

Review

A Narrative Review of Immersive Technology Enhanced Learning in Healthcare Education

Chris Jacobs ^{1,*}, Georgia Foote ¹, Richard Joiner ¹ and Michael Williams ²¹ Department for Health, University of Bath, Bath BA2 7AY, UK² School of Medicine, Queen's University Belfast, Belfast BT7 1NN, UK

* Correspondence: cj511@bath.ac.uk

Abstract: Immersive technology is a growing field in healthcare education—attracting educationalists to evaluate its utility. There has been a trend of increasing research in this field; however, a lack of quality assurance surrounding the literature prompted the narrative review. Web Of Science database searches were undertaken from 2002 to the beginning of 2022. The studies were divided into three mixed reality groups: virtual reality (VR), augmented reality (AR), 360 videos, and learning theory subgroups. Appraising 246 studies with the Medical Education Research Study Quality Instrument (MERSQI) indicated a gap in the validation of measures used to evaluate the technology. Although, those conducted in VR or those detailing learning theories scored higher according to MERSQI. There is an educational benefit to immersive technology in the healthcare setting. However, there needs to be caution in how the findings are interpreted for application beyond the initial study and a greater emphasis on research methods.

Keywords: immersive technology; technology-enhanced learning; virtual reality; augmented reality; mixed reality; simulation-based education



Citation: Jacobs, C.; Foote, G.; Joiner, R.; Williams, M. A Narrative Review of Immersive Technology Enhanced Learning in Healthcare Education. *Int. Med. Educ.* **2022**, *1*, 43–72. <https://doi.org/10.3390/ime1020008>

Academic Editor: Hideki Kasuya

Received: 19 August 2022

Accepted: 7 October 2022

Published: 15 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

“See one, do one, teach one” is an adage often associated with learning in medicine [1]. Immersive technology enables a user to experience “seeing and doing”, and, arguably, with increasing technical sophistication, “teaching one”. There is a rapid increase in the adoption of technology in healthcare education [2], and the digital transformation of health and care services opens a wide range of applications to immersive technologies. Health Education England (HEE) established a strategy in 2020 to develop and strengthen the research and evaluation of new technology [3]. Broadly this outlined the vision to improve patient safety, deliver a modern trained workforce, innovate, and collaborate outside of healthcare. However, limited evidence of pedagogical concepts and learning theory underpins the technology [4].

Immersive technology can be defined as virtual reality (VR), augmented reality (AR), mixed reality (MR), and extended reality (XR) [5]. The COVID-19 pandemic was a catalyst for many educationalists experimenting with new immersive technology [6–12]. Immersive technology has been associated with learning in numerous health disciplines using a variety of measures [13,14], though a review [5] identified a gap in understanding user experience. Educationalists would benefit from a more coherent understanding of the pedagogical principles underlying existing studies, as well as an assessment of the methods used, the results emerging, and the conclusions drawn.

This review aims to analyse the existing studies' methodologies and demonstrate our current understanding of immersive technology-enhanced learning in healthcare. Specifically, the literature will aim to answer the following:

1. What are the trends in medical education research on immersive technology?

2. What are the disciplines in healthcare (within the scope of the review) in which immersive technology has been studied?
3. What is the quality of the medical education research on immersive technology, including a formal assessment of instrument validity?
4. What are the pedagogical concepts used to understand learning and user experience in the studies reviewed?

A scoping review strategy was adopted to examine how research has been conducted to date in this field and combine this evidence in a critical appraisal and mapping of these characteristics [15]. Additionally, this type of review suits a rapidly developing research area.

Section 2 expands on the above with definitions, learning theories, and related reviews to date to clarify how this review fills a research gap. Section 3 outlines the methodology and charting of the data extraction. Section 4 relates the data outcomes to the research questions, and Section 5 brings the review together to discuss the results, highlight the key findings and limitations, and, finally, the conclusions drawn.

2. Literature Review

2.1. Definitions Relating to Immersive Technology

The definitions surrounding immersion and technology vary according to context and perspective. One view considers the sensory component of the experience, for example, how the environment simulates sounds and haptics. A valid sensorimotor response can infer a place illusion for the user [16]. Fidelity refers to how the user experiences the immersion and how similar it is to the real world [17]. The relevant technology has existed for over 50 years, as first demonstrated in a laboratory-based construct called the 'ultimate display', which considered how we might react to situations of prior knowledge, for example, that an object will fall [18]. Since then, higher fidelity has developed to create a simulated environment that feels real. Immersive technology can thus be defined as a set of applications and interfaces that create a simulation of experiences using multimodal techniques to create sensory stimuli [19].

A reality-virtuality fidelity continuum has been proposed in which VR and AR lie at opposite ends of a mixed reality spectrum [20]. AR merges real and simulated worlds, superimposing digital images or sounds onto views of the real world. In contrast, in VR, users navigate an entirely computer-generated construct. Both technologies create MR (Figure 1). Within the MR continuum, there is an experienced environment, and this model does not infer different levels of immersion.

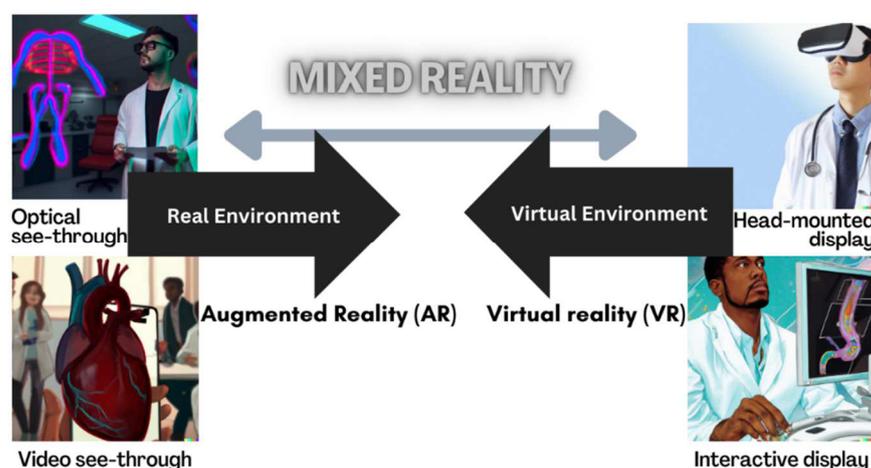


Figure 1. Adapted reality-virtuality continuum. Created by authors CJ and DALL-E.

For the purpose of the review, not all forms of VR will be considered, for example, computer screens displaying an interactive program such as a sequential game-based

virtual patient. More recent immersive VR allows users to interact via head-mounted displays [21] as VR headsets contain 3D inertial sensors that can track and monitor movement that is then reflected in the computer-generated display. Headsets thus allow virtual presence (place illusion) to be experienced as part of a more immersive experience. Recent examples of VR in healthcare use Oculus technology, such as the Rift, to recreate a patient scenario [22] and AR using Microsoft HoloLens with overlaid digital assets for blood transfusion technique [23]. However, simulators that mimic the surgical setting of keyhole instrumentation, for example, early VR innovation of MIST VR for laparoscopic procedures or later technology demonstrated by the EYESi ophthalmic simulator, are described as virtual reality as computer-generated displays with interaction using haptic controllers. These different hardwares still fall under the VR terminology seen in Figure 1.

There is a subset of VR involving passive viewing of 360-degree videos, with limited interaction, though the user can view in multiple directions by turning their head, again requiring inertial headset sensors to match head movement to screen movement. Additional digital annotations to the recording can create some interactivity, suggested to influence the audiences' cognitive access to information and facilitate learning acquisition [24].

2.2. Related Literature of Medical Education, Theories of Learning, and Immersive Technology

Immersive technology has potential use in healthcare. For example, the use of VR has been shown to be useful in acute and chronic pain management [25,26] and MR in-patient stroke rehabilitation [27,28], severe burns [29], and lower extremity injuries [30]. The virtual environment can mimic the real world and provide a stepwise approach to phobia treatment. VR in-vivo exposure can lead to improvement in phobia management [31]. The broader uses within telehealth were accelerated during the COVID-19 pandemic, which enabled healthcare professionals to remotely track patients, allowing real-time feedback and flexible instruction and promoting patient care using VR [32] or AR [33].

The implications for medical practice and direct patient benefit is numerous but depend on the education of healthcare professionals (HCP) in the use of immersive technology. A synthesis of the literature on workplace interventions to achieve adequate HCP education highlighted the challenge of the rise in patient activity and the lack of a proportionate rise in workforce numbers (33). Upscaling the number of trained HCP is a potential solution offered by immersive technology [5].

Teacher and public perceptions of immersive technology in education were explored through the analysis of 4721 social media posts to find what VR-enhanced immersion and absorption into the learning activity offered [24]. A key finding was a strong association of immersive technology in social media in general with education themes.

Suh et al. [5] considered the educational foundations of this technology by applying a stimulus-organism-response framework as a theoretical platform. This classification aimed to evaluate behaviours with "stimuli" and "organism" and reflecting concepts of immersion and presence with affective reactions. The authors reflected on "response" encompassing the learning performance and processes that occurred from which benefit could be derived. This would translate to improved cognitive states and, in turn, behavioural changes, i.e., incorporating technology [34]. Multiple learning theories arose from the 54 papers analysed, many of which were in settings outside of healthcare. Flow theory was described, whereby immersion in a flow state aids concentration, and a balance between challenges and skills is met. A meaningful experience transforms this into eventual testing of newly acquired knowledge and skills, as described in experiential theory [35]. AR technology has a digital addition to the real world and a distinct learning space, which forms the conceptual blending theory [36]. Furthermore, the real-world enhancement of digital overlay in AR can be experienced as different levels of cognitive load [37]. Constructive learning theory dictates that learning is an active process, and learners are participants in knowledge forming and connecting it to established knowledge [38]. This active process is in part stimulated by how the learning environment motivates the learner and forms satisfaction, with motivational theory describing how motivation promotes attention [39].

Self-actualised behaviour is self-regulated, and motivation is a pivotal control. Finally, as indicated in definitions of immersion, how we perceive our presence influences our learning engagement and focus, with the interaction between learning technology and ourselves. The presence theory is conceptualised in any virtual learning environment [40].

In separate work, a scoping review considered the pedagogical underpinning of 360-degree video in healthcare [4]. The authors felt two theories were of significance: the Cone of Experience [41] and the ARCS-V model of motivational learning [42]. These theories were the framework used to analyse papers' outputs. Dale's theory on experience hypothesised that purposeful experiences are the cone base, moving towards the pinnacle of verbal symbols with increasing degrees of abstraction. There is an overlap with experiential theory as three key stages are denoted as learning by: doing, observation, and abstraction. When applied to multimedia, the proposed experience by doing is accomplished with VR creating a real experience [43]. Keller's ARCS-V model is broken down into five domains: attention, relevance, confidence, satisfaction, and volition. All contribute to the motivation to achieve self-regulated goals. It was concluded that 360-degree video in healthcare has relevance to the five domains and has significant effects on users' motivation to learn.

There are two areas our study aims to discover. Firstly, the learning theories researchers considered with their work on immersive technology in healthcare education, and secondly, the author's intentions to employ a deductive stance without leaning on the analysis of preconceived learning theories.

A systematic review from 1990–2017 of the effectiveness of VR educating HCP found an improvement in postintervention VR scores of knowledge gained versus traditional learning [44]. While the quality of included papers was measured, the papers that were extracted from searches used an evaluated score to rate the quality. However, the framework used did not explicitly review the validity of the measures used for the data collection of the papers and hence, could not reliably make conclusions. Tang et al. [13] described numerous findings of value in their systematic review. They identified an upward trend of research in medical education with immersive technology and the varied disciplines covered. This was considered best explained by the Gartner Hype Cycle for technology adoption, which is an adaptation of the Kubler-Ross change curve.

Furthermore, a meta-analysis demonstrated significantly improved surgical procedure scores and operating times when using VR (13). The paper quality assessment was a derived measure of three questions relating to research quality and relevance to the study's research questions.

A 2019 review of non-technical skill acquisition regarding immersive technology in healthcare education found that the majority of papers considered technical skills that could be measured, i.e., task performance scores, and validated tools to assess non-technical skills were not widely used [45]. The authors advocated for more work to validate how we measure outputs in relation to VR studies. Similarly, an integrative review of users' experiences of VR systems noted a significant proportion of articles failed to determine the validity of the instruments used [46].

Medical education traditionally is a pedagogy related to the training and practice of a medical doctor. However, definitions have modernised and vary depending on the source. Learning and training in a clinical setting for non-medical healthcare professionals and interprofessional and non-medical clinicians form an important contribution to medical education [47].

In summary, meta-analyses show improved technical scores following VR use in surgical specialties [13], improved knowledge acquisition [44], and improved non-technical skills, for example, communication skills and teamwork [45] with immersive technology. There are numerous theoretical propositions that authors have used to attempt to explain findings [5], and many outcomes have been captured using non-validated measures [46]. Gaps inevitably exist in how we construct meaning from immersive technology in medical education, given that questions exist about the methods of research and what learning took place.

This review aims to describe the landscape and capture the nature, quality, and interpretation of existing measures.

3. Methods

3.1. Purpose and Reporting

To achieve our aims, a literature review and synthesis of the relevant studies between 2002 and early 2022 was conducted, with searches occurring over 2 weeks from 7 January 2022. The research questions were registered with the Open Science Framework in January 2022 [48], which indicated the research intention.

3.2. Search Methods

The database Web of Science (WoS) was chosen as the primary publication search. It is a trusted multi-source search engine that indexes academic journals and conference proceedings. WoS database searches included a MEDLINE search. The combination of WoS and those indexed with MEDLINE has been reported in recent medical education knowledge syntheses [49,50]. On completion of electronic searches, a hand-search of Google Scholar was undertaken to ensure all relevant articles were reviewed, which included grey literature.

3.3. Eligibility Criteria

To answer the research questions, the papers were required to be from medical-related fields with the applications of immersive technology for educational purposes. Any media within the MR spectrum could be evaluated at the screening stage, and users were identified as trainees or qualified HCPs. The study years included in the searches were 2002–early 2022.

The inclusion criteria for medical participants were expanded beyond doctors or medical students. Studies within paramedic science, nursing, physician associate, midwifery, and radiography were screened to be included. The characteristics of inclusion were healthcare professionals with a clinically orientated or procedural purpose to the research study.

Themes that pertained to technology hardware within the MR umbrella were devices described as VR if allowing for the manipulation of a virtual environment using controls or head movement with a head-mounted display. AR studies to be included required hardware that enabled interaction with the environment. Smartphone AR software that could not allow appropriate user interactivity and could not be clarified as an immersive experience was excluded.

Exclusion criteria were pre-determined and seen in Table 1.

Table 1. Exclusion criteria for review.

Exclusion Criteria	Rationale
Patient education	Study aims to explore HCP education
Use of new prototypes/designs/haptics	These papers focus on technological development
Reviews	Prior material was considered in background
Virtual patients or virtual learning environment studies	Falls outside MR scope for this study
VR or AR without user interaction	This review aimed to focus on immersive environments
Animal studies	Human participants only are relevant
Gamification without medical content	Serious gaming is within MR domain, however, there needs to be a medical focus for purpose of review
Non-medical	By definition this excluded vets, dentists, pharmacists, and optometrists

3.4. Search Terms

The string search used in WoS:

ALL = (virtual reality OR VR) OR ALL = (Mixed reality OR MR) OR ALL = (augmented reality OR AR) OR ALL = (mixed reality OR MR) OR ALL = (360 video) AND ALL = (health-care OR medic*) AND ALL = (learn* OR educat* OR train*) AND ALL = (Student).

Boolean operators and partial words maximised article inclusion.

3.5. Selection Processing

A total of 5102 articles were filtered after the removal of duplicates using Rayyan software. Author CJ screened all 4751 publications for suitability using the PRISMA scoping review guidelines [51]. Of these, 318 were sought for full article-level screening: 73 met the exclusion criteria.

The process detail is seen in the PRISMA flow diagram (Figure 2) and was supported by Rayyan online management system [52] for screening abstracts and duplicate removal.

