Title:

Augmented Reality in Medical Education: A Systematic Review

Authors

Kevin S. Tang[1,2,4,5], Derrick L. Cheng[1,2,3], Eric Mi[1], Paul B. Greenberg[4,5]

[1]The Program in Liberal Medical Education of Brown University, Providence RI
[2]The Warren Alpert Medical School of Brown University, Providence RI
[3]Lifespan Clinical Research Center, Providence RI
[4]Division of Ophthalmology, Warren Alpert Medical School, Providence RI
[5]Section of Ophthalmology, Providence VA Medical Center, Providence RI

Competing/conflicts of interest: No stated conflict of interest.

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the United States Department of Veterans Affairs or the United States government.

Presentation: Presented at the Association for Medical Education in Europe Annual Conference, August 28th, 2018, Congress Centre, Basel, Switzerland

Corresponding Author:

Paul B Greenberg, MD, MPH
Section of Ophthalmology
Providence VA Medical Center
830 Chalkstone Ave
Providence, RI 02908
Phone 401-273-7100 x1506
Email: paul_greenberg@brown.edu

The authors report no conflicts of interest

Abstract

Introduction

The field of augmented reality (AR) is rapidly growing with many new potential applications in medical education. This systematic review aims to investigate the current state of augmented reality applications (ARAs) as teaching tools in the healthcare field.

Methods

A literature search was conducted using PubMed, Embase, Web of Science, Cochrane Library, and Google Scholar. This review followed PRISMA guidelines and included publications from January 1, 2000 to June 18, 2018. Inclusion criteria were experimental studies evaluating ARAs implemented in healthcare education published in English. The quality of each study was assessed using GRADE criteria. The five stages of validity initially described by Gallagher et al. (2003) for assessment of surgical simulation were also applied to each ARA.

Results

We identified 100,807 articles in the initial literature search; 36 met inclusion criteria for final review and were categorized into three categories: Surgery (23), Anatomy (9), and Classroom (4). The overall quality of the studies was poor. No ARAs were validated at all five levels.

Conclusion

While AR technology is growing at a rapid rate, the current quality and breadth of AR research in medical training is insufficient to recommend the adoption into educational curricula. More coordinated and comprehensive research is needed to define the role of AR technology in medical education.

Commented [KT1]: E1

Introduction

Over the past decade, augmented and virtual reality technology have demonstrated the potential to transform a variety of fields. Virtual reality (VR) technology creates entirely artificial environments through headsets that isolate users from their surroundings. In comparison, augmented reality (AR) overlays digital interfaces upon physical surroundings, producing an environment that is both real and digital.[1,2] This combination of physical and virtual information allows AR to further enhance the well-established methods of procedural simulation.[3]

While the concept of AR has existed for several decades,[4,5] recent advances in visual technology and the development of new augmented reality applications (ARAs) have drawn consumer and professional attention.[6] ARAs are software and/or hardware developed explicitly with AR functionality in mind, and have already been applied in many educational settings including environmental sciences, chemistry, humanities, and the arts.[7,8] Recent studies have shown that there is a growing number of ARAs in medicine and that AR may foreshadow a new paradigm in medical education.[8,9] To date, ARAs have been adapted to every stage of medical training as anatomical teaching tools,[10] classroom study aids,[11] image training simulators,[12] and clinical skills interaction simulators.[13]

To date, there have been no systematic reviews that comprehensively describe the use of different ARAs in medical education. Prior systematic reviews have not assessed the quality of recent AR research in medical education and have focused primarily on the integration of surgical ARAs in medical training[9] or applications in general education.[8,14] The purpose of this systematic review is to (a) examine the quality of current research on AR in medical education, (b) describe prevalent ARAs in varying areas of medical education that have been studied by

Commented [KT2]: C1a

Commented [KT3]: C1b

multiple research teams, and (c) develop an analytical model to guide future research in addressing the current gaps in literature and advocating for that adoption of AR technology into current medical curricula.

Commented [KT4]: E2

Methods

We conducted a systematic literature search using PubMed, Cochrane Library, Embase, Web of Science, and Google Scholar from January 1, 2000 through June 18, 2018. Search phrases included the keyword "augmented reality" with the phrases "surgical training", "surgical education", "anatomy education", "medical education", and "medical student". Keyword and database selection were determined following independent consultations with a university librarian (E.S.) and a medical education researcher (P.G.). Search results were recorded per the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[15]

Commented [KT5]: E3

Included articles a) described ARAs in the context of medicine and medical education, b) carried out experimental studies evaluating specific ARAs, c) were obtained from peer-reviewed journals after the year 2000, and d) were written in English. Excluded results included articles that a) discussed VR or similar technologies but not AR, b) were focused on the technological basis for AR or c) discussed AR outside of medicine. Reporting bias was minimized using two independent reviewers (K.T., D.C.) with a third reviewer to break any conflicts (E.M.) in addition to consultation with a university librarian to optimize search parameters.

