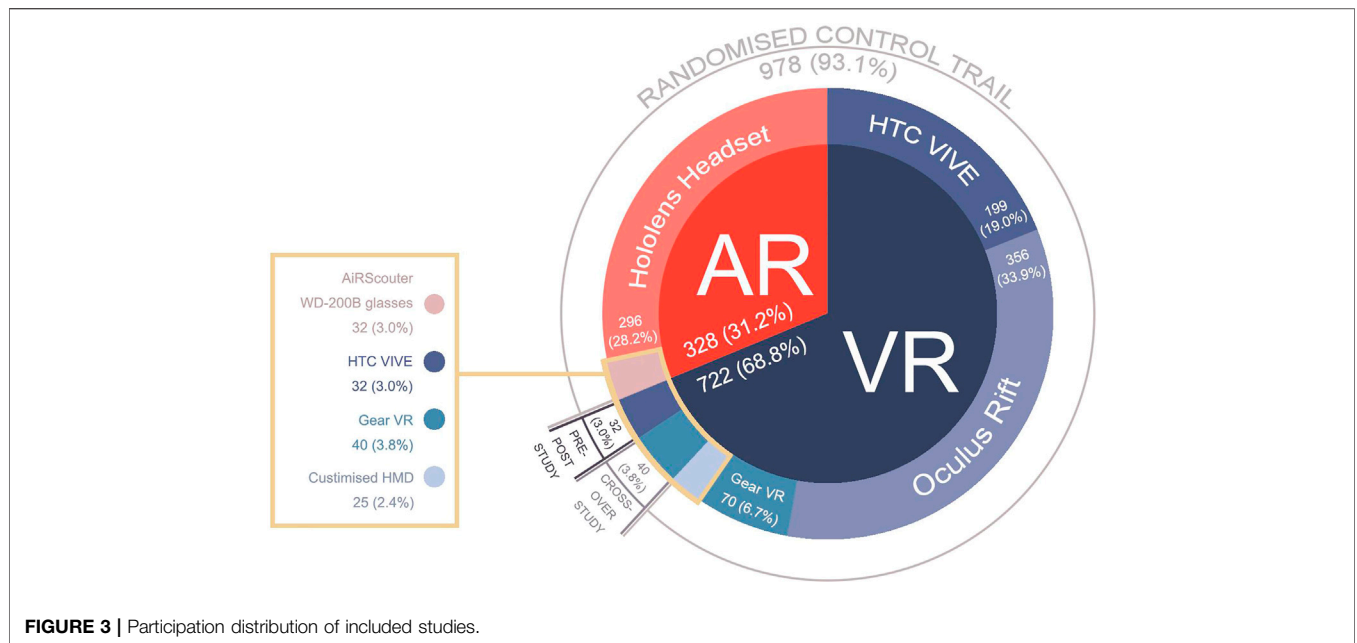


FIGURE 3 | Participation distribution of included studies.

TABLE 2 | Compared with the traditional method.

Category	Intervention	Authors	n	Results
Theoretical assessment	VR	Pulijjala et al. (2018)	91	Study group $p = 0.024$ Control group $p = 0.025$ Between groups $p > 0.05$
		Cai et al. (2019)	50	Between groups (study group performed better, $p = 0.023$)
	AR	Jiang et al. (2019)	52	Between groups $p > 0.05$
Surgical operation assessment		Wang H. et al. (2019)	120	Between groups (study group performed better, $p < 0.05$)
				Between groups (study group performed better, $p < 0.05$)
Error rate				Between groups (study group performed better, $p < 0.05$)

preferred. This demonstrates an interesting result where AR learning as a self-evaluation tool could prove useful in the future.

Rojas-Muñoz et al. (2019) investigated the benefits of an AR HMD telementoring system when compared with a conventional telestrator in surgical guidance by conducting a comparative study ($n = 20$). The study group used the HoloLens AR HMD to receive the instructions during the operation, while the control group needed to watch on a nearby screen. The result showed significant differences between the study group and the control group in placement errors (Task 1: $p < 0.001$; Task 2: $p = 0.01$) and focus shifts (Task 1: $p < 0.001$; Task 2: $p = 0.0039$). In general, the study group used more time to complete each task. It was reported in this study that the use of the HMD caused discomfort, which can be very common with older three degrees of freedom (3DoF) HMDs and is still an issue with the current six degrees of freedom (6DoF) AR/VR HMDs but appears to be steadily improving with every generation.

2.2.4.3 Supportive Usage

Five studies used VR/AR HMD as additional training to support education (Table 4). Zackoff et al. (2020) conducted a

randomized controlled prospective study ($n = 168$) to determine whether exposure to immersive VR-simulated pediatric respiratory distress environment improves students' emergency recognition. All participants received the standard curriculum with a subsequent high-fidelity mannequin simulation, while the study group underwent an additional VR curriculum using the Oculus Rift VR headset. The result showed a significant difference for consideration/interpretation of mental status ($p < 0.01$). The study group performed significantly better in the assignment of assessing appropriate respiratory status ($p < 0.01$) and recognizing a need for escalation of care. Meng et al. (2018) conducted a similar experiment ($n = 70$). Two senior lecturers taught both control and study groups by traditional teaching methods (CT image, slides, and daily operation observation). The study group used the Gear VR headset (only 3DoF compared to most VR HMD with 6DoF) to watch real operation 360° videos after the course. The postintervention test showed a significant difference between groups in the score ($p < 0.05$). Stepan et al. (2017) conducted a randomized controlled study ($n = 66$) using the Oculus Rift VR system as an additional training method to evaluate the effectiveness, satisfaction, and

TABLE 3 | Compared with the nontraditional method.