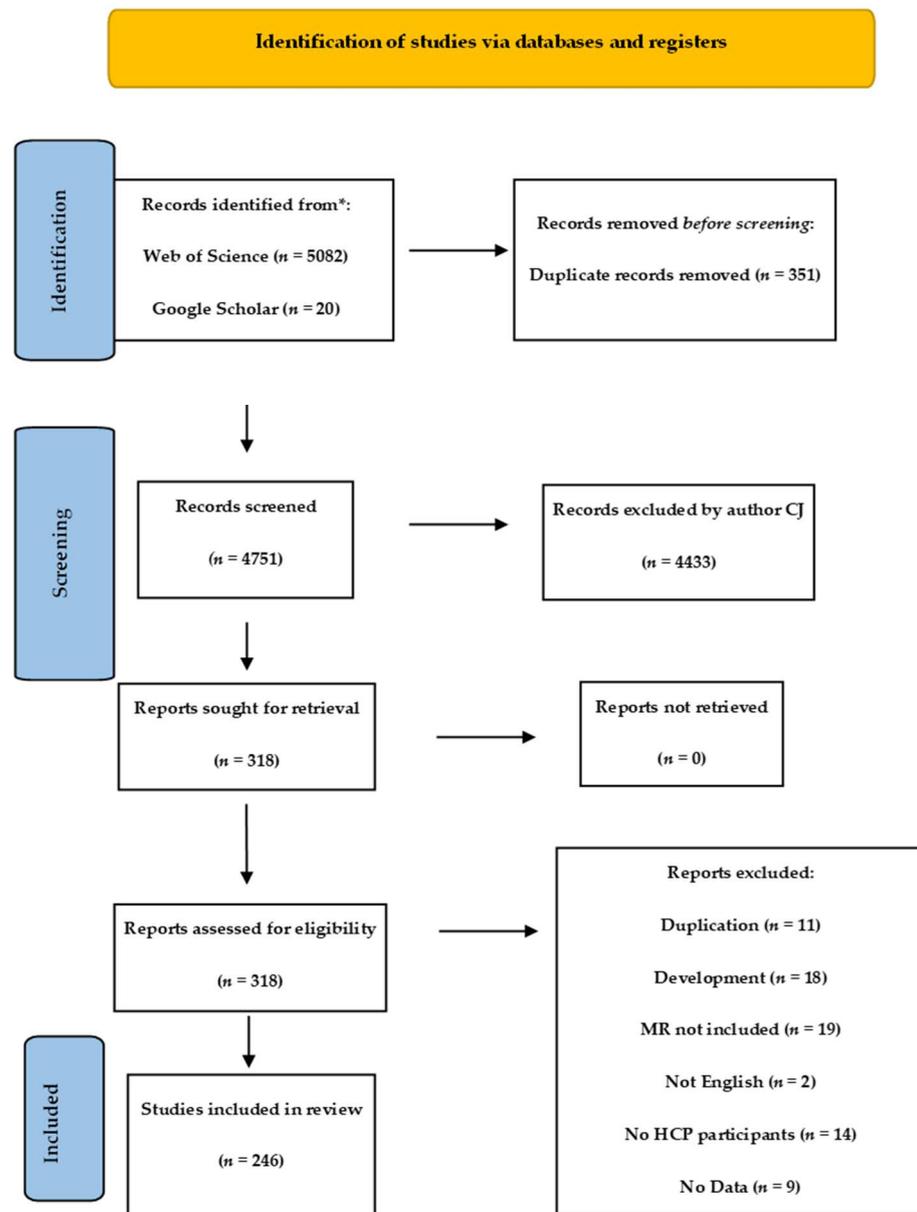


Figure 2. PRISMA flow diagram of review.

3.6. Paper Quality Assessment Instrument

All papers fulfilling the inclusion criteria were scored according to the Medical Education Research Quality Instrument (MERSQI). Internal consistency, interrater reliability, and criterion validity were demonstrated for the 10-item instrument [53,54]. In MERSQI, 6 domains reflect the study quality: study design, sampling, data type, the validity of assessments, data analysis methods, and outcomes. Total MERSQI scores range from 5 to 18. There was no cut-off score for exclusion as supported by prior MERSQI tool development [54].

Author CJ independently reviewed and scored all papers, and GF scored 20% of papers for a comparison of score agreement and discussion for those with a different score.

3.7. Charting the Data

The included papers were separated into 3 MR subtypes: VR, AR, and 360 videos. A further subtype consisted of papers that had outcomes or discussions relating to learning theory regardless of the type of technology employed. The data extracted included: medical specialty, year of publication, country of origin, sample size, study outcomes, and reviewers' judgment on paper limitations. Full data extraction is available in the Supplementary Material: Table S1 of VR, AR, 360 videos, and learning theory studies.

3.8. Data Analysis

Total MERSQI, including the study validity domain, scores were summarised with descriptive statistics.

ANOVA with Bonferroni comparisons was used to examine associations between total MERSQI scores and the MR study subtype. A p -value of < 0.05 was considered to be significant, and data were analysed with StatsDirect version 3.3.5.

4. Results

4.1. Study Characteristics—Research Questions 1 and 2

Studies eligible for review were published from 2002 to early 2022 in peer-reviewed sources.

Table 2 summarises the 246 paper characteristics, which together had 11,427 participants. The highest proportion of study participants were undergraduate students, 10,801/11,427 (92%), and the proportion of qualified physicians or surgeons was 626/11,427 (5.5%). Papers recruiting qualified doctors were predominantly in the VR subgroup ($n = 568/7879$), which reflected this being the largest paper group of 163 (66%).

Table 2. Summary of characteristics of reviewed papers. Learning theory subgroups comprises papers from VR, AR, and 360 themes.

Theme Subtype	Number of Studies	Total Participants (Number of Qualified HCP)	Sample Size Range	Average Sample Size
VR	163	7879 (568)	5–289	98
AR	68	2520 (10)	10–292	55
360	15	1028 (48)	34–207	69
Learning theory	44	2342 (115)	10–200	56

A trend of increasing research interest with time was evident in all fields of MR. There was a lag in publications with AR and 360, with the first publications in the years 2010 and 2015, respectively, which reflects the availability of AR and 360 technologies. The most active years were from 2015 to early 2022, with research in all subtypes, as seen in Figure 3. Additionally, the selected papers that considered learning theory demonstrated this pattern.

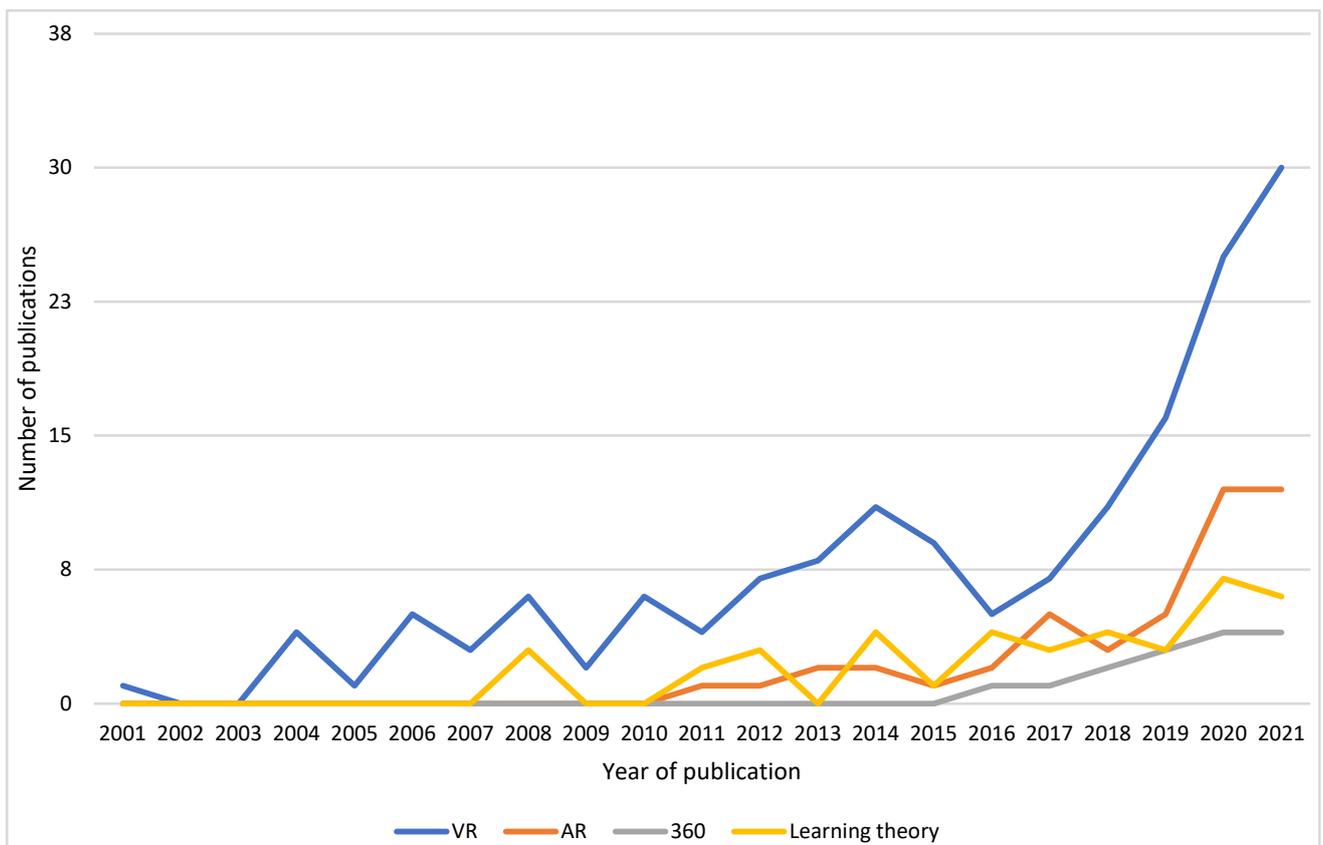


Figure 3. Number of publications by year for VR, AR, 360, and learning theory domains.

There were 18 varied specialties involved as settings for the technology, particularly general surgery ($n = 51, 31\%$) in VR and anatomy ($n = 13, 28\%$) in AR (Figures 4–6).

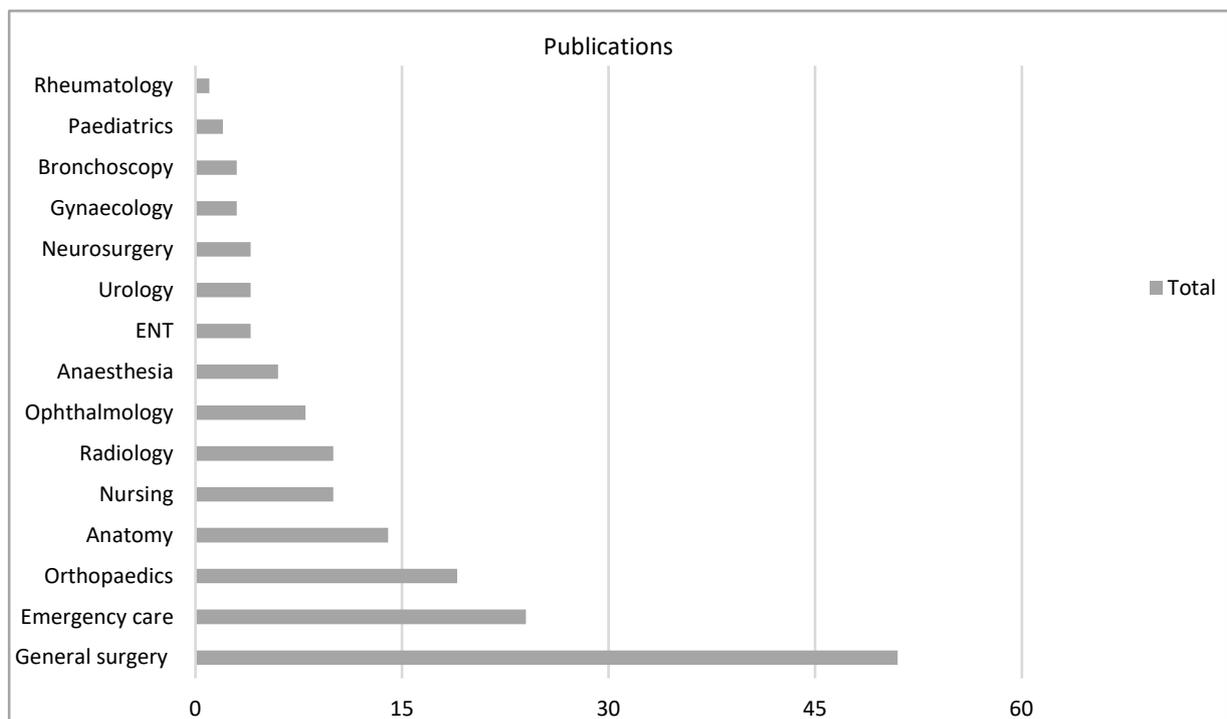


Figure 4. VR subgroup according to specialty.

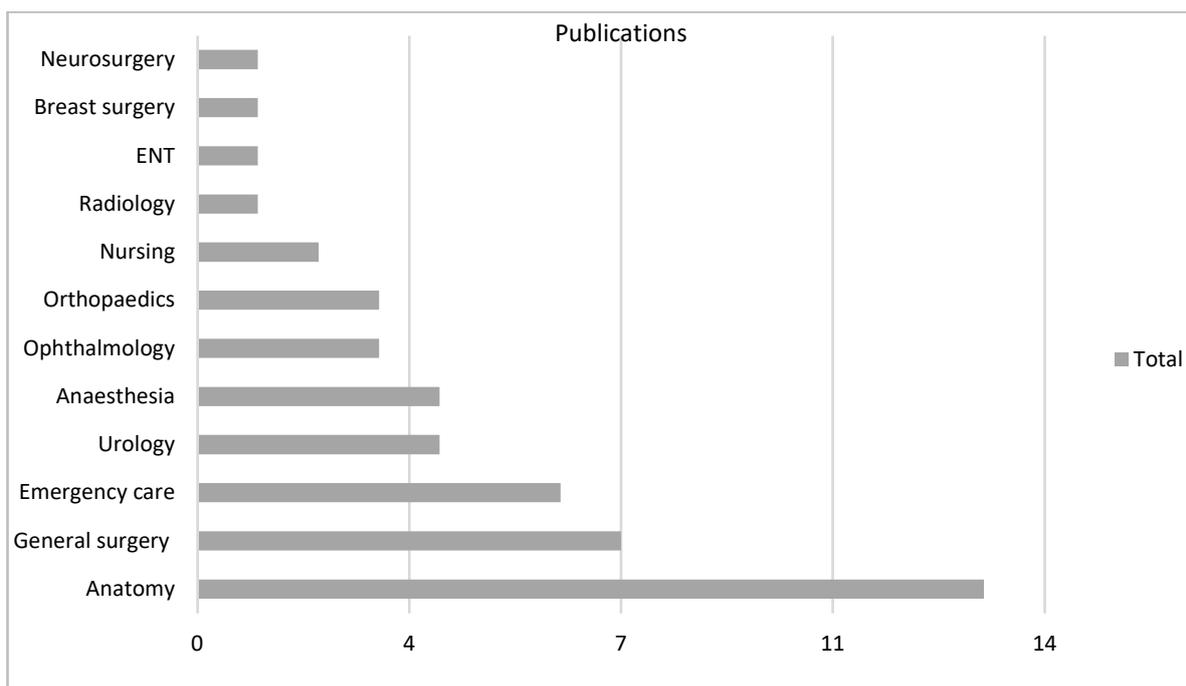


Figure 5. AR subgroup according to specialty.

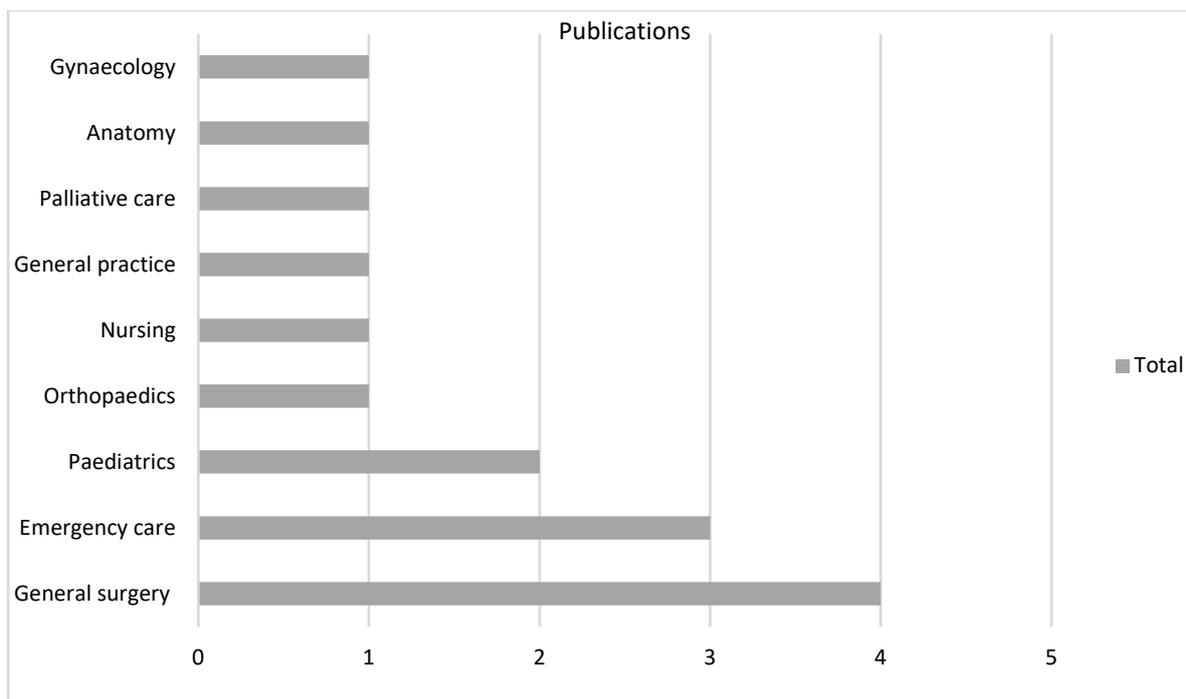


Figure 6. 360 subgroup according to specialty.

There were over 30 countries publishing work in English on immersive technology in medical education (Figure 7). The USA accounted for 25% of the published work ($n = 60$). In one paper, the country of origin could not be established.