Commented [KT6]: E5, E6

Reviewed articles were divided into three categories. "Surgical" applications were designed to train medical novices in procedural tasks such as basic laparoscopic skills, suturing, ventriculostomy, and echocardiography. "Anatomy" applications were designed to assist

students with learning human anatomy. "Other" applications were developed for general healthcare education including clinical skills, forensic medicine, dermatology, and pathology.

Commented [KT7]: C3

Studies were assessed for quality using criteria based on the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) Working Group scoring protocol.[16] Per the GRADE criteria, quality analysis is determined subjectively based on qualitative metrics including inconsistency, directness of evidence, possibility of bias, confounders, strength of association, dose response, and data quantity. In addition to GRADE quality assessment, each ARA was evaluated for validity.[17-19] We assessed five forms of validity proposed by Gallagher et. al.: face, content, construct, concurrent, and predictive validity.[9,19] A description of these five stages of validity can be found in the Appendix. No specific quantitative variables were analyzed in this review.

Commented [KT8]: E4, E6

Commented [KT9]: E4

Results

We identified 100,807 papers in the initial search. Title screening and removal of duplicates left 439 papers that were evaluated based on abstract. Second-level exclusion removed 347 papers, leaving 93 full-text papers that were reviewed in their entirety. Thirty-six articles met proposed inclusion criteria. These papers were divided into three categories -- 23 in Surgical, nine in Anatomy, and four in Other. 22 total ARAs were described: 15 in Surgical, five in Anatomy, and two in Other. Of the 36 included articles, 26 (72%) were published in the last five years and eight (22%) were published between 5-10 years ago. A PRISMA flowchart detailing this literature search is shown in Figure 1.

Evaluation of study quality is delineated in Tables 1 and 2. Three out of 22 ARAs (14%) received a quality grade above Low and one ARA received a High rating. Eleven ARAs (50%)

did not achieve any stage of validity and no application has been sufficiently tested to achieve all five stages of validity. In the following sections, we describe in detail the ARAs that have been evaluated by multiple studies. Many of the identified ARAs have only been the subject of a single study and thus remain largely untested, these include Google Glass, Microsoft Hololens, and the virtual patient (VP) DIANA, among many others; a full list of identified ARAs and study outcomes is included in Tables 3 and 4.

Commented [KT10]: C2

Surgical

ProMIS AR Laparoscopic Simulator (Haptica, Dublin, Ireland)

Of 36 studies, seven involved use of the ProMIS simulator. Composed of a torso-shaped mannequin connected to a computer, this device trained students in laparoscopic procedures and combined the benefits of haptic feedback with the ability to view simulation feedback videos. Three cameras within the mannequin identify inserted instruments from different angles. Substitution of the peritoneal cavity with plastic trays allows the simulator to be used for multiple tasks.

The ProMIS AR simulator was used to train users on sigmoid colectomies[20], suturing[21-24], and other basic laparoscopic tasks.[24-26] Overall, the ProMIS trainer was an effective educational tool. It was described as highly realistic and improved task-effectiveness across all studies.[21] Studies that measured the difference in skill between novice and experienced participants found a significant correlation between high performance metrics and experience, indicating that the ProMIS simulator is reliable for evaluation of laparoscopic skills.[20,22,24,26] It is important to note that the majority of these studies were pilot studies with low numbers of participants (n=7-28) with the exception of one (n=115)[25]. Additionally, none of the studies were randomized, only

one was controlled[20], and most depended on subjective means such as Likert-scale surveys to determine performance.

ImmersiveTouch System (ImmersiveTouch, Inc., University of Illinois, Chicago, IL, USA)

Another AR training simulator that provides haptic feedback is the ImmersiveTouch system. ImmersiveTouch involves the integration of a head-hand tracking system with a stereoscopic display and is typically used for neurosurgical training.

Two randomized controlled trials (RCTs) evaluated the ImmersiveTouch system -- one for thoracic screw placement[27] and the other for ventriculostomies.[28] Use of the ARA slightly lowered failure rate in screw placement and demonstrated a statistically significant improvement of correct catheter placement for ventriculostomies. However, these experiments had small sample sizes of 51 and 16 participants, respectively.