Category	Intervention	Authors	n	Results
Theoretical assessment	VR	Coulter et al. (2007)	25	Study group $p < 0.001$ Control group $p = 0.024$ Between groups $p > 0.05$
		Wang et al. (2019)	125	Study group $p < 0.05$ Control group $p < 0.05$ Between groups $p = 0.136$
Surgical operation Assessment		Logishetty et al. (2019a)	24	Between groups (study group performed better, $p = 0.027$)
Error rate	AR	Logishetty et al. (2019b)	24	Between groups (study group performed better, $p < 0.001$)
	VR	Logishetty et al. (2019a)	24	Study group mean error 4° Control group mean error 16°
	AR	Logishetty et al. (2019b)	24	Study group $1^\circ \pm 1^\circ$ Control group $6^\circ \pm 4^\circ$ Between groups $p < 0.001$
		Rojas-Muñoz et al. (2019)	20	Placement errors between groups Task 1: study group performed better, $p < 0.001$ Task 2: study group performed better, $p = 0.01$

TABLE 4 | Supportive usage.

Category	Intervention	Authors	n	Results
Theoretical assessment	VR	Zackoff et al. (2020)	168	Between groups (study group performed better, $p < 0.05$)
		Meng et al. (2018)	70	Between groups (study group performed better, $p < 0.05$)
		Stepan et al. (2017)	66	Between groups (study group performed better in post-test, $p = 0.87$) Between groups (study group performed better in retention test, $p = 0.47$)
Surgical operation Assessment	AR	Alismail et al. (2019)	32	Between groups (study group performed better, $p < 0.001$)
		Chen et al. (2019)	80	Study group vs. control group = 93.3 vs. 42.5% Between groups $p < 0.05$ (first-pass rate)

motivation in teaching medical students neuroanatomy. Both groups used the same teaching materials, while the study group had access to a VR headset which allowed them to view virtual brain anatomy generated from CT and MRI data. Different from the two studies mentioned above, there was no significant difference in preintervention ($p = 0.86$), postintervention ($p = 0.87$), or retention test ($p = 0.47$) between the two groups. However, the experimental group performed significantly better in the instructional materials motivational survey with greater attention, relevance, confidence, and satisfaction ($p < 0.01$).

Alismail et al. (2019) presented a study ($n = 32$) to assess the effectiveness of using the AR headset as an assistance tool to perform intubation simulation procedure. All participants watched a video and then started the intubation procedure; in the meantime, those in the study group used the AiRScouter AR headset, from which they could see the slides as a guideline additionally. The result showed a significant difference in ventilation time (seconds) between the study and control groups (280 vs. 205; $p = 0.005$). Moreover, the study group had a higher percentage adherence to following the checklist ($p < 0.001$). Chen et al. (2019) conducted a randomized experiment ($n = 80$) to evaluate the mixed reality application in lateral ventricle puncture training. The study group and the control group were taught traditionally for one month, while the

study group used HoloLens AR headset to train the puncture in a simulated environment. As a result, the study group had a significantly higher first-pass rate than the control group (93.3 vs. 42.5%, $p < 0.05$). In the meantime, the study group participants had significantly better 3D reconstruction and more confidence ($p < 0.05$).

2.2.4.4 Cognition and Emotion

Frederiksen et al. (2020) conducted an RCT ($n = 31$) to explore the cognitive load and performance changes after enhancing the immersion of laparoscopic surgery simulation training by using an HMD. The 360° videos were clipped into different stressor levels (calm, light, and severe). The study group and the control group used a conventional VR laparoscopic surgery simulator, while the study group used the Oculus Rift VR headset playing 360° videos of a real operating room in the meantime. The cognitive load was significantly different ($p < 0.001$) between the study group (15.2% in light stressor and 43.1% in severe stressor) and the control group (23.0%). The study group reported a significantly worse performance on most simulator metrics (time, blood loss, damage, and hand movement). The authors stated, “However, immersive VR offers some potential advantages over conventional VR such as more real-life conditions but we only recommend introducing immersive VR in surgical skills training after initial training in conventional VR.”

2.2.4.5 Non-RCT

Harrington et al. (2018) designed a randomized crossover study ($n = 40$) to evaluate the efficiency of immersive 360° video in surgical education when compared with traditional 2D video. The participants were divided into two different groups randomly. One group attended the 360° video experiment using the Samsung Gear VR headset first and then attended the 2D video experiment, while the other group experimented with the same content in the opposite order. The result revealed a significantly higher engagement level ($p < 0.0001$) and a higher level of focusing ($p < 0.0001$) with the 360° immersive video. There was no significant difference in information retention between the two groups ($p = 0.143$).