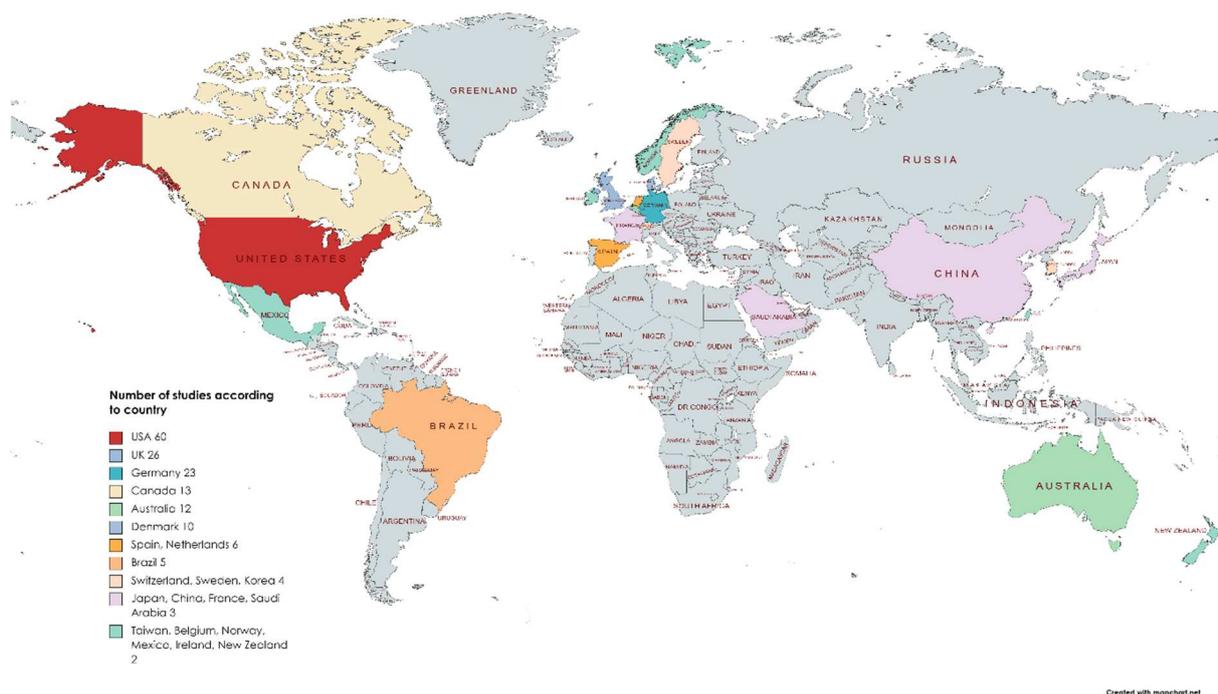


Figure 7. Count of eligible articles of all subtypes by country.

4.2. MERSQI Evaluation—Research Question 3

The interrater reliability (ICC) of MERSQI scores between the two reviewers was 0.8, indicating good ICC.

The mean (SD) total MERSQI score was 10.58 (2.33) for the VR group, 9.57 (3.17) for the AR group, and 8.86 (2.66) for the 360 group. The highest mean score of 10.85 (2.03) was calculated for the learning theory group. Furthermore, the instrument validity scores (maximum 3) were highest in the learning theory group at 1.94 (0.75). The VR group mean was 1.25 (1.42), the AR group was 1.30 (1.50), and the 360 group was 1.28 (1.54).

A one-way ANOVA revealed the total mean MERSQI scores were statistically different between groups ($F(3, 262) = [4.66], p = 0.003$). Post-hoc Bonferroni comparisons of mean MERSQI scores were significantly higher for VR than AR (1.0, 95% CI: 0.20 to 1.82, $p < 0.05$) or 360 (1.8, 95% CI: 0.51–3.12, $p < 0.05$). The highest mean was in the learning theory group, but it was not significantly different from the VR group (0.23, 95% CI: -0.60 – 1.06 , $p = 0.57$). When comparing instrument validity scores, the learning theory group reported higher validity when compared to all groups (0.64, 95% CI: 0.15–1.14, $p < 0.05$). Table 3 below summarises the descriptive statistics on MERSQI scores for all subgroups.

Table 3. Descriptive statistics of MERSQI scores and instrument scores for all groups.

Variables	Score VR	Score AR	Score 360	Score Learning Theory	Instrument Validity Score VR	Instrument Validity Score AR	Instrument Validity Score 360	Instrument Validity Score Learning Theory
Mean	10.57	9.58	8.86	10.85	1.26	1.30	1.29	1.88
Standard deviation	2.33	3.17	2.66	2.03	1.42	1.50	1.54	1.40
Standard error of mean	0.18	0.47	0.71	0.31	0.11	0.22	0.41	0.22
Upper 95% CL of mean	10.93	10.52	10.39	11.48	1.48	1.75	2.18	2.32
Lower 95% CL of mean	10.21	8.63	7.32	10.21	1.03	0.86	0.40	1.44
Maximum	15.50	14.50	13.00	14.50	3.00	3.00	3.00	3.00

4.3. Instruments in VR

There were over 50 different methods of collecting participant data in the VR group, with task performance being the most commonly reported measure, $n = 71/163$ (44%). Task performance was a term applied to any study in which the VR system recorded data

of participant interactions within a clinical procedure, for example, laparoscope extent of movements, reaction times, and accuracy of movement. The composite scores from simulators were specific to the simulator and the selected procedure of interest. This created heterogeneous outputs that, on consideration by review authors, were not suitable for meta-analysis.

Twenty studies (12%) created multiple-choice questionnaires (MCQ) as a measure of learning, objective scores by clinical assessors (i.e., OSCE) were used in 14 studies (8.5%), and a global rating scale was used in 12 papers (6%). Five papers were qualitative, and in these, a thematic analysis was undertaken.

Table 4 below separates all of the measures into four domains: procedural or objective tool, learning assessment tool, subjective measure of experience, and physical measurements recorded. This broadly relates to study outcomes within the MERSQI instrument that scores the paper to whether the outcomes are opinion, knowledge, skills, or patient health outcomes. Supplementary Table S1 for this study include this data.

Table 4. Summary of the measures in domains of primary outcomes of the VR group with number (no.) of papers reporting this tool.

Procedure/Objective Tool	No.	Learning Assessment Tool	No.	Subjective Measure	No.	Physical Measure	No.
Task performance	72	MCQ	20	User experience questionnaire	4	Heart rate	2
Global rating scale (GRS)	12	NASA TLX	4	System usability scale (SUS)	3	Cortisol levels	1
OSCE	7	Visual-spatial ability	3	Confidence rating	3	EEG	1
Objective score by an assessor	4	Cube subtest of standard intelligence test	1	Presence questionnaire	3		
Objective structured assessment of technical skills (OSATS)	3	Intrinsic motivation inventory	1	Difficulty rating	1		
Arthroscopic surgery skill evaluation tool (ASSET)	2	Learner evaluation questionnaire (LEQ)	1	Instructional materials motivation survey (IMMS)	1		
Reaction time	2	Lasater clinical judgment rubric	1	Interprofessional socialization and valuing scale (ISVS)	1		
Procedure checklist	2	Learning survey	1	iGroup presence questionnaire (IPQ)	1		
Basic life support	2	Mental rotation test	2	Scale of experience	1		
Bronchoscopy skills and tasks assessment tool (BSTAT)	2	Learning style (MIDAS)	1	Nursing anxiety and self-confidence	1		
Cannulation score	1	Pictorial surface orientation test (PicSOR)	1	Serious gaming usability scale (SGUS)	1		
Xray procedure score	1	Social validity questionnaire	1	Simulation design questionnaire	1		
Neurological exam	1	Spatial orientation test	1	Student satisfaction	1		
Cochlear implant assessment (CISAT)	1	Stress mental readiness form	1	VR experience survey	1		
Score of communication	1	Students' engagement in school four-dimensional scale	1				
Emergency medicine crisis resource management (EMCRM)	1	Technology acceptance questionnaire	1				
Modified Welling score	1						
Total measures	115		41		23		4

When analysing the papers' MERSQI scores by type of outcome measure used, the VR task performance total mean score was 11.01 (SD 1.69), with a mean instrument validity score of 1.0 (SD 1.3). Assessment of learning by an OSCE yielded a total mean score of 10.67 and an instrument validity score of 1.42 (SD 1.5). The MCQ assessment studies had a total score of 9.54 (SD 1.8) and an instrument validity score of 0.58 (SD).

Studies outside of the most commonly used measures collectively had a total MERSQI score mean of 10.15 (SD 2.48) and a mean instrument validity score of 1.76 (SD 1.43).

The post-hoc Bonferroni comparison of mean MERSQI scores of studies that used task performance were significantly higher than those using MCQs (0.62, 95% CI: 0.30–2.74, $p < 0.05$), but there was no difference in the instrument validity subscales (0.41, 95% CI: –0.80–1.64, $p = 0.50$).

The task performance group had the highest mean total scores (1.41, 95% CI: 0.65–2.16, $p < 0.05$) compared with other studies. However, the instrument validity was the highest of all groups in those outside task performance, OSCE, and MCQ papers (<0.05).

4.4. Instruments in AR

There were 20 instruments utilised in collecting participant data in the AR group (Table 5), with MCQ being the most commonly reported measure, $n = 13$ (19%). A system usability scale was reported in six studies and five other studies reported a usability assessment. Task performance was only featured in seven papers (10%). One paper was a qualitative paper in which thematic analysis was undertaken.

Table 5. Summary of the measures in domains of primary outcomes of the AR group with number (no.) of papers reporting this tool.

Procedure/Objective Tool	No.	Learning Assessment Tool	No.	Subjective Measure	No.	Physical Measure	No.
Task performance	7	MCQ	13	System usability scale	6	EEG	1
OSCE	3	Mental rotation test	2	Usability questionnaire	2		
Anesthetist non-technical skills (ANTS)	1	NASA TLX	1	AttrakDiff 2 usability score	2		
Angle of hip surgery	1	Cambridge brain psychomotor test	1	Profile of mood states (POMs)	2		
Global rating scale	1	The nine-point cognitive load scale	1	Subjective rating	2		
Kakarla scale of EVD insertion	1	Simulation effectiveness Tool	1	Experience survey	2		
Skill competency checklist	1			Overall perception of headset score	1		
				Evaluating technology-enhanced learning materials (ETELM)	1		
				Attention/confidence with scale	1		
				Written feedback	1		
				Instructional measure of motivation survey	1		
Total	15		19		21		2

Table 5 shows the summary of a breakdown of measures used in the AR group, which is presented as the VR group was. Similarly, a statistical analysis of the group characteristics follows.

Papers that selected MCQ as a measure of learning had a MERSQI mean total of 9.78 (SD 2.40), with an instrument score of 1.71 (SD 1.60). Those focusing on usability had a MERSQI total mean of 8.50 (SD 3.21) and an instrument score of 1.50 (SD 1.64).

There was no statistical difference when comparing MCQ and usability MERSQI scores (1.29, 95% CI: $-0.88-3.45$, $p = 0.24$) or their instrument validity scores (0.21, 95% CI: $-1.95-2.38$, $p = 0.85$).

4.5. Instruments in 360

This group was the smallest (Table 6): of the 15 studies, 5 (33%) assessed knowledge by MCQ. One paper was qualitative and used ethnographic analysis, and no papers measured the biometrics of participants.

The MCQ group of papers had a mean total MERSQI score of 9.10 (SD 1.75) and an instrument domain score of 1.20 (SD 1.34).

Table 6. Summary of the measures in domains of primary outcomes of the 360 group with number (no.) of papers reporting this tool.

Procedure/Objective Tool	No.	Learning Assessment Tool	No.	Subjective Measure	No.
Task performance	1	MCQ	5	System usability scale	1
Simple triage and rapid treatment protocol (START)	1	Communication accuracy	1	Engagement scores	1
		NASA TLX	1		
		Ease to learn	1		
		Motivation and learning	1		
		Task unrelated images of thoughts	1		
Total	2		10		2

So far, we have seen the trends of research output in immersive technology in medical education, which featured a rise in publications in all subgroups. Additionally, by applying an assessment of paper quality, we see varied MERSQI scores with studies in the different technologies. The next section will provide the reader with a summary and synthesis of outcomes in the VR, AR, and 360 papers included in the review, followed by the learning theory group to answer the final research question.

5. Synthesis of Main Study Findings—Research Question 4

For transparency, the MERSQI score, with the total and instrument validity subscore in brackets, will be presented with the reported study in the synthesis, i.e., Le et al. work on a VR eye simulator was appraised with a total score of 13.5, and the task performance instrument had undergone testing in all three validity areas: structure, content, and relationship to variables (13.5/3).

5.1. VR Group

Task performance as a primary outcome is a composite score that is derived from computer-based simulators [55–62]. Creating a system-based score and broadly defining thresholds for acceptable performance allows the VR intervention to be tested.

Pre- and post-scores of task performance following a VR experience improved with participants from a medical background [58–103].

Studies that recruited a mixed background of medical students and qualified surgeons demonstrated a learning curve whereby novices improved most over successive VR sessions when compared to experienced surgeons [58–60,67,68,71,75,83,88,91,94,96,97,103–106]. Le et al. [71] conducted a multi-centre trial (13.5/3) on the EyeSi ophthalmological simulator and established content validity, with participants with greater experience achieving significantly higher total scores than those who were less experienced: level of training predicted performance ($p < 0.05$). In an experiment on VR training for gynaecological surgery (13/3) [97], there was a learning curve in which performance improved over the first 10 repetitions

across multiple metrics for all the exercises. Senior trainees performed the suture exercise significantly faster than the junior trainees during the first and last repetitions ($p < 0.004$ and $p < 0.003$, respectively). It was also noted that features such as depth perception and interaction of instruments in the laparoscopic VR surgery model added realism (13/3) [102]. In a final session of a course of VR ophthalmology training (13/3), the highest median score was achieved by the most experienced vitreoretinal surgeons ($p < 0.01$). However, two of the six modules of VR performance did not correlate with experience [103]. An early plateau of improvements with experience was found with robotic VR simulation (13/3) [107], which could indicate results are dependent on the task under experiment and a ceiling effect occurred (13/3) [82]. Harden et al. [108] also noted that if VR was used prior to actual robotic surgery, then a ceiling effect occurred, and no further improvements in task performance measures were demonstrable when doing real-life robotic surgery. Although there was a higher perceived improvement by participants, the mix of learning curves and a small sample makes it difficult to make comparisons (14/3). A similar disconnect between task performance and perceived performance occurred in a VR model of basic life support training [109] when perceived performance improved, but flow time for conducting life support compressions was worse for VR compared to traditional training. Berg et al. [110] demonstrated a correct order to the ABCDE approach in life support and described VR training as non-inferior to traditional practice equipment (14.5/3).

Global rating scales (GRS) are a procedural checklist for an objective assessment of performance, which has high interrater reliability through processes of the instrument validity [111–113]. It is an anchored scale that aims to score from novice to expert with respect to the skill in question. In a study using VR as a tool to teach intubation [114], there were significant improvements in GRS scores in the expert review of real intubations in the experimental VR group when compared to a control group (15.5/3). Mulligan et al. [113] compared GRS in arthroscopy using VR and were able to differentiate between the novice, intermediate, and expert groups (12.5/3). The use of GRS in VR demonstrated improved performance in all studies within this review [63,111–119].

MCQs have been used to assess educational outcomes [120], and in comparison to control groups undergoing standard teaching, VR improved the MCQ scores of participants' learning more than standard teaching [121–126]. In a large study of 168 participants using video teaching as a control group versus VR to train medical students in the assessment of infants in respiratory distress, recognition of respiratory status, and the need to escalate care was significantly better with VR ($p < 0.01$). In this study, Zackoff et al. [126] reported a great interpretation of a patient's mental status (14/3). Whilst exploring the benefits of VR bronchoscopy training, Gopal et al. [127] demonstrated a 55% gain in students' bronchial anatomy knowledge (12/3).

Subjective measures that consider the participant reported perceived benefits as the primary outcome are valuable as they can establish additional information surrounding satisfaction, attitudes, and perceptions that are outside the scope of other outcomes [54]. A strongly positive experience was found in more than 50% of users of VR in medical education [128–130].

Evaluation of an individual's physiological response while experiencing high cognitive demands with subsequent inference on performance quality is an established concept [131] and can be applied to medical education and simulation. Using an electroencephalogram (EEG), researchers matched increasing cognitive load with participants' surgical performance in VR [132] (11/2). Alternatively, measuring serum cortisol levels as a marker for stress was used in a stress-inducing experience of a mass casualty assessment in VR [133]. Serum cortisol levels did not rise as high in the VR group when compared to in-situ simulation (13.5/3). Lower heart rates were also seen in VR simulation compared to live simulation [134], although there was no difference in the accuracy of allocating correct triage (14.5/3).

5.2. AR Group

Task performance measures have been used to measure the impact of AR. A head-mounted AR display improved the pattern and accuracy of breast examination of novices, as measured by task performance [135]. In this study, 69 participants went on to perform real patient examinations (14.5/3). Projecting an anatomical pathway for accurate liver biopsy needle insertion in an AR model reduced failure rates from 50% to 30% [136], which included surgeons in the sample (10/3).

Head-mounted AR displaying real-time video as a means of instruction has been shown to improve procedural skills, such as suturing [137] and joint injection [138], though both of these studies used measures scoring 0/3 on validity. However, when using a validated instrument for joint injection, Van Gestel et al. [106] found the AR-guided extra-ventricular drain placement as performed by inexperienced individuals was significantly more accurate compared with freehand placement on models (14.5/3).