EyeSI AR Binocular Indirect Ophthalmoscopy (BIO) Simulator (VRmagic Holding AG, Mannheim, Germany)

The EyeSI AR simulator displays virtual retinae on a model head through a lens inspired by traditional BIO lenses. The user physically adjusts the lens to look in different directions while their movements are recorded on a separate monitor.

Two RCTs compared traditional BIO lenses to the EyeSI AR simulator. Rai et al. (n=28) randomized first-year ophthalmology residents to traditional and EyeSI training methods and evaluated their performance in three tasks.[29] The AR group significantly outperformed the control group in both raw score and mean performance and was able to complete the procedure in less time. Leitritz et al. (n=37) randomized 4th year medical students with no prior experience

with BIO into control and AR groups using the EyeSI simulator.[30] All students performed the procedure the day after training and were assessed through their drawings of the patient's optic disk and arteries/veins. The AR group sketched more vessels correctly and achieved a higher Ophthalmoscopy Training Score.

Anatomy

AR Magic Book (various)

Several studies utilized a system called "MagicBook".[31,32] A number of specific ARAs fit into this category (see Table 4) but all consisted of a standard didactic textbook with cards for relevant anatomical figures. These cards could be recognized by a computer webcam or a smartphone and were able to display a virtual, interactive representation of the figure on the connected display.

Two large RCTs conducted by Ferrer-Torregrosa et al.[33,34] concluded that this type of ARA improved attention, recall, learning, structure, imaging, and understanding in university students. The AR group scored significantly higher than the traditional learning control groups on final assessments. Most respondents believed that AR was effective for studying (76.9%), that it increased motivation and interest (75%), and that their grades would improve if professors utilized the technology (67.3%). Another RCT conducted by Kucuk et al.[35] demonstrated similar results: medical students utilizing the "MagicBook" ARA scored significantly higher on an academic test with lower cognitive load compared to control and 100% of respondents reported that AR either greatly or partially facilitated learning.

Microsoft Kinect (Microsoft Corp., Redmond, WA, USA)

The Microsoft Kinect was often used as part of an "AR Magic Mirror" (ARMM) approach. The Kinect contains a high-resolution camera for video reproduction and a low-resolution camera for depth perception, allowing the device to accurately track the user's body movements. The system is often used for interactive video games but can be adapted to allow overlay of tracked virtual information onto a user's body.

There were three papers exploring the ARMM application; all were surveys directed at medical students and clinicians.[36-38] Responses from all three were positive. Varying majorities of respondents reported that ARMM increased learning motivation (58%), was beneficial in an educational setting (69.1%), stimulated active learning (82.4%), and improved 3-Dimensional understanding of anatomy (93.4%) while remaining easy to use.[36,37] A large majority (80.5%) rated the system as excellent or good, and surveyed physicians unanimously recommended that ARMM be used to supplement existing anatomy curriculums.[38]

Other

Mobile AR Blended Learning Environment (mARble) [Peter L. Reichertz Institute for Medical Informatics at the Hannover Medical School, Hanover, Germany]

The mARble is an application developed for the Apple mobile operating system that stores content separately from the program's code; this allows for the addition of modules to adapt the application for different purposes without changing its source code.

Three studies evaluated the mARble application; two were RCTs[39,40] and one was a survey.[41] All three had small sample sizes, with two recruiting ten or less participants.[39,41] Students described the application as pragmatic and enjoyable to use, but the two RCTs yielded conflicting results.

Albercht et al. concluded that mARble increased knowledge retention with lower cognitive fatigue when compared with traditional textbook material[39], but Noll et al. found no difference in knowledge gain between mARble and control groups immediately after training, although the AR group retained more knowledge in a follow-up assessment 14 days after training.[40]

Discussion

While AR technology has the potential to improve or replace traditional medical training methods, this systematic review demonstrates inconsistency in both focus and quality of published studies. Overall, most studies were of low or very low quality and failed to meet the five described validation criteria. Despite these shortcomings, the large majority of studies established positive responses toward AR and a desire by both trainees and experts to see the technology implemented in training programs. Most articles identified in this systematic review were published within the last five years, underscoring the rapidly expanding nature of the field. These findings suggest that the integration of AR in medical education is gaining momentum and consumer interest. It is important for future research to meet higher quality standards because only then may evidence supporting AR's utility in medical education facilitate the technology's implementation into current educational curricula.