Logishetty et al. (2020) designed a competency-based simulation curriculum study ($n = 32$) using a VR HMD to evaluate the skills measurement and visuospatial transfer performance. All the residents attended five consecutive VR training and assessment sessions. The outcome of each assessment was compared with four expert hip surgeons' performance in both a physical world assessment and a VR one-off assessment. The result showed that the residents progressively developed surgical skills in VR by practicing repeatedly, and it allowed them to match expert VR levels on nine out of the 10 metrics included in the study. In the preparation phase, the number of errors in instrument selection and usage errors ($p < 0.001$), number of prompts required ($p < 0.001$), and procedural time ($p < 0.001$) were reduced significantly. The performance of the residents in the VR assessments was significantly improved as the inclination error ($p < 0.005$) and anteversion error ($p < 0.001$) were reduced. In the physical world–simulated assessment, the errors in femoral osteotomy height ($p = 0.044$), in femoral osteotomy angle ($p = 0.002$), in acetabular cup inclination ($p < 0.001$), and in acetabular cup anteversion ($p < 0.001$) were significantly reduced, which indicated that the visuospatial skills were transferred from VR to the physical world successfully.

2.2.4.6 Secondary and Additional

The original proposed additional outcomes in the protocol (Section 1.1) are as follows: “side effects of applying HMD into medical education, such as headache, motion sickness, and claustrophobia,” and “learning motivation improvement by HMD VR or AR.” The measures of effect are questionnaires or interview subjective experience. Next, the additional outcomes found in the systematic review will be outlined in three aspects: motion sickness, limitations, and motivations.

The motion sickness symptoms can occur after the user uses the VR or AR HMDs, especially when the virtual space movement does not match the user's movement in reality or their mind, which can be highly dependent on the content (Saredakis et al., 2020). This is heightened if the experiences are on a device only capable of 3DoF (e.g., roll, pitch, and yaw) and not 6DoF (e.g., X, Y, Z, roll, pitch, and yaw). Other factors include frame lag or screen tearing caused by low device capability or bad software optimization, which may enhance such symptoms. Several studies included in this systematic review reported that some

participants in the VR intervention group had motion sickness after the experiment (Meng et al., 2018; Cai et al., 2019; Wang H. et al., 2019). Furthermore, the limited field of view (FOV) and the imagery of the HoloLens AR headset may produce head discomfort and ocular fatigue (Rojas-Muñoz et al., 2019). However, in Frederiksen et al. (2020)'s experiment, there were no motion sickness cases reported. The possible reason might be “minimal head movements compared to immersive VR video games where motion sickness has been an issue.”

Regarding the limitations of VR/AR HMDs summarized from the included articles, their price is generally too high to deploy in a class-scale teaching environment (Stepan et al., 2017; Pulijala et al., 2018; Wang H. et al., 2019). However, as the technology develops, the price of these devices are reducing (Logishetty et al., 2019a) and are cheaper than an orthopedic simulator, open surgical platforms, or synthetic hip models (Meng et al., 2018; Logishetty et al., 2019b, Logishetty et al., 2020). The above conclusions indicated the VR AR technologies are expensive to be applied in some cases; nevertheless, they have the potential to be a cost-effective teaching method compared with other simulators and be an alternative teaching method in the future. The other limitations reported are the lack of model details and haptic feedback (Cai et al., 2019; Logishetty et al., 2019a; Wang H. et al., 2019), the extra workload needed for the user to get familiar with the devices (Stepan et al., 2017; Jiang et al., 2019), and bad user experience caused by limited FOV or the weight of the devices (Jiang et al., 2019; Rojas-Muñoz et al., 2019; Wang H. et al., 2019). Last but not least, Wang P. et al. (2019) mentioned that as one HMD can only support one user, it is time-consuming to conduct an experiment or teaching mission and has potential health problems with devices sharing, which needs extra attention under the current COVID pandemic situation.

As for the motivation and confidence aspect, the included studies found that the usage of VR/AR HMDs could improve participants' learning motivations and self-confidence by the immersive environment and interactive teaching process. The more satisfied students are, the more engaging students are in the teaching process. Compared with the traditional teaching method, the study group participants showed significantly higher satisfaction and motivation to the teaching method than those in the control group (Cai et al., 2019; Jiang et al., 2019; Wang H. et al., 2019). The same effect also showed when the VR/AR techniques were compared with other simulators (Wang H. et al., 2019) or used as an additional teaching tool (Stepan et al., 2017). However, in Meng et al. (2018)'s experiment, there was no significant difference in the mean satisfaction score. The confidence level significantly increased in both groups in several studies, but the participants of the study group showed significantly higher self-confidence scores (Pulijala et al., 2018; Chen et al., 2019).