Studies measuring ARs impact using MCQs had mixed outcomes, with few papers presenting the validity of the instrument. Those that used validated measures included a study on AR head and neck anatomy teaching, in which there was a significant improvement in scores (15% higher, $p < 0.05$) compared to traditional screen work [139] (12/3). Moro et al. [8] had two experimental groups of mobile AR and head-mounted AR, which found no significant differences in scores between lessons on brain anatomy and physiology teaching (12/3). In assessing anatomy education by randomising groups of students to either a screencast or AR, both groups increased post-test; however, the screencast group had the highest gains [140] (12/3). In a study that focused on respiratory microanatomy, there was no significant difference in AR anatomy when compared to teacher-led microscope slides on initial MCQ scores [141]. Moreover, a follow-up test found that the experimental group assessment maintained a higher score compared with the control group, with averages of 86.7% and 83.3%, respectively (9/0).

The system usability scale (SUS) is a reliable measure to evaluate the usability of a wide range of products and services [142], which include technology in medical education [143]. After watching a projection of a video in a head-mounted display to demonstrate accurate catheter insertion, participants were assessed on performance and usability [144]. While the control group rated teacher instruction higher with the SUS, OSCE performance was better in the experimental AR group (14/3). Similarly, lower SUS scores were found when assessing AR in mass casualty triage, with a preference for simulation without immersive technology [145]. Nurses additionally scored the AR experience lower in the simulation effectiveness tool (9.5/3). However, augmented lumbar puncture training demonstrated greater SUS over a traditional monitor displaying an instructional video [138] (10.5/3).

EEG recordings of participants learning neuroanatomy from a textbook versus AR 3D images were not significantly different [146]. Short-term memory assessment found that textbook recall was superior to AR; however, long-term memory assessment found that the retention of spatial information was maintained in the AR group only (14.5/3). Several themes emerged when interviewing the same students who commented on spatial advantages to AR but expressed dissatisfaction at not being able to draw 'like on a book', and technical difficulties arose.

5.3. 360 Group

In an RCT randomising medical students to learn suture knot-tying via a 360-degree video or a traditional monitor, greater accuracy of knot tying was associated with the 360 video (median score 5.0 vs. 4.0, $p < 0.05$) [147] (9/0). Lowe et al. [148] conducted an observational study of triage training using 360-degree video in an out-of-hospital setting in paediatrics. The triage protocol scored on field intervention selected by users following the video, and scores were able to discriminate between levels of emergency medicine experience (11.5/3). Based on a 5-point Likert scale (1 lowest, 5 highest score), participants felt the 360-degree video experience was engaging (median = 5) and enjoyable (median = 5). The majority felt that 360 video was more immersive than mannequin-based simulation

training (median = 5). It should be noted that validity had not been demonstrated for this questionnaire.

A study comparing a traditional interactive lecture to 360-degree videos of effective communication with patients found higher MCQ scores out of 20 in the experimental group (17.4 vs. 15.9, p -value < 0.001) [149]. In addition, students underwent an OSCE with simulated patients and achieved higher scores in the 360 group. The overall satisfaction of the 360 experience was rated 7.36/10, and 93% agreed with the statement that 360 can be used in medical education (11/3). As with the work of Lowe et al. above, instrument validity on subjective measures was not demonstrated. Student communication featured in a sensory deprivation study using 360 to simulate deafness. In this qualitative work, themes emerged on empathy, and as an experience, it granted the user more than merely watching a video [150].

MCQ performance was non-inferior in comparing the 360-degree video to the hybrid mannequin simulation in a clinical scenario of opioid overdose and administration of naloxone [151] (11/3). Similarly, in gynaecology education, the 360-degree video had no statistically significant difference in MCQ scores when comparing conventional teaching [152]. Indeed, those in the 360 group wanted more 360-degree content (8/0).

The system usability scale (SUS) results were high in the 360-degree presentation of sepsis prevention, with it being relaxing and acceptable functionality [153]. A cognitive load questionnaire was reported with the imposed mental demand categorised as a medium mental demand, which implied an acceptable mental load range to learn (9.5/3).

5.4. Learning Theory

Pedagogical concepts and theories were identified in an extensive number of studies in this review [11,17,72,84,88,90,93,132,134,146,154–184].

Hu et al. [163] demonstrated that after completing a course of ultrasonographic training in VR, there was improved performance in identifying anatomical structures. Authors applied the principles of the psychomotor domain of Bloom's taxonomy [185] and theorised that VR anatomical training improves learner's ability to use sensory cues to guide motor activity (Perception), increase action readiness (Set), and the ability to learn complex motor skills by imitation, trial, and error (Guided response). Hence, VR can aid the user in appropriate procedural clinical instrument movements required for medical investigations.

Furthermore, if using VR to teach pelvic anatomy, the recalling of information aligns with the 'remember' level, and functional questions align with 'understand' by inferring a response [183]. The complexity of understanding is fostered by VR users comparing structures spatially in the environment. In this study, matching visuospatial ability with a physical model found equivalent short-term recall and matched scores on mental rotation tests to immersive technology. However, those with the lowest visuospatial scores fared best on a real model. Sommer et al. [11] demonstrated visual-spatial ability as a predictor for performance in a VR simulator, those with low visuospatial results caught up with those with pre-existing high scores after two training sessions. Three-dimensional visualisation is not limited to VR, as a study of AR neuroanatomy was found to be non-inferior when compared with traditional cross-sectional images when testing MCQ and mental rotation scores [162]. Thematic analysis of a group of students' comments found limitations with the AR technology, which introduced complications of holding a device and could not fully replicate a cross-section at all angles (unlike real anatomy). Hence, MR as an instructional tool may need to consider individual abilities and preferences.

The work by Henssen et al. [162] on learning neuroanatomy additionally applied a cognitive load measure, which had similar scores between AR and real cross-sections. Cognitive load (CL) theory suggests that the designs of instruction impose cognitive load on learners' limited working memory and that the CL, in turn, influences learning outcomes [186]. CL scores were much higher when using real cross-sections, either because AR created a less demanding learning tool or the cross-sections were too complex for the level of the student. When a 9-point CL score was utilised in a study of AR anatomy

versus traditional lessons, higher MCQ results with reduced cognitive load were found in the AR group [169]. The National Aeronautics and Space Administration Task Load Index (NASA TLX) can also be used to assess workload, which can be seen to reduce scores in VR simulations of surgery [184]. Mass triage scenarios were tested using VR and live simulation by Mills et al. [134]. Lower NASA TLX scores, reduced heart rate, and reduced simulation design scale scores were associated with the VR simulation. The heart rate measurement was to capture stress levels. At present, subjective measures, such as NASA TLX, are the most common methods of measuring cognitive load. Advantages exist with objective psychophysiological measurements, such as EEG, that can be used to detect subtle changes in alpha and beta waves that can be inferred as load change [132]. Using an EEG, scores from the NASA-TLX score, and heart rate, it was demonstrated on three fundamental surgical tasks in VR that CL reduces as participants' skill proficiency improves with subsequent practice. Yu et al. [132] considered that developed skills might indicate a lower cognitive load.

EEGs have been used, for example, in the comparison of anatomy learning of visualisations from textbooks, computer screen 3D visualisations, and in an MR device. EEG waveforms were interpreted as a marker for wakefulness and potential CL and did not differ between any of the three groups. An anatomy test indicated better performance with textbook and worse with MR, though subsequent retesting at over a month showed the MR group had higher retention, especially in spatial recall [146]. There may be added benefits to participants in being able to manipulate anatomy images, hence, the improved spatial recall in the MR group.

Students' motivational reactions to self-directed learning of anatomy using AR have been studied. Researchers applied the Instructional Measure of Motivation Survey (IMMS), which is based on Keller's ARCS model. In this study, there was no statistical difference in IMMS scores between participants learning anatomy from AR or cross-sections [162]. VR was found to be non-inferior in anatomy learning compared to textbook learning [178]. However, in this study, IMMS scores and self-reported perceived learning were significantly higher in the VR group ($p < 0.01$). An alternative motivation subjective questionnaire that captured interest, value, and importance, based on Ryan's intrinsic motivation inventory [187], enabled authors to see if multiple-player VR could benefit anatomy learning. Both single-player and multiple-player VR stimulated and motivated users most when compared to using textbooks [160].

Constructive learning describes the active process of learning in immersive environments, as motivation, attention, and knowledge forming are connected. MR creates the perception of being dissociated from the real-world (presence) and the user's loss of time (flow). Sense of presence can be associated with the realism of the image and user control and reduces with prolonged exposure [188,189]. Measuring presence in the IGroup Presence Questionnaire (IPQ) in VR gave higher scores in a semi-autonomous ward environment compared to a PC setup [17]. With more experienced doctors describing a feeling of greater involvement and immersion. Additionally, an immersive tendencies questionnaire (ITQ) applied to a mass casualty immersive VR simulation found a high level of sense of presence and was correlated with individual immersion propensity [190].

Deliberate practice is an instructional method that relates to skill acquisition [191]. This method is based on four domains: motivation, goal orientation, repetition, and focused feedback, to create an iterative cycle of progressing from a novice to an expert. Furthermore, these procedures can be broken down into a series of microskills that have deliberate steps [192] and eventually can form mastery learning. Feedback can be implemented at different stages of the learning curves demonstrated in VR [84]. A study of the complex surgical task investigated by Paschold et al. [84] found that separating lower performers in VR and providing tailored instruction could close the gap in performance to higher performers early on. Learning curves, as measured by a procedural task performance, are demonstrated in a number of studies [58,72,75,93,98,117,177,193–195]. There is varia-

tion in the number of repetitions to accurately perform a procedure at a matched expert measured performance.

Kolb's experiential learning theory, in part, maps on the deliberate practice model as both represent an iterative cycle that experience is converted to knowledge and skill, although the experiential cycle does not explicitly refer to feedback. A simulation curriculum in VR was assessed using experiential learning as the model for the development of a training program [175]. A mixed methods study evaluated the clinical judgement and competency of nurses in the virtual environment. Self-reported measures found the learning experience transformed the thinking process of critical thinking, clinical judgement, and metacognition. Similarly, using group self-practice as a form of instructor feedback in completing an ABCDE (airway, breathing, circulation, disability, and exposure) assessment in VR showed it was non-inferior to practice with physical equipment [196]. Additionally, the VR software provided feedback to the users related to knowledge and passive positions.

Finally, learning styles have been described as how we might learn in different ways and how technology can be designed to accommodate our differences [181]. Multiple intelligence theory complements learning styles and holds that learning is pluralistic. Categorising these and relating them to surgical experience found that the kinaesthetic learning style was associated with the best VR laparoscopic performance. The majority of novices exhibited a malleability to learning, and styles were convergent and accommodative, which favour hands-on experience.

6. Discussion

A total of 246 papers were selected in this review for a more detailed appraisal from an initial search resulting in 5102 papers.

The research on MR is steadily increasing (Figure 2) and at a series of points could be interpreted as matching the Gartner Hype Cycle as Tang et al. [2] proposed. In particular, the period 2014–2016 saw a downward turn in output in VR, which could be understood as the point of disillusionment whereby this phase is identified as a time with greater difficulty applying the technology to current medical education needs. The following increase in publications is then a re-emergence of ideas as technology evolves and acceptability in the medical education arena increases, though the time period evident in this review in relation to technology adoption differs from that described by Tang et al. [2]. Additionally, the differences in graphs could relate to search criteria, and this review had a greater number of included papers. It may be simpler to consider the process as waves, with technological utilisation depending on affordability, availability, accessibility, and appropriateness according to the region and specialty [197]. This considers digital adoption not only in the context of users but also in wider influences, such as the economy and COVID-19. Extrapolating from Figure 2, there will be an increase, with new technology becoming a trigger for successive interest within the spectrum of MR.

A total of $n = 11,427$ participants were studied across the 246 papers. However, only 5% ($n = 573$) were qualified medical professionals, with the overwhelming majority of participants being students ($n = 10,801$, 92%). This proportion is lower than other reviews have identified [198]; however, a significantly higher number of participants contributed to the study synthesis when compared to other reviews. There is a challenge in trying to study the population of medical professionals [199] when medical students offer a relatively easily accessed and well-motivated group, usually attached to a setting in which there is a culture of and structure for research.

General surgery teaching in VR and anatomy teaching in AR were the most common applications, which mirrors other reviews [2,198,200]. Eighteen medical specialties were evaluated with MR, and this represents a broad mix of different areas of medicine testing the utility of these technologies in education. Of the 30 countries in the collected data, the USA was the most common country of research origin (25%). A possible explanation is the mandatory requirement of simulation training in North American surgical training [88].

How we assess learning in MR environments was a primary focus of this review. The MERSQI instrument was selected for this purpose. Mean MERSQI scores from subgroup analysis were highest in papers on VR (10.57) and learning theory (10.85), then AR (9.58) and 360 (8.86). MERSQI scores were assessed as lower than in the recent review of immersive technology in healthcare student education by Ryan et al. [198]. However, MeSh terms differed, and 29 papers were analysed compared with 246 in this review. An explanation for subgroup differences could relate to Figure 2, which tracks trends in publications. VR is the most established research area, whilst 360 is the newest. Although development papers that only assessed the technology without measuring an area of educational benefit were excluded, the lower scores reflect the pilot study nature of 360 papers and the lower instrument validity scores found. This is the first review to categorise subgroups in MERSQI and to critically look at instrument methods to assess learning in detail. Instrument validity scores were statistically higher ($p < 0.05$) in the learning theory group (1.88) and similar in the MR subgroups (range 1.26–1.30). This confirms previous research that studies do not fully report validity assessments for the instruments used [54,201–203]. In a review of the quality of evidence in serious gaming in medical education, only 28.5% of instruments had internal validity tests [204]. The MERSQI average score of 10 was similar to this review. It is worth being aware of the validity of instruments used to capture outcomes when assessing evidence in this (or any) field.

Overall, in this review, the use of immersive technology has been associated with improved learning, echoing the association of high-fidelity simulation-based medical education with learning [205]. However, MR must be studied thoughtfully. For example, feedback is necessary for the development of non-technical and technical skills but cannot be fully derived purely from computer-derived procedural scores.

A major challenge in studying MR is that many different categories and examples of outcomes exist. Task performance was the most commonly reported outcome, followed by OSCE scores and GRS. MCQs were the most common method of knowledge assessment. MCQs and task performance measures had the lowest MERSQI scores for validity. The composite scores generated by simulators are individual to the device and task at hand. Furthermore, a high score in a simulator cannot be assumed as an assessment to reflect real-world performance. If a novice and an experienced surgeon are at similar levels by the 10th repetition, one cannot infer a novice is as capable and experienced in surgery but only on the metrics of the task performance analysed (95). The final MERSQI question related to the Kirkpatrick level [206], with the highest score achieved if 'patient benefit' was an outcome. While difficult to demonstrate, patient benefit was shown in only 5% ($n = 8$) of papers in VR, compared to the impact on knowledge/skills being measured in 66% ($n = 111$) of papers.

GRS and OSCE had higher instrument validity scores as those measures more commonly than others had undergone validity testing in all three domains (structure, relationship to variables, and content). Numerous other instruments had been previously studied, for example, OSAT, NASA TLX, SUS, IMMS, and psychophysical EEG data.

AR and 360 subgroups were of a similar nature, with task performance and MCQ as the primary outcomes. There was no difference evident in validity scores between the measures used in AR and 360 studies, probably partially due to the smaller size of these two groups.

Limitations and biases exist in the field. For example, there is a high proportion of novice users in immersive technology, skewing the samples tested. Learning as tested with mental rotation tests and procedural tasks in VR simulators [174] have produced contrary results [207]. Isolating the effect of MR per se may be difficult, as MR and traditional methods used as controls differ in several ways, such as variability in the information given by instructors or the MR software [10,100,178,195,208].

The learning theories considered by the papers in this review were extensive, and the author selected these papers during the full-text review as ones that detailed pedagogy and was a subjective admission and risk to bias. There was a finding of higher MERSQI

and instrument validity scores, which is not unexpected as the MERSQI is a measure of paper quality. The theories discussed in this review focus on different processes of learning. Presence and immersion can be captured with questionnaires, CLT can be measured by various instruments, e.g., NASA-TLX, and motivation is considered an important domain in how users approach simulation and practice. The multifaceted and abstract nature of learning will not be fully captured by one measure; however, this synthesis of papers supports the argument that immersive technology can create a learning environment, and the applications are broad. It also creates a dialogue on how we might begin to understand the evolution of this technology and where it might sit with traditional learning models, e.g., Dale's cone of experience.

This review has limitations, firstly, only English language papers were considered, and the reflexivity of the substantive author as a medical educationalist influenced which papers were included and how they were interpreted. Methodologically, the MERSQI is a well-established scale to rate paper quality; however, it groups knowledge and skill as one for assessing outcomes [204]. The validity assessment is limited to three traditional conceptions. Messick [209] described this viewpoint as fragmented and suggested six distinguishable aspects of construct validity (content, substantive, structural, generalisability, external, and consequential) [210]. Further work considering the newer concepts of validity assessment integrated into an updated MERSQI could yield a more detailed assessment of paper research methodology. Researchers could use a MERSQI framework in research design to ensure a robust methodology is in place prior to conducting trials. Furthermore, there was numerous additional information that could have been gleaned from MERSQI, which was not included in this review, i.e., the proportion of studies using RCT design. Analysis of the use of controls could add to the evaluation of immersive technology in education.