Commented [KT11]: E8

This review primarily used grading criteria developed by the GRADE Working Group[16] and simulation validation criteria described by Gallagher et al.[19] to assess study and ARA quality. Per the GRADE criteria, the most important component of study quality is study design: randomized trials constitute the highest grade of evidence while uncontrolled or observational studies are characterized as low or very low grade. Points are subtracted for methodological limitations, inconsistencies in results between multiple studies (or lack of multiple studies), high probability

of reporting bias, and uncertainty about directness of evidence. Points are added for significant outcomes based on consistent evidence across two or more studies. These characteristics are combined to offer a final grade of recommendation for an intervention's estimate of effect. Application of these criteria to our findings provided uninspiring results: the majority of ARAs were graded low or very low quality, where only three papers provided high quality evidence supporting a single ARA.[33-35] Points were primarily lost for study design and inconsistency. Most reviewed studies were pilot studies with small sample sizes utilizing hardware that is not commercially viable. Of these, only 31% were RCTs and 55% had sample sizes <50. Studies also evaluated widely divergent, research-oriented ARAs that were not designed for consumer use; therefore, applications were unable to achieve consistent outcomes across multiple experiments. Only the ProMIS simulator, Microsoft Kinect ARMM test system, EyeSI, and mARble were evaluated by more than one study.

The five validity criteria proposed by Gallagher et. al. were initially adopted to evaluate testing instruments in surgical training.[19] More recently, these criteria were used to provide systematic objective validation for surgical simulators and to assess readiness for implementation in surgical curricula.[9,17,18] The surgical origins of modern validation techniques reflects the higher levels of validity achieved by surgical ARAs in this review. Several articles explicitly aimed to demonstrate specific stages of validity. However, we contend that these tests of validity should be modified and adapted to all uses of AR in medical education.

The state and quality of research also varied between surgical, anatomical, and classroom-based ARAs. Surgical ARAs included a variety of laparoscopic simulators (ProMIS, ImmersiveTouch), AR glasses (Google Glass©, Microsoft Hololens©, etc.), and AR telementoring systems (ART, STAR). This diversity reflects the well-documented use of simulation as a surgical training

tool.[9,42,43] The development of AR hardware by leading technology corporations such as Google, Microsoft, Brother, and Epson also indicates the potential integration of consumer products into medical settings. While recent technological advances have made AR simulation more viable for surgical training, further developments will need to broaden in scope to focus on more than technical skill.[9,44] A holistic approach to training effective surgeons will require the integration of knowledge and attitude education[45,46] as well as development of standardized assessments of simulation training in the operating room.[47,48]

Anatomical ARAs generally used a "MagicBook" or ARMM approach. AR technology is easily applied to anatomy learning due to its heavy reliance on spatial and 3-dimensional conceptualization – a hallmark of digital simulation. Consequently, the use of digital technology to enhance anatomical learning has already been studied for over a decade.[49] This extensive history is reflected by higher quality evidence: anatomical studies include several large RCTs, specifically for "MagicBook" experiences.[33-35] Three studies found that the use of this technology significantly improved student assessment scores post-training, indicating reproducible potential and high quality evidence by GRADE criteria.[16]

Studies in the Other category did not offer compelling evidence for AR implementation. There was a lack of consistently positive outcomes and high-quality studies for both mARble[39-41] and DIANA.[50] Study sample sizes were also small. Outcomes of mARble were conflicting: Albrecht et al. concluded that mARble was superior to traditional textbook learning[39] while Noll et al. demonstrated that mARble did not produce better knowledge retention than mobile phone applications.[40] The VP DIANA produced worse assessment and empathy scores than traditional SP experiences.[50] This may be a result of the unrealistic design of the system; adjustments to

enhance the realism of the VP DIANA module and incorporation of more modern AR simulation technology (including AR glasses) may improve student outcomes.

The breadth of projects identified in this review highlights both the adaptability of AR technology and the lack of standardized assessment tools. To this end, we propose an analytical model to assess future research and guide ARA integration into current educational paradigms. This model utilizes elements from Cook et al.'s approach to evaluating the implementation of technology-enhanced learning (TEL) in medical education[51] as well as the quality criteria and validity metrics described previously.[19] However, our model differs from these examples in that we seek to holistically address all barriers to AR technology's eventual implementation into medical educational curricula. We divide our model into four primary components illustrated in Figure 2: *quality, application content, outcome,* and *feasibility*.

Commented [KT12]: E9

Quality references the caliber of study design and consistency of evidence.[16] As recommended by GRADE criteria, future AR research should utilize more rigorous study designs and larger study sizes as well as conduct more studies on existing ARAs to provide further feedback and high-quality evidence supporting curricular integration. Importantly, subjective metrics such as "realism" proposed by GRADE criteria were not included in this model.

Application content refers to the quality and design of the application itself. Future ARAs should be designed to closely mimic or enhance the desired procedure/setting and should add value to the teaching experience. Furthermore, to be implemented in educational curricula, applications should provide feedback and be consumer-oriented. This may be assessed by both novices and experts in the area an ARA is designed to simulate. Positive user input on the points listed in Figure 2 demonstrate support by the ARA's intended audience.