2.3 Risk of Bias Within Studies

To reduce the bias of language, this systematic review included English studies and Chinese studies. Among 15 RCTs, nine were in English and six were Chinese. All RCT studies' risk of bias was assessed by using the RoB 2.0 tool (Figure 4). Four English

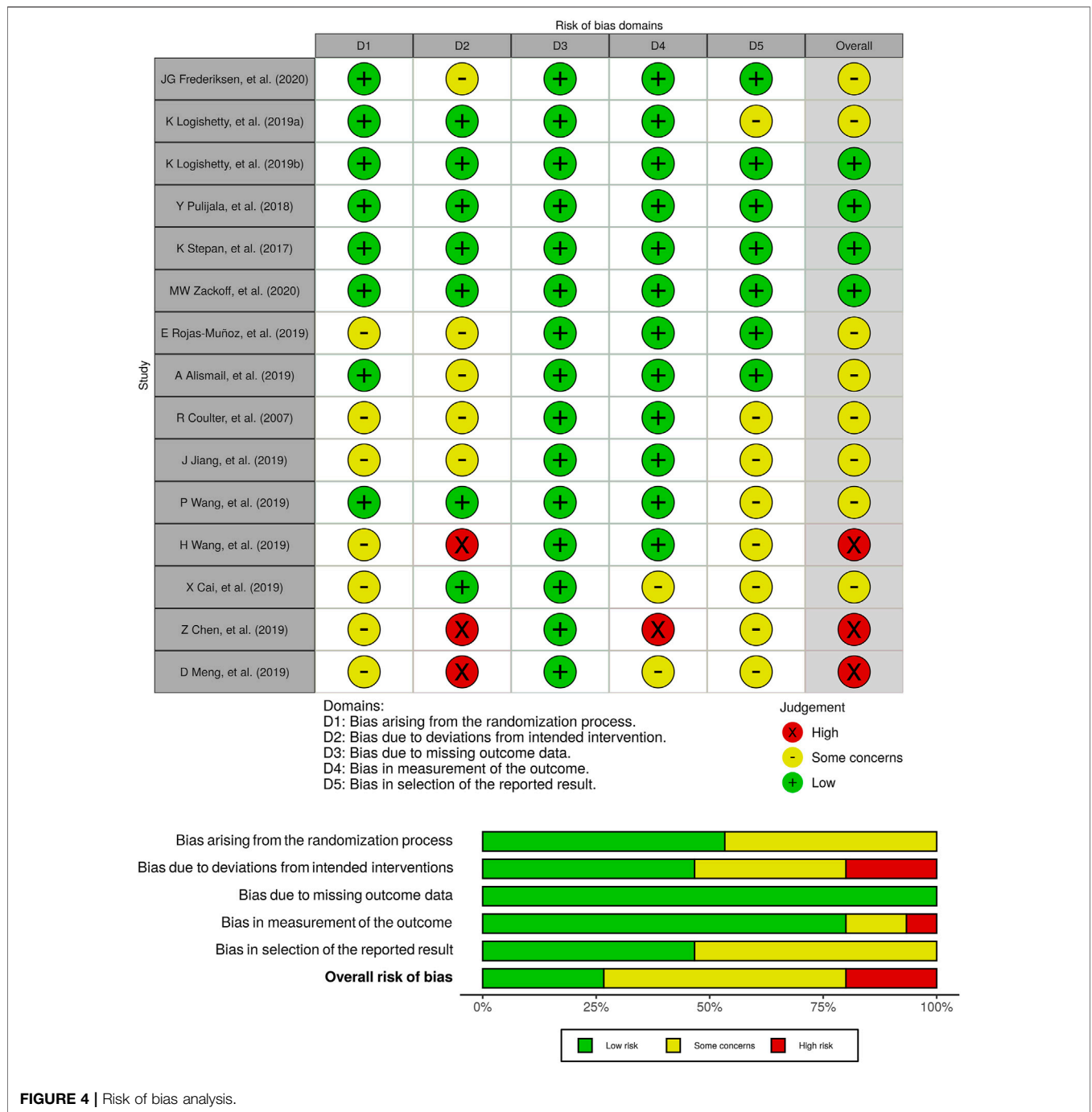


FIGURE 4 | Risk of bias analysis.

articles had a low risk of bias, while the other five English articles had some bias concerns. The primary concern was bias due to deviations from the intended intervention. For instance, they did not mention whether the participants or the data accessors were blinded to the random assignment. Some studies did not clarify whether the allocation sequence was random, and some did not mention a prespecified analysis plan for analyzing the result.

Meanwhile, none of the Chinese articles had a low risk of bias. Three studies had a high risk of bias due to deviations from

intended intervention as there was no information about clarifying the assignment process or analysis after the assignment. All Chinese studies lacked a prespecified analysis plan or an experiment protocol, and the majority of them did not specify the detail of random sequences. Three studies in Chinese had a bias in the measurement of the reported result. Overall, four English studies had a low risk of bias, three studies in Chinese had a high risk of bias, and the rest studies had some concerns.

3 DISCUSSION

The review will first summarize the evidence for both advantages and disadvantages of VR/AR HMD application in medical training. These are the first two questions of this study while the third question will be answered in the proof subsection, demonstrating that these approaches do indeed increase the efficiency of teaching in medical education. Finally, the use of AR/VR as a support tool will be addressed in the final subsection. After summarizing all the evidence, the limitations of this review will be discussed.