Although research questions were proposed, scoping reviews have methodological limitations and the risk of bias [15]. In particular, an assessment of bias was not conducted, and in this review, only one author conducted the screening process. The importance of identifying knowledge gaps was demonstrated in this review. The breadth of papers included in the synthesis was the largest to date in this field; however, the primary multi-source database (WoS) will invariably miss relevant studies. A MEDLINE search was included, and this incorporated most but not all PubMed publications. Further inclusion of educational databases, such as the Education Resources Information Center (ERIC) online library, would strengthen this review. Not all medical education journals are indexed in MEDLINE, and this needs to be considered when understanding the scope of this review. The definition of medical education varies, and the author's interpretation of this and the inclusion criteria of studies at the screening stage will invariably bias this review. Furthermore, a scoping protocol was not published alongside the OSF title and research aims. A priori protocol registration and publication are best practices.

The taxonomy of MR is expanding as new hardware and software are created. The inclusion of all available technologies was not possible in this review, and immersion is a subjective quality. The review appraised the existing evidence for the different MR technologies included and created a discussion about how we define learning.

The immersive field is rapidly developing. In the adoption of such technology, there is a need for rigorous validation that empirically evaluates the methodology employed and the assumptions made. The following recommendations are suggested to assist those evaluating MR as a learning tool:

1. Apply a methodology that is grounded in learning theory and literature.
2. Avoid basing conclusions on a newly created instrument but consider existing validated measures for further analysis.
3. Do not assume a novice versus expert investigation equates to a thorough study of validity.
4. Reframe the position on evidencing the superiority of TEL over traditional teaching methods and apply a non-inferior mindset.

5. Technology is a tool, and the way technology is used to supplement learning is varied. MR may help some learning environments but not others.

7. Conclusions

Presence and the motivated learner that employs a deliberate action of experiential practice are important concepts in relation to the application of immersive technology in healthcare education. This review adds to the literature in that technology can improve learning in a broad number of medical specialties with evidence at mid-level strength, according to the MERSQI tool. Currently, this study is the only review that uses a validated measure for evaluating paper quality in immersive technology for clinical-based education and a novel approach to present a synthesis of findings to a reader that incorporates paper quality scores. The pedagogy supports the use of immersive technology that can replicate clinical experiences and engage healthcare learners, with objective and subjective outcomes being evident. If we are to progress beyond a relationship with critically uncertain assumptions, rigorous methodology needs to be applied to future studies.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ime1020008/s1>, Table S1 of all synthesized articles and MERSQI scores.

Author Contributions: Conceptualisation, C.J.; methodology, C.J.; software, C.J.; validation, C.J. and G.F.; formal analysis, C.J.; resources, C.J.; data curation, C.J.; writing—original draft preparation, C.J.; writing—review and editing, C.J., M.W. and R.J.; visualisation, C.J.; supervision, M.W. and R.J.; project administration, C.J.; funding acquisition, C.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by South West Simulation Network Healthcare Education England as part of a larger body of work on appraising immersive technology in education.

Institutional Review Board Statement: Ethical review was not required for a systematic review of existing literature.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kotsis, S.V.; Chung, K.C. Application of the “See One, Do One, Teach One” Concept in Surgical Training. *Plast. Reconstr. Surg.* **2013**, *131*, 1194–1201. [CrossRef] [PubMed]
2. Tang, K.S.; Cheng, D.L.; Mi, E.; Greenberg, P.B. Augmented reality in medical education: A systematic review. *Can. Med. Educ. J.* **2020**, *11*, e81–e96. [CrossRef] [PubMed]
3. Baxendale, B.; Shinn, S.; Munsch, C.; Ralph, N. *Enhancing Education, Clinical Practice and Staff Wellbeing. A National Vision for the Role of Simulation and Immersive Learning Technologies in Health and Care*; Health Education England: Leeds, UK, 2020. Available online: <https://www.hee.nhs.uk/> (accessed on 20 February 2022).
4. Blair, C.; Walsh, C.; Best, P. Immersive 360° videos in health and social care education: A scoping review. *BMC Med. Educ.* **2021**, *21*, 590. [CrossRef]
5. Suh, A.; Prophet, J. The state of immersive technology research: A literature analysis. *Comput. Hum. Behav.* **2018**, *86*, 77–90. [CrossRef]
6. Swann, G.; Jacobs, C. Augmented reality medical student teaching within primary care. *Future Healthc. J.* **2021**, *8*, 15–16. [CrossRef]
7. Kuhn, S.; Huettl, F.; Deutsch, K.; Kirchgassner, E.; Huber, T.; Kneist, W.; Kuhn, S.; Huettl, F.; Deutsch, K.; Kirchgassner, E.; et al. Surgical Education in the Digital Age—Virtual Reality, Augmented Reality and Robotics in the Medical School. *Zent. Fur Chir.* **2021**, *146*, 37–43. [CrossRef]
8. Moro, C.; Phelps, C.; Redmond, P.; Stromberga, Z.; Moro, C.; Phelps, C.; Redmond, P.; Stromberga, Z. HoloLens and mobile augmented reality in medical and health science education: A randomised controlled trial. *Br. J. Educ. Technol.* **2021**, *52*, 680–694. [CrossRef]
9. Bala, L.; Kinross, J.; Martin, G.; Koizia, L.J.; Kooner, A.S.; Shimshon, G.J.; Hurkxkens, T.J.; Pratt, P.J.; Sam, A.H.; Bala, L.; et al. A remote access mixed reality teaching ward round. *Clin. Teach.* **2021**, *18*, 386–390. [CrossRef]

10. Fukuta, J.; Gill, N.; Rooney, R.; Coombs, A.; Murphy, D.; Fukuta, J.; Gill, N.; Rooney, R.; Coombs, A.; Murphy, D. Use of 360 degrees Video for a Virtual Operating Theatre Orientation for Medical Students. *J. Surg. Educ.* **2021**, *78*, 391–393. [[CrossRef](#)] [[PubMed](#)]
11. Sommer, G.M.; Broschewitz, J.; Huppert, S.; Sommer, C.G.; Jahn, N.; Jansen-Winkel, B.; Gockel, I.; Hau, H.M.; Sommer, G.M.; Broschewitz, J.; et al. The role of virtual reality simulation in surgical training in the light of COVID-19 pandemic Visual spatial ability as a predictor for improved surgical performance: A randomized trial. *Medicine* **2021**, *100*, e27844. [[CrossRef](#)]
12. Chan, V.; Larson, N.D.; Moody, D.A.; Moyer, D.G.; Shah, N.L.; Chan, V.; Larson, N.D.; Moody, D.A.; Moyer, D.G.; Shah, N.L. Impact of 360 degrees vs 2D Videos on Engagement in Anatomy Education. *Cureus* **2021**, *13*, e14260.
13. Tang, Y.M.; Chau, K.Y.; Kwok, A.P.K.; Zhu, T.; Ma, X. A systematic review of immersive technology applications for medical practice and education—Trends, application areas, recipients, teaching contents, evaluation methods, and performance. *Educ. Res. Rev.* **2022**, *35*, 100429. [[CrossRef](#)]
14. Chavez, O.L.; Rodriguez, L.F.; Gutierrez-Garcia, J.O.; Lopez Chavez, O.; Rodriguez, L.-F.; Octavio Gutierrez-Garcia, J. A comparative case study of 2D, 3D and immersive-virtual-reality applications for healthcare education. *Int. J. Med. Inform.* **2020**, *141*, 104226. [[CrossRef](#)]
15. Munn, Z.; Peters, M.D.J.; Stern, C.; Tufanaru, C.; McArthur, A.; Aromataris, E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med. Res. Methodol.* **2018**, *18*, 143. [[CrossRef](#)]
16. Slater, M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. R. Soc. B Biol. Sci.* **2009**, *364*, 3549–3557. [[CrossRef](#)] [[PubMed](#)]
17. Ochs, M.; Mestre, D.; de Montcheuil, G.; Pergandi, J.M.; Saubesty, J.; Lombardo, E.; Francon, D.; Blache, P.; Ochs, M.; Mestre, D.; et al. Training doctors' social skills to break bad news: Evaluation of the impact of virtual environment displays on the sense of presence. *J. Multimodal User Interfaces* **2019**, *13*, 41–51. [[CrossRef](#)]
18. Sutherland, I. The ultimate display. In Proceedings of the Congress of the International Federation of Information Processing (IFIP), New York, NY, USA, 24–29 May 1965; Volume 2, pp. 506–508.
19. Forrest, K.; McKimm, J. *Healthcare Simulation at A Glance*; John Wiley & Sons: Hoboken, NJ, USA, 2019.
20. Milgram, P.; Kishino, F. A taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.* **1994**, *77*, 1321–1329.
21. Lee, H.-G.; Chung, S.; Lee, W.-H. Presence in virtual golf simulators: The effects of presence on perceived enjoyment, perceived value, and behavioral intention. *New Media Soc.* **2013**, *15*, 930–946. [[CrossRef](#)]
22. Zackoff, M.W.; Real, F.J.; Cruse, B.; Davis, D.; Klein, M.; Zackoff, M.W.; Real, F.J.; Cruse, B.; Davis, D.; Klein, M. Medical Student Perspectives on the Use of Immersive Virtual Reality for Clinical Assessment Training. *Acad. Pediatr.* **2019**, *19*, 849–851. [[CrossRef](#)]
23. Kim, S.K.; Lee, Y.; Yoon, H.; Choi, J.; Kim, S.K.; Lee, Y.; Yoon, H.; Choi, J. Adaptation of Extended Reality Smart Glasses for Core Nursing Skill Training Among Undergraduate Nursing Students: Usability and Feasibility Study. *J. Med. Internet Res.* **2021**, *23*, e24313. [[CrossRef](#)]
24. Lampropoulos, G.; Barkoukis, V.; Burden, K.; Anastasiadis, T. 360-degree video in education: An overview and a comparative social media data analysis of the last decade. *Smart Learn. Environ.* **2021**, *8*, 20. [[CrossRef](#)]
25. Chuan, A.; Zhou, J.J.; Hou, R.M.; Stevens, C.J.; Bogdanovych, A. Virtual reality for acute and chronic pain management in adult patients: A narrative review. *Anaesthesia* **2021**, *76*, 695–704. [[CrossRef](#)]
26. Hoffman, H.G.; Patterson, D.R.; Carrougher, G.J.; Sharar, S.R. Effectiveness of Virtual Reality—Based Pain Control With Multiple Treatments. *Clin. J. Pain* **2001**, *17*, 229–235. [[CrossRef](#)]
27. Fong, K.N.K.; Tang, Y.M.; Sie, K.; Yu, A.K.H.; Lo, C.C.W.; Ma, Y.W.T. Task-specific virtual reality training on hemiparetic upper extremity in patients with stroke. *Virtual Real.* **2022**, *26*, 453–464. [[CrossRef](#)]
28. Cortés-Pérez, I.; Nieto-Escamez, F.A.; Obrero-Gaitán, E. Immersive Virtual Reality in Stroke Patients as a New Approach for Reducing Postural Disabilities and Falls Risk: A Case Series. *Brain Sci.* **2020**, *10*, 296. [[CrossRef](#)] [[PubMed](#)]
29. Scapin, S.; Echevarría-Guanilo, M.E.; Boeira Fuculo Junior, P.R.; Gonçalves, N.; Rocha, P.K.; Coimbra, R. Virtual Reality in the treatment of burn patients: A systematic review. *Burns* **2018**, *44*, 1403–1416. [[CrossRef](#)] [[PubMed](#)]
30. Reilly, C.A.; Greeley, A.B.; Jevsevar, D.S.; Gitajn, I.L. Virtual reality-based physical therapy for patients with lower extremity injuries: Feasibility and acceptability. *OTA Int.* **2021**, *4*, e132. [[CrossRef](#)] [[PubMed](#)]
31. Freitas, J.R.S.; Velosa, V.H.S.; Abreu, L.T.N.; Jardim, R.L.; Santos, J.A.V.; Peres, B.; Campos, P.A.-O. Virtual Reality Exposure Treatment in Phobias: A Systematic Review. *Psychiatr. Q.* **2021**, *92*, 1685–1710. [[CrossRef](#)]
32. Matamala-Gomez, M.; Bottiroli, S.; Realdon, O.; Riva, G.; Galvagni, L.; Platz, T.; Sandrini, G.; De Icco, R.; Tassorelli, C. Telemedicine and Virtual Reality at Time of COVID-19 Pandemic: An Overview for Future Perspectives in Neurorehabilitation. *Front. Neurol.* **2021**, *12*, 646902. [[CrossRef](#)]
33. Park, S.; Bokijonov, S.; Choi, Y. Review of microsoft hololens applications over the past five years. *Appl. Sci.* **2021**, *11*, 7259. [[CrossRef](#)]
34. Kourouthanassis, P.; Boletsis, C.; Bardaki, C.; Chasanidou, D. Tourists responses to mobile augmented reality travel guides: The role of emotions on adoption behavior. *Pervasive Mob. Comput.* **2015**, *18*, 71–87. [[CrossRef](#)]
35. Kolb, D. *Experiential Learning: Experience as the Source of Learning and Development*; Prentice-Hall: Hoboken, NJ, USA, 1984.
36. Fauconnier, G. Conceptual blending and analogy. In *The Analogical Mind: Perspectives from Cognitive Science*; Bradford Books: Cambridge, UK, 2001; pp. 255–286.

37. Sweller, J. Cognitive load theory. In *Psychology of Learning and Motivation*; Elsevier: Amsterdam, The Netherlands, 2011; Volume 55, pp. 37–76.
38. Vermunt, J.D. The regulation of constructive learning processes. *Br. J. Educ. Psychol.* **1998**, *68*, 149–171. [[CrossRef](#)]
39. Dweck, C.S. Motivational processes affecting learning. *Am. Psychol.* **1986**, *41*, 1040–1048. [[CrossRef](#)]
40. Dengel, A.; Mägdefrau, J. *Presence Is the Key to Understanding Immersive Learning*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 185–198.
41. Dale, E. *Audio-Visual Methods in Teaching*; Dryden Press: Fort Worth, TX, USA, 1969.
42. Keller, J.M. Motivation, learning, and technology: Applying the ARCS-V motivation model. *Particip. Educ. Res.* **2016**, *3*, 1–15. [[CrossRef](#)]
43. Mayer, R.E. Incorporating motivation into multimedia learning. *Learn. Instr.* **2014**, *29*, 171–173. [[CrossRef](#)]
44. Kyaw, B.M.; Saxena, N.; Posadzki, P.; Vseteckova, J.; Nikolaou, C.K.; George, P.P.; Divakar, U.; Masiello, I.; Kononowicz, A.A.; Zary, N.; et al. Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. *J. Med. Internet Res.* **2019**, *21*, e12959. [[CrossRef](#)] [[PubMed](#)]
45. Bracq, M.-S.; Michinov, E.; Jannin, P. Virtual Reality Simulation in Nontechnical Skills Training for Healthcare Professionals: A Systematic Review. *Simul. Healthc.* **2019**, *14*, 188–194. [[CrossRef](#)]
46. Mäkinen, H.; Haavisto, E.; Havola, S.; Koivisto, J.-M. User experiences of virtual reality technologies for healthcare in learning: An integrative review. *Behav. Inf. Technol.* **2022**, *41*, 1–17. [[CrossRef](#)]
47. Gilmour, J.; Huntington, A.; Bogossian, F.; Leadbitter, B.; Turner, C. Medical education and informal teaching by nurses and midwives. *Int. J. Med. Educ.* **2014**, *5*, 171–175. [[CrossRef](#)]
48. Jacobs, C. Immersive Technology in Healthcare Education: A Scoping Review. Available online: <https://osf.io/tpjyw/> (accessed on 17 January 2022).
49. Tricco, A.C.; Lillie, E.; Zarin, W.; O'Brien, K.K.; Colquhoun, H.; Levac, D.; Moher, D.; Peters, M.D.J.; Horsley, T.; Weeks, L.; et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann. Intern. Med.* **2018**, *169*, 467–473. [[CrossRef](#)] [[PubMed](#)]
50. Maggio, L.A.; Costello, J.A.; Norton, C.; Driessen, E.W.; Artino, A.R., Jr. Knowledge syntheses in medical education: A bibliometric analysis. *Perspect. Med. Educ.* **2021**, *10*, 79–87. [[CrossRef](#)] [[PubMed](#)]
51. Pei, L.; Wu, H. Does online learning work better than offline learning in undergraduate medical education? A systematic review and meta-analysis. *Med. Educ. Online* **2019**, *24*, 1666538. [[CrossRef](#)]
52. Ouzzani, M.; Hammady, H.; Fedorowicz, Z.; Elmagarmid, A. Rayyan—A web and mobile app for systematic reviews. *Syst. Rev.* **2016**, *5*, 210. [[CrossRef](#)] [[PubMed](#)]
53. Reed, D.A.; Cook, D.A.; Beckman, T.J.; Levine, R.B.; Kern, D.E.; Wright, S.M. Association Between Funding and Quality of Published Medical Education Research. *JAMA* **2007**, *298*, 1002–1009. [[CrossRef](#)] [[PubMed](#)]
54. Reed, D.A.; Beckman, T.J.; Wright, S.M.; Levine, R.B.; Kern, D.E.; Cook, D.A. Predictive Validity Evidence for Medical Education Research Study Quality Instrument Scores: Quality of Submissions to JGIM's Medical Education Special Issue. *J. Gen. Intern. Med.* **2008**, *23*, 903–907. [[CrossRef](#)] [[PubMed](#)]
55. Debes, A.J.; Aggarwal, R.; Balasundaram, I.; Jacobsen, M.B.; Debes, A.J.; Aggarwal, R.; Balasundaram, I.; Jacobsen, M.B. A tale of two trainers: Virtual reality versus a video trainer for acquisition of basic laparoscopic skills. *Am. J. Surg.* **2010**, *199*, 840–845. [[CrossRef](#)] [[PubMed](#)]
56. Elessawy, M.; Wewer, A.; Guenther, V.; Heilmann, T.; Eckmann-Scholz, C.; Schem, C.; Maass, N.; Noe, K.G.; Mettler, L.; Alkatout, I.; et al. Validation of psychomotor tasks by Symbionix LAP Mentor simulator and identifying the target group. *Minim. Invasive Ther. Allied Technol.* **2017**, *26*, 262–268. [[CrossRef](#)] [[PubMed](#)]
57. Fiedler, M.J.; Chen, S.J.; Judkins, T.N.; Oleynikov, D.; Stergiou, N. Virtual Reality for Robotic Laparoscopic Surgical Training. In *Studies in Health Technology and Informatics, Proceedings of 15th Conference on Medicine Meets Virtual Reality, Long Beach, CA, USA, 6–9 February 2007*; IOS Press: Amsterdam, The Netherlands, 2007; Volume 125, pp. 127–129.
58. Bartlett, J.D.; Lawrence, J.E.; Yan, M.; Guevel, B.; Stewart, M.E.; Audenaert, E.; Khanduja, V.; Bartlett, J.D.; Lawrence, J.E.; Yan, M.; et al. The learning curves of a validated virtual reality hip arthroscopy simulator. *Arch. Orthop. Trauma Surg.* **2020**, *140*, 761–767. [[CrossRef](#)]
59. Gelinias-Phaneuf, N.; Choudhury, N.; Al-Habib, A.R.; Cabral, A.; Nadeau, E.; Mora, V.; Pazos, V.; Debergue, P.; DiRaddo, R.; Del Maestro, R.F.; et al. Assessing performance in brain tumor resection using a novel virtual reality simulator. *Int. J. Comput. Assist. Radiol. Surg.* **2014**, *9*, 1–9. [[CrossRef](#)]
60. Sandy, N.S.; da Cruz, J.A.S.; Passerotti, C.C.; Nguyen, H.; dos Reis, S.T.; Gouveia, E.M.; Duarte, R.J.; Bruschini, H.; Srougi, M.; Sandy, N.S.; et al. Can the learning of laparoscopic skills be quantified by the measurements of skill parameters performed in a virtual reality simulator? *Int. Braz. J. Urol.* **2013**, *39*, 371–376. [[CrossRef](#)] [[PubMed](#)]
61. Cychosz, C.C.; Tofte, J.N.; Johnson, A.; Gao, Y.B.; Phisitkul, P.; Cychosz, C.C.; Tofte, J.N.; Johnson, A.; Gao, Y.; Phisitkul, P. Fundamentals of Arthroscopic Surgery Training Program Improves Knee Arthroscopy Simulator Performance in Arthroscopic Trainees. *Arthrosc.-J. Arthrosc. Relat. Surg.* **2018**, *34*, 1543–1549. [[CrossRef](#)] [[PubMed](#)]
62. Gallagher, A.G.; Lederman, A.B.; McGlade, K.; Satava, R.M.; Smith, C.D.; Gallagher, A.G.; Lederman, A.B.; McGlade, K.; Satava, R.M.; Smith, C.D. Discriminative validity of the Minimally Invasive Surgical Trainer in Virtual Reality (MIST-VR) using criteria levels based on expert performance. *Surg. Endosc. Other Interv. Tech.* **2004**, *18*, 660–665. [[CrossRef](#)] [[PubMed](#)]