Outcome assesses the nature of study results: statistically significant values favoring ARA use over traditional teaching methods and positive user feedback on usability and didactic potential are both needed for strong outcome metrics. While ARAs that successfully address '*Application Content*' demonstrate qualitative support for curricular integration, '*Outcome*' metrics provide additional quantitative support.

Finally, the *feasibility* module highlights the rarely-discussed factors of interest, cost, and ARA adoption outcomes. While this may be a topic better suited to entrepreneurs and application developers, future research should also understand the balance between an application's value and its barriers to implementation. Many of the ARAs described in this article, such as the VP DIANA, were not designed for consumer or educator use and therefore have less potential for curricular integration. Developing consumer-oriented applications and maintaining industry awareness of the resources required for new technologies will inform program decisions and help ensure sustainability.[51-53]

Researchers interested in developing or testing new AR technology should address each of these four categories or provide a rationale for exclusion prior to inclusion in any standardized medical curriculum. Relationships between these four components should be explored and identified to offer a holistic perspective of an ARA's qualifications for supplanting or augmenting traditional medical training methods.

Our study has several limitations. Inherent flaws to systematic reviews are the possibility that some studies were missed due to search criteria (i.e. foreign studies) and reporting bias based on method (i.e. keywords, scope, databases used, etc.). Bias is further possible in this review due to the subjectivity inherent in quality analysis and validity assessment. However, measures were

taken to minimize bias through number of independent reviewers and outside consultations as described in the Methods. Given the rapid growth of AR technology in recent years, it is also probable that research involving certain cutting-edge applications have not yet been published or are under patent/copyright restrictions, precluding their inclusion in this review. Finally, many criteria put forth in this paper regarding study quality and training potential are inherently subjective and may not be broadly applicable to every program or student population.

Commented [KT13]: E5

Conclusion

The use of AR technology in medical education is in its early stages presently lacks evidence-based support for its widespread implementation. Future research should adopt long-term and large-scale RCT or cohort study designs in keeping with the proposed model to evaluate ARA efficacy. Rigorous and standardized validation of commercially viable applications will allow the technology to be more readily integrated into medical educational curricula.

Table 1. Quality Assessment of Augmented Reality Applications in Surgery

Application (# of studies)	Design	Purpose	Stages of Validity[€]					Quality[§]
			1	2	3	4	5	
ProMIS (7)	Observational study (6) Survey (1)	Basic laparoscopic skills	x		x			Moderate
		Suturing	x		x	x		
		Laparoscopic colectomy						
ImmersiveTouch (2)	Observational studies	Ventriculostomy	x					Low
		Thoracic pedicle screw placement						
ARToolKit (1)	Observational study	Echocardiography						Very low
Vuzix 920AR (1)	Observational study	Tumor resection planning						Very low
STAR (1)	Observational study	Surgical telementoring						Very low
Brother AiRScouter WD-200B (1)	RCT	Central line insertion						Low
EyeSI (2)	RCTs	Binocular indirect ophthalmoscopy	x		x	x		Moderate
HoST UVA (1)	RCT	Urethrovesical anastomosis	x			x		Low
Google Glass (1)	Survey	Inflatable penile prosthesis placement						Very low
Prototype simulator (1)	Survey	Ultrasound-guided needle placement						Very low
Epson Moverio BT-200 (1)	Survey	Central line insertion	x					Very low
MicronTracker 2 (1)	Survey	Spinal needle insertion	x					Very low
ART (1)	RCT	Surgical telementoring	x			x		Low
Microsoft Hololens (1)	Observational study	Surgical telementoring	x			x		Very low
FLS (1)	Observational study	Peg transfer task						Very low

[€]For stages of validity, see Appendix
[§]Quality rank based on GRADE guidelines[16]

Table 2. Quality Assessment of Augmented Reality Applications in Anatomy and Classroom

Application (# of studies)	Design	Stages of Validity[€]					Quality[§]
		1	2	3	4	5	
Anatomy							
Unspecified application (1)	Survey						Very low
AR Magic Books (3)	RCTs				x		High
AR Magic Mirror (3)	Surveys	x	x				Low
Unity v5 (1)	Observational study				x		Low
BARETA (1)	Survey						Very low
Classroom							
mARble (3)	RCTs (2) Survey (1)						Low
DIANA virtual patient (1)	RCT						Low

[€]For stages of validity, see Appendix
[§]Quality rank based on GRADE guidelines[16]