3.1 Summary of Evidence

This systematic review focused on clinical educational studies related to VR/AR HMD application in medicine. It revealed that compared with traditional teaching media and other additional teaching methods, the application of VR and AR HMDs improved students' learning curve and motivation. The participant who used virtual HMDs had a better performance in the theory test and the operation examination. Furthermore, the HMDs also provided immersion for the simulated learning environment to increase students' cognition load, maintaining students' performance in the real-life study case.

3.1.1 When Compared with the Standard Teaching Method or Other Types of Simulators, What Are the Comparative Effectiveness of HMD VR or AR Usage in Medical Education?

The standard teaching method refers to the case that lecturers give out the course by using paper-based teaching materials, slides, and videos, while other types of simulators, in this review case, could be 3D print solid or silicon models, PC/phone educational software, and conventional simulator without HMD such as LapSim^{®4}. This systematic review found three aspects of comparative effectiveness of VR and AR HMD application: motivation, learning efficiency, and space efficiency.

Firstly, this review has found that the immersive environment provided by the HMDs increases student's learning motivation and course satisfaction. The results of studies show that students who use HMD intervention during the study process are more satisfied and motivated (Stepan et al., 2017; Pulijala et al., 2018; Cai et al., 2019; Jiang et al., 2019; Wang H. et al., 2019). Furthermore, the simulation offers the residents a chance to experience the test or operational environment before entering an actual one. It increases residents' self-confidence in the knowledge they gained (Pulijala et al., 2018; Chen et al., 2019). With the mental status enhanced, the knowledge can be transferred more effectively (Zackoff et al., 2020). Secondly, the HMDs provide a stereoscopic view, which would potentially benefit the curriculum that needs students to reconstruct spatial information. According to the result of this review, residents who used VR or AR HMDs performed better in 3D reconstruction (Meng et al., 2018; Cai et al., 2019; Jiang et al., 2019) and had a better understanding of the new information

(Harrington et al., 2018). Based on those benefits, the system can generate detailed operation replay or a high-quality 3D virtual model to support the learning process. The student will have unlimited chances to learn and practice without considering any waste of cadaver resources, which maximizes learning opportunities while cutting down the cost at the same time (Meng et al., 2018; Cai et al., 2019; Chen et al., 2019; Logishetty et al., 2019a). The interactive learning mode and hands-on learning experience can benefit student's learning curve (Meng et al., 2018; Cai et al., 2019; Jiang et al., 2019), because of which, the student can conduct unsupervised self-driven learning (Logishetty et al., 2019a; Logishetty et al., 2019b) with live feedback (Alismail et al., 2019; Logishetty et al., 2019b). Finally, the usage of VR or AR HMDs as teaching supportive material is space-efficient compared with actual 3D models and simulators and causes fewer collisions during the practice when compared with other media (Rojas-Muñoz et al., 2019).

3.1.2 What Are the Disadvantages of Using HMD VR or AR? Which One Has a Lower Side Effect?

The results of included studies (Section 2.2.4.6) give answers to the second question proposed in the protocol. This review discovered two disadvantageous aspects of HMDs usage in medical education: motion sickness and other limitations. Motion sickness symptoms cases were reported in several studies, while there was no specific figure to reflect the scale (Meng et al., 2018; Cai et al., 2019; Wang H. et al., 2019). The VR HMD motion sickness can be eased by minimizing head movements (Frederiksen et al., 2020). According to the result, AR HMD has a lower side effect as only one AR study reported head discomfort and ocular fatigue (Rojas-Muñoz et al., 2019).

Except for the potential motion sickness issue, state-of-the-art VR and AR HMDs have some other limitations. As commented in several included articles, the cost of VR HMDs is too high to apply in a class-scale teaching scenario (Stepan et al., 2017; Pulijala et al., 2018; Wang H. et al., 2019); however, the price of VR HMD is reducing when the technique is developing (Logishetty et al., 2019a), and the price of AR HMD is lower than an orthopedic simulator, open surgical platforms, or synthetic hip models (Meng et al., 2018; Logishetty et al., 2019b; Logishetty et al., 2020). The majority of studies that mentioned price limitations are those using VR HMDs intervention; however, this review cannot conclude that AR HMDs are easier to deploy as HoloLens AR HMD is not a commercial product and its price is much higher than a commercial VR HMD. One of the other limitations reported is the lack of model details and tactile feedback in the VR environment (Cai et al., 2019; Logishetty et al., 2019a; Wang H. et al., 2019). The AR devices may potentially have similar limitations due to their lower capacity in graphics computation. However, those limitations are not reported in the included articles. The reason might be the different functionalities between VR and AR applications. AR is generally used as a reference tool that provides extra information to the real object or person, while VR is more isolated so that the virtual environment detail affects the learning process directly. Moreover, the HMD design itself can lead to a bad user experience caused by limited

⁴<https://surgicalsociety.com/systems/lapsim/>

FOV and the extra weight on a user's head (Jiang et al., 2019; Rojas-Muñoz et al., 2019; Wang H. et al., 2019); this issue as mentioned before is becoming less of an issue due to the rapid improvements in HMD design.