63. Banaszek, D.; You, D.; Chang, J.; Pickell, M.; Hesse, D.; Hopman, W.M.; Borschneck, D.; Bardana, D. Virtual Reality Compared with Bench-Top Simulation in the Acquisition of Arthroscopic Skill A Randomized Controlled Trial. *J. Bone Jt. Surg.-Am. Vol.* **2017**, *99*, e34. [[CrossRef](#)] [[PubMed](#)]
64. Chien, J.H.; Suh, I.H.; Park, S.H.; Mukherjee, M.; Oleynikov, D.; Siu, K.C. Enhancing Fundamental Robot-Assisted Surgical Proficiency by Using a Portable Virtual Simulator. *Surg. Innov.* **2013**, *20*, 198–203. [[CrossRef](#)] [[PubMed](#)]
65. Gandhi, M.J.; Anderton, M.J.; Funk, L.; Gandhi, M.J.; Anderton, M.J.; Funk, L. Arthroscopic Skills Acquisition Tools: An Online Simulator for Arthroscopy Training. *Arthrosc.-J. Arthrosc. Relat. Surg.* **2015**, *31*, 1671–1679. [[CrossRef](#)] [[PubMed](#)]
66. Gasco, J.; Patel, A.; Ortega-Barnett, J.; Branch, D.; Desai, S.; Kuo, Y.F.; Luciano, C.; Rizzi, S.; Kania, P.; Matuyauskas, M.; et al. Virtual reality spine surgery simulation: An empirical study of its usefulness. *Neurol. Res.* **2014**, *36*, 968–973. [[CrossRef](#)]
67. Hart, R.; Doherty, D.A.; Karthigasu, K.; Garry, R.; Hart, R.; Doherty, D.A.; Karthigasu, K.; Garry, R. The value of virtual reality-simulator training in the development of laparoscopic surgical skills. *J. Minim. Invasive Gynecol.* **2006**, *13*, 126–133. [[CrossRef](#)] [[PubMed](#)]
68. Hudak, S.J.; Landt, C.L.; Hernandez, J.; Soderdahl, D.W.; Hudak, S.J.; Landt, C.L.; Hernandez, J.; Soderdahl, D.W. External Validation of a Virtual Reality Transurethral Resection of the Prostate Simulator. *J. Urol.* **2010**, *184*, 2018–2022. [[CrossRef](#)] [[PubMed](#)]
69. Johnston, T.J.; Tang, B.; Alijani, A.; Tait, I.; Steele, R.J.; Ker, J.; Nabi, G. Laparoscopic Surgical Skills are Significantly Improved by the Use of a Portable Laparoscopic Simulator: Results of a Randomized Controlled Trial. *World J. Surg.* **2013**, *37*, 957–964. [[CrossRef](#)] [[PubMed](#)]
70. Kanumuri, P.; Ganai, S.; Wohaibi, E.M.; Bush, R.W.; Grow, D.R.; Seymour, N.E. Virtual reality and computer-enhanced training devices equally improve laparoscopic surgical skill in novices. *Jsls-J. Soc. Laparoendosc. Surg.* **2008**, *12*, 219–226.
71. Le, T.D.B.; Adatia, F.A.; Lam, W.-C. Virtual reality ophthalmic surgical simulation as a feasible training and assessment tool: Results of a multicentre study. *Can. J. Ophthalmol.-J. Can. D Ophthalmol.* **2011**, *46*, 56–60. [[CrossRef](#)] [[PubMed](#)]
72. LeBel, M.-E.; Haverstock, J.; Cristancho, S.; van Eimeren, L.; Buckingham, G. Observational Learning During Simulation-Based Training in Arthroscopy: Is It Useful to Novices? *J. Surg. Educ.* **2018**, *75*, 222–230. [[CrossRef](#)] [[PubMed](#)]
73. Matzke, J.; Ziegler, C.; Martin, K.; Crawford, S.; Sutton, E. Usefulness of virtual reality in assessment of medical student laparoscopic skill. *J. Surg. Res.* **2017**, *211*, 191–195. [[CrossRef](#)] [[PubMed](#)]
74. McClusky, D.A.; Ritter, E.M.; Lederman, A.B.; Gallagher, A.G.; Smith, C.D. Correlation between perceptual, visuo-spatial, and psychomotor aptitude to duration of training required to reach performance goals on the MIST-VR surgical simulator. *Am. Surg.* **2005**, *71*, 13–20. [[CrossRef](#)] [[PubMed](#)]
75. Moore, A.K.; Grow, D.R.; Bush, R.W.; Seymour, N.E. Novices Outperform Experienced Laparoscopists on Virtual Reality Laparoscopy Simulator. *JSLs-J. Soc. Laparoendosc. Surg.* **2008**, *12*, 358–362.
76. Mulla, M.; Sharma, D.; Moghul, M.; Kailani, O.; Dockery, J.; Ayis, S.; Grange, P. Learning Basic Laparoscopic Skills: A Randomized Controlled Study Comparing Box Trainer, Virtual Reality Simulator, and Mental Training. *J. Surg. Educ.* **2012**, *69*, 190–195. [[CrossRef](#)] [[PubMed](#)]
77. Neumann, E.; Mayer, J.; Russo, G.I.; Amend, B.; Rausch, S.; Deininger, S.; Harland, N.; da Costa, I.A.; Hennenlotter, J.; Stenzl, A.; et al. Transurethral Resection of Bladder Tumors: Next-generation Virtual Reality Training for Surgeons. *Eur. Urol. Focus* **2019**, *5*, 906–911. [[CrossRef](#)] [[PubMed](#)]
78. Newmark, J.; Dandolu, V.; Milner, R.; Grewal, H.; Harbison, S.; Hernandez, E. Correlating virtual reality and box trainer tasks in the assessment of laparoscopic surgical skills. *Am. J. Obstet. Gynecol.* **2007**, *197*, e1–e4. [[CrossRef](#)]
79. Nielsen, M.R.; Kristensen, E.Q.; Jensen, R.O.; Mollerup, A.M.; Pfeiffer, T.; Graumann, O. Clinical Ultrasound Education for Medical Students Virtual Reality Versus e-Learning, a Randomized Controlled Pilot Trial. *Ultrasound Q.* **2021**, *37*, 292–296. [[CrossRef](#)] [[PubMed](#)]
80. Nomura, T.; Mamada, Y.; Nakamura, Y.; Matsutani, T.; Hagiwara, N.; Fujita, I.; Mizuguchi, Y.; Fujikura, T.; Miyashita, M.; Uchida, E.; et al. Laparoscopic skill improvement after virtual reality simulator training in medical students as assessed by augmented reality simulator. *Asian J. Endosc. Surg.* **2015**, *8*, 408–412. [[CrossRef](#)] [[PubMed](#)]
81. Orland, M.D.; Patetta, M.J.; Wieser, M.; Kayupov, E.; Gonzalez, M.H. Does Virtual Reality Improve Procedural Completion and Accuracy in an Intramedullary Tibial Nail Procedure? A Randomized Control Trial. *Clin. Orthop. Relat. Res.* **2020**, *478*, 2170–2177. [[CrossRef](#)] [[PubMed](#)]
82. Paschold, M.; Huber, T.; Maedge, S.; Zeissig, S.R.; Lang, H.; Kneist, W. Laparoscopic assistance by operating room nurses: Results of a virtual-reality study. *Nurse Educ. Today* **2017**, *51*, 68–72. [[CrossRef](#)]
83. Paschold, M.; Huber, T.; Zeissig, S.R.; Kauff, D.W.; Lang, H.; Kneist, W. Man versus machine: Is verbal instructor feedback in virtual-reality laparoscopic simulation dispensable? Results of a prospective randomised study. *Biomed. Eng.-Biomed. Tech.* **2014**, *59*, S1213–S1215.
84. Paschold, M.; Huber, T.; Zeissig, S.R.; Lang, H.; Kneist, W. Tailored instructor feedback leads to more effective virtual-reality laparoscopic training. *Surg. Endosc. Other Interv. Tech.* **2014**, *28*, 967–973. [[CrossRef](#)] [[PubMed](#)]
85. Peng, C.; Ng, K.M.; Roszczynialski, K.N.; Warrington, S.J.; Schertzer, K. Rapid Cycle Deliberate Practice in Virtual Reality: Teaching Transvenous Pacemaker Insertion to Emergency Medicine Residents. *Cureus* **2021**, *13*, e18503. [[CrossRef](#)] [[PubMed](#)]
86. Pereferrer, F.S.; Gonzalez, M.H.; Garcia, A.M.; Vilanova, A.C.; Dejardin, D.D. Evaluation of surgical skills in medical students using a virtual simulator. *Cir. Esp.* **2013**, *91*, 177–183.

87. Perez-Gutierrez, B.; Uribe-Quevedo, A.; Vega-Medina, L.; Salgado, J.S.; Jaimes, N.; Perez, O. Immersive and Non-Immersive VR Percutaneous Coronary Intervention Simulation for Acute Myocardial Infarction. In Proceedings of the 2020 IEEE 8th International Conference On Serious Games and Applications for Health (Segah 20), Vancouver, BC, Canada, 12–14 August 2020.
88. Rivard, J.D.; Vergis, A.S.; Unger, B.J.; Hardy, K.M.; Andrew, C.G.; Gillman, L.M.; Park, J. Construct validity of individual and summary performance metrics associated with a computer-based laparoscopic simulator. *Surg. Endosc. Other Interv. Tech.* **2014**, *28*, 1921–1928. [[CrossRef](#)] [[PubMed](#)]
89. Roehr, M.; Wu, T.; Maykowski, P.; Munter, B.; Hoebee, S.; Daas, E.; Kang, P. The Feasibility of Virtual Reality and Student-Led Simulation Training as Methods of Lumbar Puncture Instruction. *Med. Sci. Educ.* **2021**, *31*, 117–124. [[CrossRef](#)]
90. Roitberg, B.Z.; Kania, P.; Luciano, C.; Dharmavaram, N.; Banerjee, P. Evaluation of Sensory and Motor Skills in Neurosurgery Applicants Using a Virtual Reality Neurosurgical Simulator: The Sensory-Motor Quotient. *J. Surg. Educ.* **2015**, *72*, 1165–1171. [[CrossRef](#)]
91. Rose, K.; Pedowitz, R. Fundamental Arthroscopic Skill Differentiation With Virtual Reality Simulation. *Arthrosc.-J. Arthrosc. Relat. Surg.* **2015**, *31*, 299–305. [[CrossRef](#)]
92. Rosenthal, R.; Gantert, W.A.; Scheidegger, D.; Oertli, D. Can skills assessment on a virtual reality trainer predict a surgical trainee's talent in laparoscopic surgery? *Surg. Endosc. Other Interv. Tech.* **2006**, *20*, 1286–1290. [[CrossRef](#)] [[PubMed](#)]
93. Sanchez-Peralta, L.F.; Sanchez-Margallo, F.M.; Moyano-Cuevas, J.L.; Pagador, J.B.; Enciso, S.; Gomez-Aguilera, E.J.; Uson-Gargallo, J. Learning curves of basic laparoscopic psychomotor skills in SINERGIA VR simulator. *Int. J. Comput. Assist. Radiol. Surg.* **2012**, *7*, 881–889. [[CrossRef](#)] [[PubMed](#)]
94. Savran, M.M.; Nielsen, A.B.; Poulsen, B.B.; Thorsen, P.B.; Konge, L. Using virtual-reality simulation to ensure basic competence in hysteroscopy. *Surg. Endosc. Other Interv. Tech.* **2019**, *33*, 2162–2168. [[CrossRef](#)]
95. Scerbo, M.W.; Bliss, J.P.; Schmidt, E.A.; Thompson, S.N. The efficacy of a medical virtual reality simulator for training phlebotomy. *Hum. Factors* **2006**, *48*, 72–84. [[CrossRef](#)]
96. Seixas-Mikelus, S.A.; Adal, A.; Kesavadas, T.; Baheti, A.; Srimathveeravalli, G.; Hussain, A.; Chandrasekhar, R.; Wilding, G.E.; Guru, K.A. Can Image-Based Virtual Reality Help Teach Anatomy? *J. Endourol.* **2010**, *24*, 629–634. [[CrossRef](#)]
97. Sheth, S.S.; Fader, A.N.; Tergas, A.I.; Kushnir, C.L.; Green, I.C. Virtual Reality Robotic Surgical Simulation: An Analysis of Gynecology Trainees. *J. Surg. Educ.* **2014**, *71*, 125–132. [[CrossRef](#)] [[PubMed](#)]
98. Shetty, S.; Panait, L.; Baranoski, J.; Dudrick, S.J.; Bell, R.L.; Roberts, K.E.; Duffy, A.J. Construct and face validity of a virtual reality-based camera navigation curriculum. *J. Surg. Res.* **2012**, *177*, 191–195. [[CrossRef](#)] [[PubMed](#)]
99. Snyder, C.W.; Vandromme, M.J.; Tyra, S.L.; Hawn, M.T. Proficiency-Based Laparoscopic and Endoscopic Training With Virtual Reality, Simulators: A Comparison of Proctored and Independent Approaches. *J. Surg. Educ.* **2009**, *66*, 201–207. [[CrossRef](#)]
100. Sugand, K.; Wescott, R.A.; Carrington, R.; Hart, A.; van Duren, B.H. Training and Transfer Effect of FluoroSim, an Augmented Reality Fluoroscopic Simulator for Dynamic Hip Screw Guidewire Insertion A Single-Blinded Randomized Controlled Trial. *J. Bone Jt. Surg.-Am. Vol.* **2019**, *101*, e88. [[CrossRef](#)]
101. Tanoue, K.; Ieiri, S.; Konishi, K.; Yasunaga, T.; Okazaki, K.; Yamaguchi, S.; Yoshida, D.; Kakeji, Y.; Hashizume, M. Effectiveness of endoscopic surgery training for medical students using a virtual reality simulator versus a box trainer: A randomized controlled trial. *Surg. Endosc. Other Interv. Tech.* **2008**, *22*, 985–990. [[CrossRef](#)]
102. Van Ginkel, M.P.H.; Schijven, M.P.; van Grevenstein, W.M.U.; Schreuder, H.W.R. Bimanual Fundamentals: Validation of a New Curriculum for Virtual Reality Training of Laparoscopic Skills. *Surg. Innov.* **2020**, *27*, 523–533. [[CrossRef](#)] [[PubMed](#)]
103. Vergmann, A.S.; Vestergaard, A.H.; Grauslund, J. Virtual vitreoretinal surgery: Validation of a training programme. *Acta Ophthalmol.* **2017**, *95*, 60–65. [[CrossRef](#)] [[PubMed](#)]
104. Connolly, M.; Seligman, J.; Kastenmeier, A.; Goldblatt, M.; Gould, J.C. Validation of a virtual reality-based robotic surgical skills curriculum. *Surg. Endosc. Other Interv. Tech.* **2014**, *28*, 1691–1694. [[CrossRef](#)]
105. Fiedler, M.J.; Chen, S.-J.; Judkins, T.N.; Oleynikov, D.; Stergiou, N. Virtual Reality for Robotic Laparoscopic Surgical Training. In *Medicine Meets Virtual Reality 15: In Vivo, In Vitro, In Silico: Designing The Next In Medicine*; IOS Press: Amsterdam, The Netherlands, 2007; Volume 125, pp. 127–129.
106. Van Gestel, F.; Frantz, T.; Vannerom, C.; Verhellen, A.; Gallagher, A.G.; Elprama, S.A.; Jacobs, A.; Buyl, R.; Bruneau, M.; Jansen, B.; et al. The effect of augmented reality on the accuracy and learning curve of external ventricular drain placement. *Neurosurg. Focus* **2021**, *51*, E8. [[CrossRef](#)]
107. Phe, V.; Cattarino, S.; Parra, J.; Bitker, M.O.; Ambrogi, V.; Vaessen, C.; Roupert, M. Outcomes of a virtual-reality simulator-training programme on basic surgical skills in robot-assisted laparoscopic surgery. *Int. J. Med. Robot. Comput. Assist. Surg.* **2017**, *13*, 183–202. [[CrossRef](#)]
108. Hardon, S.F.; Kooijmans, A.; Horeman, R.; van der Elst, M.; Bloemendaal, A.L.A.; Horeman, T. Validation of the portable virtual reality training system for robotic surgery (PoLaRS): A randomized controlled trial. *Surg. Endosc. Other Interv. Techniques.* **2022**, *36*, 5282–5292. [[CrossRef](#)] [[PubMed](#)]
109. Issleib, M.; Kromer, A.; Pinnschmidt, H.O.; Suss-Havemann, C.; Kubitz, J.C. Virtual reality as a teaching method for resuscitation training in undergraduate first year medical students: A randomized controlled trial. *Scand. J. Trauma Resusc. Emerg. Med.* **2021**, *29*, 27. [[CrossRef](#)]