Table 3. Augmented Reality Applications in Surgery

Augmented reality application	Sample size	Outcome	
ProMIS Augmented Reality Laparoscopic Simulator (Haptica, Dublin, Ireland)	55	Laparoscopic skills	Realism considered good to excellent by all participants, mixed evaluations of didactic value[21]
	18	Suturing	Significant improvement in knot scores following training with the simulator[23]
	15	Laparoscopic skills	Improvement in task completion with greater efficiency[26]
	46	Laparoscopic skills	Significant correlation between experience and performance[24]
	24	Suturing	Experienced participants had higher performance scores than novice participants[22]
	35	Laparoscopic colectomy	Simulator model rated as easier than cadaver model[20]
	115	Laparoscopic skills	Experience levels correlated strongly with simulation scores[25]

ImmersiveTouch System (ImmersiveTouch, Inc., University of Illinois, Chicago, IL, USA)	16	Ventriculostomy	AR group more likely to succeed on first attempt. Residents praised the simulator for its realism[28]
	51	Thoracic pedicle screw placement	Non-significant reduction of failure rate in screw placement[27]
ARToolKit (ARToolWorks Inc., Seattle, WA, USA)	10	Echocardiography	Trainees were able to successfully perform an ECG test[55]
Vuzix 920AR goggles (Vuzix Corp., Rochester, NY, USA)	21	Tumor resection planning	Improved non-clinician performance and significantly improved time to task completion for clinicians[56]
System for Telementoring with Augmented Reality (STAR) [Purdue University, West Lafarette, IN, USA]	20	Surgical telementoring	Less placement errors and fewer focus shifts, but took more time for each task[57]
Brother AiRScouter WD-200B AR glasses (Brother International Corp., Bridgewater, NJ, USA)	32	Central line insertion	No difference in median total procedure time between AR and control groups[58]
EyeSi augmented reality binocular indirect ophthalmoscopy simulator (VYmagic Holding AG, Mannheim, Germany)	28	Binocular indirect ophthalmoscopy (BIO)	AR group demonstrated superior total scores and performance[29]
	37	BIO	More correct sketched vessels and higher Ophthalmoscopy Training Score for AR group[30]
Hand-on Surgical Training (HoST) urethrovesical anastomosis (UVA) AR module (Roswell Park Cancer Institute and the State University of New York at Buffalo Virtual Reality Laboratory, New York, NY, USA)	52	UVA	HoST group outperformed control group on multiple measures while having lower temporal demand and mental fatigue[59]
Google Glass (Google Inc., Mountain View, CA, USA)	30	Inflatable penile prosthesis placement	81% of participants recommended implementation of application into training program; 93% felt Google Glass has a place in the operating room[60]
Unspecified prototype AR simulator	60	Ultrasound-guided needle placement	Majority positive responses for usability and training feasibility[61]
Epson Moverio BT-200 Smart Glasses (Epson America, Inc., Long Beach, CA, USA)	40	Central line insertion	Participants reported that simulation was realistic, easy to use and useful for training; 59.3% responded that AR was better than other training methods[62]
MicronTracker2 (Claron Technologies, Toronto, ON, Canada)	10	Spinal needle insertion	Overall positive responses to the system by trainees[63]
Augmented reality telementoring (ART) platform (University of Nevada School of Medicine, Las Vegas, NV, USA)	18	Surgical telementoring	After training, ART group was faster and had fewer failed attempts[64]
Microsoft Hololens (Microsoft Corp., Redmond, WA, USA)	24	Surgical telementoring	Mixed feedback on Hololens versus full telemedicine setup, no statistical difference in performance[65]
Fundamentals of Laparoscopic Surgery (FLS) module (Society of American Gastrointestinal and Endoscopic Surgeons, Los Angeles, CA, USA)	20	Standard peg transfer	Participants preferred using the timed overlay over no feedback; no difference in time to task completion or muscle fatigue[66]