3.1.3 Is There a Definitive Advantage of HMD VR and AR When Used for Increasing the Efficiency of Teaching in Medical Education?

The third question proposed in the protocol is addressed in the outcome section (Section 2.2.4). Some of the VR or AR HMDs intervention groups performed significantly better than the control groups in the theoretical posttest (Meng et al., 2018; Cai et al., 2019; Wang H. et al., 2019), while some studies did not find a significant difference between the two groups in the theory test, but both groups had significant improvements and the study group performed better (Stepan et al., 2017; Pulijala et al., 2018; Jiang et al., 2019).

Regarding the actual or simulated surgical exam, the included articles' study groups had significantly higher scores (Alismail et al., 2019; Chen et al., 2019; Logishetty et al., 2019a; Wang H. et al., 2019; Wang P. et al., 2019; Zackoff et al., 2020) and a lower error rate than the control groups (Logishetty et al., 2019a; Logishetty et al., 2019b; Rojas-Muñoz et al., 2019; Wang H. et al., 2019). Even when the control group was guided by a surgical expert individually, the improvement of the study group using the AR HMD self-study system was still comparable (Logishetty et al., 2019b), which indicates the potential of using AR HMD in an alternative supportive teaching role. Besides the improvements to learning outcomes, the VR HMD intervention could aid in the development of more real-life skills as they have been shown to increase cognitive load due to the stress of experiencing a more realistic environment than other teaching methods. One example of this effect is in the study by Frederiksen et al., (2020), where the VR study group had significantly worse performance on most simulator metrics (time, blood loss, damage, and hand movement) due to the extra cognitive load when compared to the control. This was due to the real-life operational 360° video the participants were immersed in. This indicates that the usage of VR HMD could help guarantee the skill transfer from the simulators to a real-life case, but basic skills should still be taught more abstractly. This abstraction could still be taught in VR, and it is the power of this medium that allows changes to fidelity at will. Finally, by repeatedly practicing with the VR HMD operation simulator, the novice surgeons could gradually build up their skills until they performed as same as an expert level within the same VR assessment; nevertheless, the knowledge gain could also be transferred to the physical world—simulated assessment (Logishetty et al., 2020).

3.1.4 Do HMD VR and AR Have the Potential to be Support Tools for Medical Education?

The above evidence can also be used to answer the last question proposed as although current stage VR and AR HMDs have some limitations such as motion sickness and can still be relatively costly if an entire class needs access to multiple HMD's, they still have great potential to be supportive medical education tools

(Stepan et al., 2017; Harrington et al., 2018; Logishetty et al., 2019a; Frederiksen et al., 2020).

With ongoing hardware development, the motion sickness issue should be eased and even completely avoided by making the headset lighter and smaller and increasing the rendering capacity. Looking ahead in terms of accessibility, the high-performance hardware's price is reducing and is getting similar to a high-end smartphone.

Several researchers in the included studies within this review pointed out that the VR and AR applications would never replace the traditional teaching method but could provide supportive teaching materials (Logishetty et al., 2019b; Wang H. et al., 2019). As the skills and knowledge gained in the virtual world can be successfully transferred to the physical world (Logishetty et al., 2020; Zackoff et al., 2020), the current medical and veterinary anatomy education challenges, such as the lack of anatomy cadaver resources, could be eased with the introduction by merging VR and AR technique into the teaching curriculum.

3.2 Limitations

The main limitations of this systematic review are the following three points:

- Language bias. The search strategy includes English and Chinese articles to reduce language bias. However, to avoid language bias, more languages need to be added to the search strategy.
- Risk of bias for the RCTs. According to the risk of bias analysis chart (Figure 4), over half of the included article has some bias concerns. Due to the publication format difference, most Chinese articles cannot meet the requirements of the RoB framework (Sterne et al., 2019), which makes three included Chinese articles high risk of bias.
- Abstracts covered. This systematic review did not include the articles or studies that only provided abstract because it is hard to judge whether the study meets all the inclusive criteria. However, this fact became one of the limitations in this review as it did not cover all the articles, including gray publications and clinical trial protocols.

3.3 Conclusion

VR and AR HMD's applications in medical training are moving slowly into the mainstream as with their reduced cost and increased availability, researchers have taken notice in their search to improve education efficiency. Compared with traditional teaching methods and other non-HMD VR simulators, VR and AR HMDs stimulate students' learning motivation, increase their satisfaction, and improve students' learning outcomes. The immersive VR-simulated environment prepares students' better mentally before dealing with emotionally challenging real-life medical situations, which can help guarantee the skill transferred from virtuality to reality. Motion sickness and some hardware limitations are reported in this review, but with every passing year, innovations in this field mean these limitations are either being reduced or becoming not existent.

The future study directions can be divided into two aspects: HMDs as tools to support students' theoretical knowledge gain in the curriculum and be simulators to training students' surgical skills. The current studies concentrate on developing theoretical knowledge. However, in future, these studies need to be expanded, and more extensive study groups are needed to evaluate the training efficiency to integrate the interventions into the traditional teaching process. In terms of skill training, the future VR/AR HMD intervention in medical education will be more commonly combined with actual surgical equipment to bridge the gap between simulation and reality. Thus, future studies can target the actual skill and knowledge transfer rate from virtuality to reality with larger intervention groups. As the included articles all focused on some particular scenario, more wide-ranging and longitudinal studies are needed to validate this type of intervention.