110. Berg, H.; Steinsbekk, A. Is individual practice in an immersive and interactive virtual reality application non-inferior to practicing with traditional equipment in learning systematic clinical observation? A randomized controlled trial. *BMC Med. Educ.* **2020**, *20*, 123. [[CrossRef](#)] [[PubMed](#)]
111. Bishop, M.E.; Ode, G.E.; Hurwit, D.J.; Zmugg, S.; Rauck, R.C.; Nguyen, J.T.; Ranawat, A.S. The Arthroscopic Surgery Skill Evaluation Tool Global Rating Scale is a Valid and Reliable Adjunct Measure of Performance on a Virtual Reality Simulator for Hip Arthroscopy. *Arthrosc.-J. Arthrosc. Relat. Surg.* **2021**, *37*, 1856–1866. [[CrossRef](#)] [[PubMed](#)]
112. Chang, J.; Banaszek, D.C.; Gambrel, J.; Bardana, D. Global Rating Scales and Motion Analysis Are Valid Proficiency Metrics in Virtual and Benchtop Knee Arthroscopy Simulators. *Clin. Orthop. Relat. Res.* **2016**, *474*, 956–964. [[CrossRef](#)] [[PubMed](#)]
113. Mulligan, A.; Vaghela, K.R.; Jeyaseelan, L.; Lee, J.; Akhtar, K. Transferable Global Rating Scales in the Validation of the ArthroSim (TM) Virtual Reality Arthroscopy Simulator. *Surg. Technol. Int.-Int. Dev. Surg. Surg. Res.* **2020**, *37*, 306–311.
114. Wong, D.T.; Mehta, A.; Singh, K.P.; Leong, S.M.; Ooi, A.; Niazi, A.; You-Ten, E.; Okrainec, A.; Patel, R.; Singh, M.; et al. The effect of virtual reality bronchoscopy simulator training on performance of bronchoscopic-guided intubation in patients A randomised controlled trial. *Eur. J. Anaesthesiol.* **2019**, *36*, 227–233. [[CrossRef](#)] [[PubMed](#)]
115. Chidambaram, S.; Erridge, S.; Leff, D.; Purkayastha, S. A Randomized Controlled Trial of Skills Transfer: From Touch Surgery to Laparoscopic Cholecystectomy. *J. Surg. Res.* **2019**, *234*, 217–223. [[CrossRef](#)]
116. Hoogenes, J.; Wong, N.; Al-Harbi, B.; Kim, K.S.; Vij, S.; Bolognone, E.; Quantz, M.; Guo, Y.B.; Shayegan, B.; Matsumoto, E.D.; et al. A Randomized Comparison of 2 Robotic Virtual Reality Simulators and Evaluation of Trainees' Skills Transfer to a Simulated Robotic Urethrovesical Anastomosis Task. *Urology* **2018**, *111*, 110–115. [[CrossRef](#)] [[PubMed](#)]
117. Janse, J.A.; Goedegebuure, R.S.A.; Veersema, S.; Broekmans, F.J.M.; Schreuder, H.W.R. Hysteroscopic Sterilization Using a Virtual Reality Simulator: Assessment of Learning Curve. *J. Minim. Invasive Gynecol.* **2013**, *20*, 775–782. [[CrossRef](#)]
118. Kulcsar, Z.; O'Mahony, E.; Lovquist, E.; Aboulafia, A.; Sabova, D.; Ghori, K.; Iohom, G.; Shorten, G. Preliminary evaluation of a virtual reality-based simulator for learning spinal anesthesia. *J. Clin. Anesth.* **2013**, *25*, 98–105. [[CrossRef](#)] [[PubMed](#)]
119. Mok, T.-N.; Chen, J.Y.; Pan, J.H.; Ming, W.K.; He, Q.Y.; Sin, T.H.; Deng, J.L.; Li, J.R.; Zha, Z.Z. Use of a Virtual Reality Simulator for Tendon Repair Training: Randomized Controlled Trial. *JMIR Serious Games* **2021**, *9*, e27544. [[CrossRef](#)] [[PubMed](#)]
120. Brady, A.-M. Assessment of learning with multiple-choice questions. *Nurse Educ. Pract.* **2005**, *5*, 238–242. [[CrossRef](#)] [[PubMed](#)]
121. Gutierrez, F.; Pierce, J.; Vergara, V.; Coulter, R.; Saland, L.; Caudell, T.; Goldsmith, T.; Alverson, D. The impact of the degree of immersion upon learning performance in virtual reality simulations for medical education. *J. Investig. Med.* **2007**, *55*, S91. [[CrossRef](#)]
122. Ebert, J.; Tutschek, B. Virtual reality objects improve learning efficiency and retention of diagnostic ability in fetal ultrasound. *Ultrasound Obstet. Gynecol.* **2019**, *53*, 525–528. [[CrossRef](#)] [[PubMed](#)]
123. Lesch, H.; Johnson, E.; Peters, J.; Cendan, J.C. VR Simulation Leads to Enhanced Procedural Confidence for Surgical Trainees. *J. Surg. Educ.* **2020**, *77*, 213–218. [[CrossRef](#)] [[PubMed](#)]
124. Galvez, R.; Wallon, R.C.; Shackelford, L.; Amos, J.R.; Rowen, J.L. Use of Virtual Reality to Educate Undergraduate Medical Students on Cardiac Peripheral and Collateral Circulation. *Med. Sci. Educ.* **2021**, *31*, 19–22. [[CrossRef](#)] [[PubMed](#)]
125. Maresky, H.S.; Oikonomou, A.; Ali, I.; Ditkofsky, N.; Pakkal, M.; Ballyk, B. Virtual Reality and Cardiac Anatomy: Exploring Immersive Three-Dimensional Cardiac Imaging, A Pilot Study in Undergraduate Medical Anatomy Education. *Clin. Anat.* **2019**, *32*, 238–243. [[CrossRef](#)]
126. Zackoff, M.W.; Real, F.J.; Sahay, R.D.; Fei, L.; Guiot, A.; Lehmann, C.; Tegtmeier, K.; Klein, M. Impact of an Immersive Virtual Reality Curriculum on Medical Students' Clinical Assessment of Infants With Respiratory Distress*. *Pediatr. Crit. Care Med.* **2020**, *21*, 477–485. [[CrossRef](#)] [[PubMed](#)]
127. Gopal, M.; Skobodzinski, A.A.; Sterbling, H.M.; Rao, S.R.; LaChapelle, C.; Suzuki, K.; Litle, V.R. Bronchoscopy Simulation Training as a Tool in Medical School Education. *Ann. Thorac. Surg.* **2018**, *106*, 280–286. [[CrossRef](#)] [[PubMed](#)]
128. O'Connor, M.; Stowe, J.; Potocnik, J.; Giannotti, N.; Murphy, S.; Rainford, L. 3D virtual reality simulation in radiography education: The students' experience. *Radiography* **2021**, *27*, 208–214. [[CrossRef](#)] [[PubMed](#)]
129. Atli, K.; Selman, W.; Ray, A. A Comprehensive Multicomponent Neurosurgical Course with use of Virtual Reality: Modernizing the Medical Classroom. *J. Surg. Educ.* **2021**, *78*, 1350–1356. [[CrossRef](#)] [[PubMed](#)]
130. Zackoff, M.W.; Young, D.; Sahay, R.D.; Fei, L.; Real, F.J.; Guiot, A.; Lehmann, C.; Klein, M. Establishing Objective Measures of Clinical Competence in Undergraduate Medical Education Through Immersive Virtual Reality. *Acad. Pediatr.* **2020**, *21*, 575–579. [[CrossRef](#)] [[PubMed](#)]
131. Wilson, G.F.; Russell, C.A. Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Hum. Factors* **2003**, *45*, 635–644. [[CrossRef](#)] [[PubMed](#)]
132. Yu, P.; Pan, J.J.; Wang, Z.X.; Shen, Y.; Wang, L.L.; Li, J.L.; Hao, A.M.; Wang, H.P. Cognitive Load/flow and Performance in Virtual Reality Simulation Training of Laparoscopic Surgery. In Proceedings of the 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW 2021), Lisbon, Portugal, 27 March–1 April 2021; pp. 468–469.
133. Mariana, F.P.; Damian, E.T.; Antonio, N.F.P.; Nuria, P.A.; Joaquin, C.M.J.; Rafael, M.R.; Joaquin, G.C.A.; Manuel, P.R.; Laura, J.R. Comparative study of a simulated incident with multiple victims and immersive virtual reality. *Nurse Educ. Today* **2018**, *71*, 48–53. [[CrossRef](#)]

134. Mills, B.; Dyks, P.; Hansen, S.; Miles, A.; Rankin, T.; Hopper, L.; Brook, L.; Bartlett, D. Virtual Reality Triage Training Can Provide Comparable Simulation Efficacy for Paramedicine Students Compared to Live Simulation-Based Scenarios. *Prehospital Emerg. Care* **2020**, *24*, 525–536. [[CrossRef](#)]
135. Kotranza, A.; Lind, D.S.; Lok, B. Real-Time Evaluation and Visualization of Learner Performance in a Mixed-Reality Environment for Clinical Breast Examination. *IEEE Trans. Vis. Comput. Graph.* **2012**, *18*, 1101–1114. [[CrossRef](#)]
136. Gierwialo, R.; Witkowski, M.; Kosieradzki, M.; Lisik, W.; Groszkowski, L.; Sitnik, R. Medical Augmented-Reality Visualizer for Surgical Training and Education in Medicine. *Appl. Sci.-Basel* **2019**, *9*, 2732. [[CrossRef](#)]
137. Peden, R.G.; Mercer, R.; Tatham, A.J. The use of head-mounted display eyeglasses for teaching surgical skills: A prospective randomised study. *Int. J. Surg.* **2016**, *34*, 169–173. [[CrossRef](#)] [[PubMed](#)]
138. Choueib, S.; McGarry, C.; Jaeger, M.; Ungi, T.; Janssen, N.; Fichtinger, G.; Patterson, L. Assessment of skill translation of intrathecal needle insertion using real-time needle tracking with an augmented reality display. In Proceedings of the Medical Imaging Conference—Image-Guided Procedures, Robotic Interventions, and Modeling, Houston, TX, USA, 16–19 February 2020.
139. Weeks, J.K.; Pakpoor, J.; Park, B.J.; Robinson, N.J.; Rubinstein, N.A.; Prouty, S.M.; Nachiappan, A.C. Harnessing Augmented Reality and CT to Teach First-Year Medical Students Head and Neck Anatomy. *Acad. Radiol.* **2021**, *28*, 871–876. [[CrossRef](#)] [[PubMed](#)]
140. Pickering, J.D.; Panagiotis, A.; Ntakakis, G.; Athanassiou, A.; Babatsikos, E.; Bamidis, P.D. Assessing the difference in learning gain between a mixed reality application and drawing screencasts in neuroanatomy. *Anat. Sci. Educ.* **2022**, *15*, 628–635. [[CrossRef](#)] [[PubMed](#)]
141. Robinson, B.L.; Mitchell, T.R.; Brenseke, B.M. Evaluating the Use of Mixed Reality to Teach Gross and Microscopic Respiratory Anatomy. *Med. Sci. Educ.* **2020**, *30*, 1745–1748. [[CrossRef](#)] [[PubMed](#)]
142. Bangor, A.; Kortum, P.T.; Miller, J.T. An Empirical Evaluation of the System Usability Scale. *Int. J. Hum.-Comput. Interact.* **2008**, *24*, 574–594. [[CrossRef](#)]
143. Vlachogianni, P.; Tselios, N. Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review. *J. Res. Technol. Educ.* **2021**, *54*, 1–18. [[CrossRef](#)]
144. Schoeb, D.S.; Schwarz, J.; Hein, S.; Schlager, D.; Pohlmann, P.F.; Frankenschmidt, A.; Gratzke, C.; Miernik, A. Mixed reality for teaching catheter placement to medical students: A randomized single-blinded, prospective trial. *BMC Med. Educ.* **2020**, *20*, 510. [[CrossRef](#)]
145. Anderson, M.; Guido-Sanz, F.; Diaz, D.A.; Lok, B.; Stuart, J.; Akinola, I.; Welch, G. Augmented Reality in Nurse Practitioner Education: Using a Triage Scenario to Pilot Technology Usability and Effectiveness. *Clin. Simul. Nurs.* **2021**, *54*, 105–112. [[CrossRef](#)]
146. Wang, C.; Daniel, B.K.; Asil, M.; Khwaounjoo, P.; Cakmak, Y.O. A Randomised Control Trial and Comparative Analysis of Multi-Dimensional Learning Tools in Anatomy. *Sci. Rep.* **2020**, *10*, 6120. [[CrossRef](#)] [[PubMed](#)]
147. Yoganathan, S.; Finch, D.A.; Parkin, E.; Pollard, J. 360° virtual reality video for the acquisition of knot tying skills: A randomised controlled trial. *Int. J. Surg.* **2018**, *54*, 24–27. [[CrossRef](#)] [[PubMed](#)]
148. Lowe, J.; Peng, C.; Winstead-Derlega, C.; Curtis, H. 360 virtual reality pediatric mass casualty incident: A cross sectional observational study of triage and out-of-hospital intervention accuracy at a national conference. *J. Am. Coll. Emerg. Physicians Open* **2020**, *1*, 974–980. [[CrossRef](#)] [[PubMed](#)]
149. Sultan, L.; Abuznadah, W.; Al-Jifree, H.; Khan, M.A.; Alsaywid, B.; Ashour, F. An Experimental Study On Usefulness Of Virtual Reality 360 degrees In Undergraduate Medical Education. *Adv. Med. Educ. Pract.* **2019**, *10*, 907–916. [[CrossRef](#)] [[PubMed](#)]
150. McLaughlin, N.; Rogers, J.; D’Arcy, J.; Gormley, G. ‘Sorry doctor . . . I didn’t hear that . . .’: Phenomenological analysis of medical students’ experiences of simulated hearing impairment through virtual reality. *BMJ Simul. Technol. Enhanc. Learn.* **2021**, *7*, 207–215. [[CrossRef](#)] [[PubMed](#)]
151. Giordano, N.A.; Whitney, C.E.; Axson, S.A.; Cassidy, K.; Rosado, E.; Hoyt-Brennan, A.M. A pilot study to compare virtual reality to hybrid simulation for opioid-related overdose and naloxone training. *Nurse Educ. Today* **2020**, *88*, 104365. [[CrossRef](#)]
152. Arents, V.; de Groot, P.C.M.; Struben, V.M.D.; van Stralen, K.J. Use of 360 degrees virtual reality video in medical obstetrical education: A quasi-experimental design. *BMC Med. Educ.* **2021**, *21*, 202. [[CrossRef](#)]
153. Sankaran, N.K.; Nisar, H.J.; Zhang, J.; Formella, K.; Amos, J.; Barker, L.T.; Vozenilek, J.A.; LaValle, S.M.; Kesavadas, T. Efficacy Study on Interactive Mixed Reality (IMR) Software with Sepsis Prevention Medical Education. In Proceedings of the 26th IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23–27 March 2019; pp. 664–670.
154. Birbara, N.S.; Pather, N. Instructional Design of Virtual Learning Resources for Anatomy Education. In *Biomedical Visualisation*; Springer: Berlin/Heidelberg, Germany, 2021; Volume 9, pp. 75–110.
155. Aebersold, M.; Tschannen, D.; Bathish, A. Innovative Simulation Strategies in Education. *Nurs. Res. Pract.* **2012**, *2012*, 765212. [[CrossRef](#)]
156. Andersen, S.A.W.; Mikkelsen, P.T.; Konge, L.; Caye-Thomasen, P.; Sorensen, M.S. Cognitive load in distributed and massed practice in virtual reality mastoidectomy simulation. *Laryngoscope* **2016**, *126*, E74–E79. [[CrossRef](#)]
157. Chiu, H.-Y.; Kang, Y.-N.; Wang, W.-L.; Chen, C.-C.; Hsu, W.; Tseng, M.-F.; Wei, P.-L. The Role of Active Engagement of Peer Observation in the Acquisition of Surgical Skills in Virtual Reality Tasks for Novices. *J. Surg. Educ.* **2019**, *76*, 1655–1662. [[CrossRef](#)] [[PubMed](#)]