Table 4. Augmented Reality Applications in Anatomy and Classroom

Augmented reality application	Sample size	Outcome
Anatomy		
Unspecified ARA	28	Positive responses for understandability and ease of use; most (70%) felt it was useful in anatomy education[67]
AR Magic Books (Various)	211	AR group scored significantly better on final assessment; most participants responded positively to AR[33]
	70	AR group scored significantly higher on academic test with lower cognitive load; all participants reported that AR facilitated learning[34]
	171	AR group had significantly higher scores than the video and notes groups; 76.9% of participants considered AR effective for studying[35]
AR Magic Mirror (ARMM) system using Microsoft Kinect (Microsoft Corp., Redmond, WA, USA)	748	Majority responded that ARMM stimulated active learning and improved structural understanding[36]
	79	Majority positive responses (80%)[37]
	68	Majority (82%) reported that ARMM facilitated knowledge retention and was easy to use[38]
Unity v5 (Unity Technologies ApS, San Francisco, CA, USA)	59	No significant difference in test scores between AR, VR and 3D modeling groups[68]
Bangor Augmented Reality Education Tool for Anatomy (BARETA) [Bangor University, Bangor, Gwynedd, UK]	34	Majority reported that BARETA helped them learn anatomical structures and was easier to use than a mouse-and-keyboard interface[69]
Classroom		
Mobile AR blended learning environment (mARble) [Peter L. Reichertz Institute for Medical Informatics at the Hannover Medical School, Hanover, Germany] (39-41)	44	AR group scored slightly higher in post-training exam but had lower knowledge retention at 14 days[41]
	10	AR group scored slightly higher in post-training exam with lower cognitive load and significantly higher hedonistic scores[39]
	6	Pragmatic quality of mARble was rated averagely, while hedonic aspects were rated above average[40]
Digital Animated Avator (DIANA) virtual patient (Medical College of Georgia, Augusta, GA, USA) (50)	84	AR group scored significantly lower in empathy and overall rating[50]

References

1. Azuma RT. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments.* 1997;6(4):355-385.
2. Zhou F, Duh HBL, Billinghurst M. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. *IEEE International Symposium on Mixed and Augmented Reality.* 2008:15-18.
3. Okuda Y, Bryson EO, Jr SD, et al. The Utility of SImulation in Medical Education: What Is the Evidence? *Mount Sinai Journal of Medicine.* 2009;76:330-343.
4. Heilig ML, Inventor. Sensorama Simulator. 1961.
5. Cummings JJ, Bailenson JN. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology.* 2015;19(2):272-309.
6. Wu H-K, Lee SW-Y, Chang H-Y, Liang J-C. Current status, opportunities and challenges of augmented reality in education. *Computers & Education.* 2013;62:41-49.
7. Zhu EG, Hadadgar A, Masiello I, Zary N. Augmented reality in healthcare education: an integrative review. *Peerj.* 2014;2.
8. Bacca J, Baldiris S, Fabregat R, Graf S. Augmented Reality Trends in Education: A Systematic Review of Research and Applications. *Educational Technology & Society.* 2014;17(4):133-149.
9. Barsom EZ, Graafland M, Schijven MP. Systematic review on the effectiveness of augmented reality applications in medical training. *Surgical Endoscopy and Other Interventional Techniques.* 2016;30(10):4174-4183.
10. Jain N, Youngblood P, Hasel M, Srivastava S. An augmented reality tool for learning spatial anatomy on mobile devices. *Clinical Anatomy.* 2017;30(6):736-741.
11. Wang LL, Wu HH, Bilici N, Tenney-Soeiro R. Gunner Goggles: Implementing Augmented Reality into Medical Education. *Studies in Health Technology Information.* 2016;220:446-449.
12. Kamphuis C, Barsom EZ, Schijven MP, Christoph N. Augmented reality in medical education? *Perspectives in Medical Education.* 2014;3:300-311.

13. Chaballout B, Maolloy M, Vaughn J, Brisson III R. Feasibility of Augmented Reality in Clinical Simulations: Using Google Glass with Manikins. *Journal of Medical Internet Research.* 2016;18(3):42.

14. Dey A, Billinghurst M, Lindeman RW, Swan JE. A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014. *Frontiers in Robotics and Ai.* 2018;5.

15. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of internal medicine.* 2009;151(4):164-169.

16. Group GW. Grading quality of evidence and strength of recommendations. *BMJ.* 2004;328:1490-1494.

17. Schijven MP, Jakimowicz JJ. Validation of virtual reality simulators: Key to the successful integration of a novel teaching technology into minimal access surgery. *Minim Invasive Ther Allied Technol.* 2005;14(4):244-246.

18. van Dongen KW, Tournoij E, van der Zee DC, Schijven MP, Broeders IA. Construct validity of the LapSim: can the LapSim virtual reality simulator distinguish between novices and experts? *Surg Endosc.* 2007;21(8):1413-1417.

19. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surg Endosc.* 2003;17(10):1525-1529.

20. LeBlanc F, Champagne BJ, Augestad KM, et al. A comparison of human cadaver and augmented reality simulator models for straight laparoscopic colorectal skills acquisition training. *J Am Coll Surg.* 2010;211(2):250-255.

21. Botden S, Buzink SN, Schijven MP, Jakimowicz JJ. ProMIS Augmented Reality Training of Laparoscopic Procedures Face Validity. *Simulation in Healthcare.* 2008;3(2):97-102.