Due to the pandemic, remote learning, which already was on the rise before the crisis, has accelerated. It is not just in education as countries such as Ireland have passed laws to give the legal right to request home working. Working from home now has become part of society's fabric, in conjunction with the move to requiring continuous professional development for most professions. Research into alternatives to the traditional physical labs could be essential, not just for medical education but for all of education. VR and AR intervention can potentially be a supportive tool for lecturers' teaching, students' self-learning, and professional practitioners' self-evaluation.

Few studies evaluate remote learning using VR or AR interventions, so this is still an ongoing research area. The future experiment direction in this area should concentrate on how online remote teaching could increase the teaching efficiency in medical and veterinary education. The rise of the use of VR/AR within academia, even allowing remote conferences (MacIntyre 2020) to be held in VR, has helped demonstrate its future. Remote learning will still flourish after the pandemic is over, as this natural experiment has demonstrated that these approaches can be successful. With the increasing adoption of VR/AR within remote learning, these successes can be built upon. This trend complements the fact that VR/AR HMD's are also becoming more inexpensive, thus allowing for increasing equity and access

to education across the world with these new technologies if the lessons from many of the experiments outlined in this review are heeded.

At this current stage, VR and AR intervention cannot replace actual cadaver learning material due to their lack of fidelity and lack of tactical feedback will affect students' cognition when faced with actual surgical cases. However, along with ongoing HMD development, the interventions will be more accessible and easier to blend into medical education in the future. Furthermore, the high-fidelity model and haptic innovations will blur the edge between virtuality and reality; but crucially, more experiments are needed to gauge educational efficiency gain and evaluate and verify whether the VR and AR simulators can be a possible replacement to cadavers, avoiding existing ethical problems and resource limitations. Medical education, in particular, has always suffered the problem of having more qualified applicants than places across the world due to resource limitations. Removing these resource limitations could significantly impact equity and access to medical education in the future.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

Conception and design, XX and AC; data collection, XX; article screening, XX, EM, and AC; analysis and interpretation, XX; writing the article, XX and AC; critical revision of the article, XX, EM, and AC; final approval of the article, EM and AC; overall responsibility: XX, EM, and AC.

FUNDING

Author XX has been supported by the China Scholarship Council (201908300021).

REFERENCES

- 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.ijssu.2016.03.034
- Alismail, A., Thomas, J., Daher, N. S., Cohen, A., Almutairi, W., Terry, M. H., et al. (2019). Augmented Reality Glasses Improve Adherence to Evidence-Based Intubation Practice. *Amp* Vol. 10, 279–286. doi:10.2147/amep.s201640
- Cai, X., Qin, J., Huang, S., and Yi, X. (2019). Application of Virtual Reality Technology in Clinical Teaching of Spinal Surgery. *China Contin. Med. Educ.* 11, 18. doi:10.3969/j.issn.1674-9308.2019.23.008
- Chen, Z., Liu, Y., He, B., Huang, S., Hong, W., Liao, Z., et al. (2019). Application of Ventricle Puncture Training System Based on Mixed Reality in Medical Education and Training. *J. TRAUMA EMERG.* 7, 5. doi:10.16746/j.cnki.11-9332/r.2019.01.002
- Coulter, R., Saland, L., Caudell, T., Goldsmith, T. E., and Alverson, D. (2007). The Effect of Degree of Immersion upon Learning Performance in Virtual Reality Simulations for Medical Education. *In Medicine Meets Virtual Reality* 15, 155.

- Frederiksen, J. G., Sørensen, S. M. D., Konge, L., Svendsen, M. B. S., Nobel-Jørgensen, M., Bjerrum, F., et al. (2020). Cognitive Load and Performance in Immersive Virtual Reality versus Conventional Virtual Reality Simulation Training of Laparoscopic Surgery: a Randomized Trial. *Surg. Endosc.* 34, 1244–1252. doi:10.1007/s00464-019-06887-8
- Harrington, C. M., Kavanagh, D. O., Wright Ballester, G., Wright Ballester, A., Dicker, P., Traynor, O., et al. (2018). 360° Operative Videos: A Randomised Cross-Over Study Evaluating Attentiveness and Information Retention. *J. Surg. Educ.* 75, 993–1000. doi:10.1016/j.jsurg.2017.10.010
- Jensen, L., and Konradsen, F. (2018). A Review of the Use of Virtual Reality Head-Mounted Displays in Education and Training. *Educ. Inf. Technol.* 23, 1515–1529. doi:10.1007/s10639-017-9676-0
- Jiang, J., Sun, J., Luo, X., Sun, X., Wang, Y., Xu, X., et al. (2019). Effect Analysis of Applying Mixed Reality Technology in Spinal Surgery Teaching. *China Med. Educ. Technol.* 34, 230. doi:10.13566/j.cnki.cmet.cn61-1317/g4.202002028
- Larsen, C. R., Oestergaard, J., Ottesen, B. S., and Soerensen, J. L. (2012). The Efficacy of Virtual Reality Simulation Training in Laparoscopy: a Systematic