158. Cone, D.C.; Serra, J.; Kurland, L. Comparison of the SALT and Smart triage systems using a virtual reality simulator with paramedic students. *Eur. J. Emerg. Med.* **2011**, *18*, 314–321. [[CrossRef](#)]
159. De Faria, J.W.V.; Teixeira, M.J.; Sousa Junior, L.d.M.; Otoch, J.P.; Figueiredo, E.G. Virtual and stereoscopic anatomy: When virtual reality meets medical education. *J. Neurosurg.* **2016**, *125*, 1105–1111. [[CrossRef](#)] [[PubMed](#)]
160. Du, Y.-C.; Fan, S.-C.; Yang, L.-C. The impact of multi-person virtual reality competitive learning on anatomy education: A randomized controlled study. *BMC Med. Educ.* **2020**, *20*, 343. [[CrossRef](#)]
161. Gockel, I.; Hakman, P.; Beardi, J.; Schutz, M.; Heinrichs, W.; Messow, C.M.; Junginger, T. New perspectives in laparoscopic simulation: From Students' Skills Lab to stress evaluation. *Zent. Fur Chir.* **2008**, *133*, 244–249. [[CrossRef](#)] [[PubMed](#)]
162. Henssen, D.J.H.A.; van den Heuvel, L.; De Jong, G.; Vorstenbosch, M.; Van Walsum, A.M.V.; Van den Hurk, M.M.; Kooloos, J.G.M.; Bartels, R. Neuroanatomy Learning: Augmented Reality vs. Cross-Sections. *Anat. Sci. Educ.* **2020**, *13*, 350–362. [[CrossRef](#)] [[PubMed](#)]
163. Hu, K.-C.; Salcedo, D.; Kang, Y.-N.; Lin, C.-W.; Hsu, C.-W.; Cheng, C.-Y.; Suk, F.-M.; Huang, W.-C. Impact of virtual reality anatomy training on ultrasound competency development: A randomized controlled trial. *PLoS ONE* **2020**, *15*, e0242731. [[CrossRef](#)] [[PubMed](#)]
164. Ijgosse, W.M.; van Goor, H.; Luursema, J.-M. Saving robots improves laparoscopic performance: Transfer of skills from a serious game to a virtual reality simulator. *Surg. Endosc. Other Interv. Tech.* **2018**, *32*, 3192–3199. [[CrossRef](#)]
165. Johannesson, E.; Silen, C.; Kvist, J.; Hult, H. Students' experiences of learning manual clinical skills through simulation. *Adv. Health Sci. Educ.* **2013**, *18*, 99–114. [[CrossRef](#)]
166. Jungmann, E.; Gockel, I.; Hecht, H.; Kuhr, K.; Rasanen, J.; Sihvo, E.; Lang, H. Impact of Perceptual Ability and Mental Imagery Training on Simulated Laparoscopic Knot-Tying in Surgical Novices Using a Nissen Fundoplication Model. *Scand. J. Surg.* **2011**, *100*, 78–85. [[CrossRef](#)]
167. Keskitalo, T. Students' expectations of the learning process in virtual reality and simulation-based learning environments. *Australas. J. Educ. Technol.* **2012**, *28*, 841–856. [[CrossRef](#)]
168. Kowalewski, K.-F.; Hendrie, J.D.; Schmidt, M.W.; Proctor, T.; Paul, S.; Garrow, C.R.; Kenngott, H.G.; Muller-Stich, B.P.; Nickel, F. Validation of the mobile serious game application Touch Surgery (TM) for cognitive training and assessment of laparoscopic cholecystectomy. *Surg. Endosc. Other Interv. Tech.* **2017**, *31*, 4058–4066. [[CrossRef](#)] [[PubMed](#)]
169. Kucuk, S.; Kapakin, S.; Goktas, Y. Learning Anatomy via Mobile Augmented Reality: Effects on Achievement and Cognitive Load. *Anat. Sci. Educ.* **2016**, *9*, 411–421. [[CrossRef](#)]
170. Liaw, S.Y.; Ooi, S.W.; Bin Rusli, K.D.; Lau, T.C.; San Tam, W.W.; Chua, W.L. Nurse-Physician Communication Team Training in Virtual Reality Versus Live Simulations: Randomized Controlled Trial on Team Communication and Teamwork Attitudes. *J. Med. Internet Res.* **2020**, *22*, e17279. [[CrossRef](#)] [[PubMed](#)]
171. Macnamara, A.F.; Bird, K.; Rigby, A.; Sathyapalan, T.; Hepburn, D. High-fidelity simulation and virtual reality: An evaluation of medical students' experiences. *BMJ Simul. Technol. Enhanc. Learn.* **2021**, *7*, 528–535. [[CrossRef](#)]
172. Makransky, G.; Petersen, G.B. Investigating the process of learning with desktop virtual reality: A structural equation modeling approach. *Comput. Educ.* **2019**, *134*, 15–30. [[CrossRef](#)]
173. Noll, C.; von Jan, U.; Raap, U.; Albrecht, U.-V. Mobile Augmented Reality as a Feature for Self-Oriented, Blended Learning in Medicine: Randomized Controlled Trial. *JMIR Mhealth Uhealth* **2017**, *5*, e139. [[CrossRef](#)]
174. Rahm, S.; Wieser, K.; Wicki, I.; Holenstein, L.; Fucentese, S.F.; Gerber, C. Performance of medical students on a virtual reality simulator for knee arthroscopy: An analysis of learning curves and predictors of performance. *BMC Surg.* **2016**, *16*, 14. [[CrossRef](#)]
175. Rim, D.; Shin, H. Development and Assessment of a Multi-User Virtual Environment Nursing Simulation Program: A Mixed Methods Research Study. *Clin. Simul. Nurs.* **2022**, *62*, 31–41. [[CrossRef](#)]
176. Rosas-Paredes, K.; Esquicha-Tejada, J.; Manrique Morante, H.; del Carpio Toia, A.M. Usability and Learning Environment of a Virtual Reality Simulator for Laparoscopic Surgery Training. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*, 760–768. [[CrossRef](#)]
177. Samosorn, A.B.; Gilbert, G.E.; Bauman, E.B.; Khine, J.; McGonigle, D. Teaching Airway Insertion Skills to Nursing Faculty and Students Using Virtual Reality: A Pilot Study. *Clin. Simul. Nurs.* **2020**, *39*, 18–26. [[CrossRef](#)]
178. Stepan, K.; Zeiger, J.; Hanchuk, S.; Del Signore, A.; Shrivastava, R.; Govindaraj, S.; Illoreta, A. Immersive virtual reality as a teaching tool for neuroanatomy. *Int. Forum Allergy Rhinol.* **2017**, *7*, 1006–1013. [[CrossRef](#)]
179. Uemura, M.; Tomikawa, M.; Nagao, Y.; Yamashita, N.; Kumashiro, R.; Tsutsumi, N.; Ohuchida, K.; Ieiri, S.; Ohdaira, T.; Hashizume, M.; et al. Significance of metacognitive skills in laparoscopic surgery assessed by essential task simulation. *Minim. Invasive Ther. Allied Technol.* **2014**, *23*, 165–172. [[CrossRef](#)] [[PubMed](#)]
180. Vincent, D.S.; Sherstyuk, A.; Burgess, L.; Connolly, K.K. Teaching Mass Casualty Triage Skills Using Immersive Three-dimensional Virtual Reality. *Acad. Emerg. Med.* **2008**, *15*, 1160–1165. [[CrossRef](#)] [[PubMed](#)]
181. Windsor, J.A.; Diener, S.; Zoha, F. Learning style and laparoscopic experience in psychomotor skill performance using a virtual reality surgical simulator. *Am. J. Surg.* **2008**, *195*, 837–842. [[CrossRef](#)] [[PubMed](#)]
182. Wainman, B.; Pukas, G.; Wolak, L.; Mohanraj, S.; Lamb, J.; Norman, G.R. The Critical Role of Stereopsis in Virtual and Mixed Reality Learning Environments. *Anat. Sci. Educ.* **2020**, *13*, 398–405. [[CrossRef](#)] [[PubMed](#)]
183. Wainman, B.; Aggarwal, A.; Birk, S.K.; Gill, J.S.; Hass, K.S.; Fenesi, B. Virtual Dissection: An Interactive Anatomy Learning Tool. *Anat. Sci. Educ.* **2021**, *14*, 788–798. [[CrossRef](#)] [[PubMed](#)]

184. Bric, J.; Connolly, M.; Kastenmeier, A.; Goldblatt, M.; Gould, J.C. Proficiency training on a virtual reality robotic surgical skills curriculum. *Surg. Endosc. Other Interv. Tech.* **2014**, *28*, 3343–3348. [[CrossRef](#)] [[PubMed](#)]
185. Simpson, E.J. *The Classification of Educational Objectives, Psychomotor Domain*; University of Illinois: Champaign, IL, USA, 1966.
186. Cierniak, G.; Scheiter, K.; Gerjets, P. Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Comput. Hum. Behav.* **2009**, *25*, 315–324. [[CrossRef](#)]
187. Ryan, R.M.; Mims, V.; Koestner, R. Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *J. Personal. Soc. Psychol.* **1983**, *45*, 736–750. [[CrossRef](#)]
188. Witmer, B.G.; Singer, M.J. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators Virtual Environ.* **1998**, *7*, 225–240. [[CrossRef](#)]
189. Makransky, G.; Lilleholt, L.; Aaby, A. Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Comput. Hum. Behav.* **2017**, *72*, 276–285. [[CrossRef](#)]
190. Servotte, J.-C.; Goosse, M.; Campbell, S.H.; Dardenne, N.; Pilote, B.; Simoneau, I.L.; Guillaume, M.; Bragard, I.; Ghuysen, A. Virtual Reality Experience: Immersion, Sense of Presence, and Cybersickness. *Clin. Simul. Nurs.* **2020**, *38*, 35–43. [[CrossRef](#)]
191. Anders Ericsson, K. Deliberate Practice and Acquisition of Expert Performance: A General Overview. *Acad. Emerg. Med.* **2008**, *15*, 988–994. [[CrossRef](#)]
192. Petrosoniak, A.; Lu, M.; Gray, S.; Hicks, C.; Sherbino, J.; McGowan, M.; Monteiro, S. Perfecting practice: A protocol for assessing simulation-based mastery learning and deliberate practice versus self-guided practice for bougie-assisted cricothyroidotomy performance. *BMC Med. Educ.* **2019**, *19*, 100. [[CrossRef](#)] [[PubMed](#)]
193. Andersen, S.A.; Konge, L.; Caye-Thomasen, P.; Sorensen, M.S.; Andersen, S.A.; Konge, L.; Caye-Thomasen, P.; Sorensen, M.S. Learning Curves of Virtual Mastoidectomy in Distributed and Massed Practice. *JAMA Otolaryngol.-Head Neck Surg.* **2015**, *141*, 913–918. [[CrossRef](#)] [[PubMed](#)]
194. Cisse, C.; Angioi, K.; Luc, A.; Berrod, J.-P.; Conart, J.-B. EYESI surgical simulator: Validity evidence of the vitreoretinal modules. *ACTA Ophthalmol.* **2019**, *97*, E277–E282. [[CrossRef](#)] [[PubMed](#)]
195. Lia, H.; Paulin, G.; Yeo, C.T.; Andrews, J.; Yi, N.; Haq, H.; Emmanue, S.; Ludig, K.; Keri, Z.; Lasso, A.; et al. HoloLens in Suturing Training. In Proceedings of the Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling, Houston, TX, USA, 10–15 February 2018; p. 10576.
196. Berg, H.; Steinsbekk, A. The effect of self-practicing systematic clinical observations in a multiplayer, immersive, interactive virtual reality application versus physical equipment: A randomized controlled trial. *Adv. Health Sci. Educ.* **2021**, *26*, 667–682. [[CrossRef](#)]
197. Pears, M.; Konstantinidis, S. The Future of Immersive Technology in Global Surgery Education. *Indian J. Surg.* **2022**, *84*, 281–285. [[CrossRef](#)] [[PubMed](#)]
198. Ryan, G.V.; Callaghan, S.; Rafferty, A.; Higgins, M.F.; Mangina, E.; McAuliffe, F. Learning Outcomes of Immersive Technologies in Health Care Student Education: Systematic Review of the Literature. *J Med. Internet Res.* **2022**, *24*, e30082. [[CrossRef](#)] [[PubMed](#)]
199. Carney, P.A.; Nierenberg, D.W.; Pipas, C.F.; Brooks, W.B.; Stukel, T.A.; Keller, A.M. Educational Epidemiology Applying Population-Based Design and Analytic Approaches to Study Medical Education. *JAMA* **2004**, *292*, 1044–1050. [[CrossRef](#)] [[PubMed](#)]
200. Barrie, M.; Socha, J.J.; Mansour, L.; Patterson, E.S. Mixed Reality in Medical Education: A Narrative Literature Review. In Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care, Chicago, IL, USA, 24–27 March 2019; Volume 8, pp. 28–32. [[CrossRef](#)]
201. Beckman, T.J.; Cook, D.A.; Mandrekar, J.N. What is the validity evidence for assessments of clinical teaching? *J. Gen. Intern. Med.* **2005**, *20*, 1159–1164. [[CrossRef](#)] [[PubMed](#)]
202. Golnari, P.; Sodagari, F.; Baradaran, H.R. Quality of published Iranian medical education research studies: A systematic review. *Med. J. Islam. Repub. Iran* **2014**, *28*, 79. [[PubMed](#)]
203. All, A.; Castellar, E.P.N.; Van Looy, J. Assessing the effectiveness of digital game-based learning: Best practices. *Comput. Educ.* **2016**, *92*, 90–103. [[CrossRef](#)]
204. Gorbanev, I.; Agudelo-Londoño, S.; González, R.A.; Cortes, A.; Pomares, A.; Delgadillo, V.; Yepes, F.J.; Muñoz, Ó. A systematic review of serious games in medical education: Quality of evidence and pedagogical strategy. *Med. Educ. Online* **2018**, *23*, 1438718. [[CrossRef](#)]
205. McGaghie, W.C.; Issenberg, S.B.; Petrusa, E.R.; Scalese, R.J. Effect of practice on standardised learning outcomes in simulation-based medical education. *Med. Educ.* **2006**, *40*, 792–797. [[CrossRef](#)] [[PubMed](#)]
206. Kirkpatrick, D.L.; Kirkpatrick, J.D. *Implementing the Four Levels: A Practical Guide for Effective Evaluation of Training Programs*; ReadHowYouWant.com, Limited: Sydney, Australia, 2009.
207. Ritter, E.M.; McClusky, D.A.; Gallagher, A.G.; Enochsson, L.; Smith, C.D. Perceptual, visuospatial, and psychomotor abilities correlate with duration of training required on a virtual-reality flexible endoscopy simulator. *Am. J. Surg.* **2006**, *192*, 379–384. [[CrossRef](#)] [[PubMed](#)]
208. Taubert, M.; Webber, L.; Hamilton, T.; Carr, M.; Harvey, M. Virtual reality videos used in undergraduate palliative and oncology medical teaching: Results of a pilot study. *BMJ Support. Palliat. Care* **2019**, *9*, 281–285. [[CrossRef](#)] [[PubMed](#)]

-
209. Messick, S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am. Psychol.* **1995**, *50*, 741–749. [[CrossRef](#)]
 210. Cook, D.A.; Hatala, R. Validation of educational assessments: A primer for simulation and beyond. *Adv. Simul.* **2016**, *1*, 31. [[CrossRef](#)] [[PubMed](#)]