- Review of Randomized Trials. *Acta obstetrica gynecologica Scand.* 91, 1015–1028. doi:10.1111/j.1600-0412.2012.01482.x
- Logishetty, K., Gofton, W. T., Rudran, B., Beaulé, P. E., and Cobb, J. P. (2020). Fully Immersive Virtual Reality for Total Hip Arthroplasty. *JBJS* 102, e27. doi:10.2106/jbjs.19.00629
- Logishetty, K., Rudran, B., and Cobb, J. P. (2019a). Virtual Reality Training Improves Trainee Performance in Total Hip Arthroplasty: a Randomized Controlled Trial. *Bone Jt. J.* 101-B, 1585–1592. doi:10.1302/0301-620x.101b12.bjj-2019-0643.r1
- Logishetty, K., Western, L., Morgan, R., Iranpour, F., Cobb, J. P., and Auvinet, E. (2019b). Can an Augmented Reality Headset Improve Accuracy of Acetabular Cup Orientation in Simulated Tha? a Randomized Trial. *Clin. Orthop. Relat. Res.* 477, 1190–1199. doi:10.1097/corr.0000000000000542
- MacIntyre, B. (2020). Remote conference participation in social virtual worlds.
- McGuinness, L. A., and Higgins, J. P. T. (2021). Risk-of-bias VISualization (Robvis): An R Package and Shiny Web App for Visualizing Risk-of-bias Assessments. *Res. Syn Meth* 12, 55–61. doi:10.1002/jrsm.1411
- Meng, D., Zhao, L., Ouyang, Y., Lin, H., and Huihao, C. (2018). Application of Visual Reality Technology in the Clinical Teaching of Orthopedics Interns. *Med. Inf.* 31, 17. doi:10.3969/j.issn.1006-1959.2018.22.006
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and Group, P. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the Prisma Statement. *Plos Med.* 6, e1000097. doi:10.1371/journal.pmed.1000097
- Pringle, A., Campbell, A. G., Hutka, S., and Keane, M. T. (2018). “Using an Industry-Ready AR HMD on a Real Maintenance Task: AR Benefits Performance on Certain Task Steps More Than Others,” in 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), (Munich, Germany, 16–20. October 2018 (IEEE). doi:10.1109/ISMAR-Adjunct.2018.00075
- Pulijala, Y., Ma, M., Pears, M., Peebles, D., and Ayoub, A. (2018). Effectiveness of Immersive Virtual Reality in Surgical Training-A Randomized Control Trial. *J. Oral Maxillofac. Surg.* 76, 1065–1072. doi:10.1016/j.joms.2017.10.002
- Rojas-Muñoz, E., Cabrera, M. E., Andersen, D., Popescu, V., Marley, S., Mullis, B., et al. (2019). Surgical Telementoring without Encumbrance. *Ann. Surg.* 270, 384–389. doi:10.1097/sla.0000000000002764
- Saredakis, D., Szpak, A., Birkhead, B., Keage, H. A. D., Rizzo, A., and Loetscher, T. (2020). Factors Associated with Virtual Reality Sickness in Head-Mounted Displays: a Systematic Review and Meta-Analysis. *Front. Hum. Neurosci.* 14, 96. doi:10.3389/fnhum.2020.00096
- Stepan, K., Zeiger, J., Hanchuk, S., Del Signore, A., Shrivastava, R., Govindaraj, S., et al. (2017). Immersive Virtual Reality as a Teaching Tool for Neuroanatomy. *Int. Forum Allergy Rhinol.* 7, 1006–1013. Wiley Online Library. doi:10.1002/alr.21986
- Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., et al. (2019). Rob 2: a Revised Tool for Assessing Risk of Bias in Randomised Trials. *bmj* 366, l4898. doi:10.1136/bmj.l4898
- Wang, H.Ou, Y., Hu, P., Deng, Y., Zou, L., Deng, Q., et al. (2019). Application Effect of Mixed Reality in the Teaching of Hepatobiliary Surgery. *Chin. J. Med. Edu Res.* 18, 1230. doi:10.3760/cma.j.issn.2095-1485.2019.12.011
- Wang, P.Yan, W., Wang, F., Wang, L., Cai, H., and Chen, W. (2019). The Application of Immersive Virtual Reality Technology in Experimental Teaching of Intravenous Injection. *Chin. Nurs. Manage.* 20, 176. doi:10.3969/j.issn.1672-1756.2020.02.006
- Zackoff, M. W., Real, F. J., Sahay, R. D., Fei, L., Guiot, A., Lehmann, C., et al. (2020). Impact of an Immersive Virtual Reality Curriculum on Medical Students’ Clinical Assessment of Infants with Respiratory Distress*. *Pediatr. Crit. Care Med.* 21, 477–485. doi:10.1097/pcc.0000000000002249

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.

Copyright © 2021 Xu, Mangina and Campbell. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.