22. Botden S, de Hingh I, Jakimowicz JJ. Meaningful assessment method for laparoscopic suturing training in augmented reality. *Surgical Endoscopy and Other Interventional Techniques.* 2009;23(10):2221-2228.

23. Botden S, de Hingh I, Jakimowicz JJ. Suturing training in Augmented Reality: gaining proficiency in suturing skills faster. *Surgical Endoscopy and Other Interventional Techniques.* 2009;23(9):2131-2137.

24. Oostema JA, Abdel MP, Gould JC. Time-efficient laparoscopic skills assessment using an augmented-reality simulator. *Surgical Endoscopy and Other Interventional Techniques.* 2008;22(12):2621-2624.

25. Nugent E, Shirilla N, Hafeez A, et al. Development and evaluation of a simulator-based laparoscopic training program for surgical novices. *Surgical Endoscopy and Other Interventional Techniques.* 2013;27(1):214-221.

26. Feifer A, Delisle J, Anidjar M. Hybrid augmented reality simulator: Preliminary construct validation of laparoscopic smoothness in a urology residency program. *Journal of Urology.* 2008;180(4):1455-1459.

27. Luciano CJ, Banerjee P, Bellotte B, et al. Learning retention of thoracic pedicle screw placement using a high-resolution augmented reality simulator with haptic feedback. *Neurosurgery.* 2011;69.

28. Yudkowsky R, Luciano C, Banerjee P, et al. Practice on an Augmented Reality/Haptic Simulator and Library of Virtual Brains Improves Residents' Ability to Perform a Ventriculostomy. *Simulation in Healthcare-Journal of the Society for Simulation in Healthcare.* 2013;8(1):25-31.

29. Rai AS, Rai AS, Mavrikakis E, Lam WC. Teaching binocular indirect ophthalmoscopy to novice residents using an augmented reality simulator. *Canadian Journal of Ophthalmology-Journal Canadien D Ophtalmologie.* 2017;52(5):430-434.

30. Leitritz MA, Ziemssen F, Suesskind D, et al. Critical evaluation of the usability of augmented reality ophthalmoscopy for the training of inexperienced examiners. *Retina.* 2014;34:785-791.

31. Billinghurst M, Kato H, Poupyrev I. The MagicBook: a transitional AR interface. *Computers & Graphics*. 2001;25(5):745-753.
32. Lee K. Augmented Reality in Education and Training. *TechTrends*. 2012;56(2):13-21.
33. Ferrer-Torregrosa J, Jimenez-Rodriguez MA, Torralba-Estelles J, Garzon-Farinos F, Perez-Bermejo M, Fernandez-Ehrling N. Distance learning ects and flipped classroom in the anatomy learning: comparative study of the use of augmented reality, video and notes. *Bmc Medical Education.* 2016;16.
34. Ferrer-Torregrosa J, Torralba J, Jimenez MA, Garcia S, Barcia JM. ARBOOK: Development and Assessment of a Tool Based on Augmented Reality for Anatomy. *Journal of Science Education and Technology.* 2015;24(1):119-124.
35. Kucuk S, Kapakin S, Goktas Y. Learning Anatomy via Mobile Augmented Reality: Effects on Achievement and Cognitive Load. *Anatomical Sciences Education.* 2016;9(5):411-421.
36. Kugelmann D, Stratmann L, Nuhlen N, et al. An Augmented Reality magic mirror as additive teaching device for gross anatomy. *Annals of Anatomy-Anatomischer Anzeiger.* 2018;215:71-77.
37. Manrique-Juan C, Grostieta-Dominguez ZVE, Rojas-Ruiz R, Alencastre-Miranda M, Munoz-Gomez L, Silva-Munoz C. A Portable Augmented-Reality Anatomy Learning System Using a Depth Camera in Real Time. *American Biology Teacher.* 2017;79(3):176-183.
38. Ma M, Fallavollita P, Seelbach I, et al. Personalized augmented reality for anatomy education. *Clinical Anatomy.* 2016;29(4):446-453.
39. Albrecht UV, Folta-Schoofs K, Behrends M, von Jan U. Effects of Mobile Augmented Reality Learning Compared to Textbook Learning on Medical Students: Randomized Controlled Pilot Study. *Journal of Medical Internet Research.* 2013;15(8).
40. Noll C, von Jan U, Raap U, Albrecht UV. Mobile Augmented Reality as a Feature for Self-Oriented, Blended Learning in Medicine: Randomized Controlled Trial. *Jmir Mhealth and Uhealth.* 2017;5(9).

41. Albrecht UV, Noll C, Von Jan U. Explore and experience mobile augmented reality for medical training. *MEDINFO.* 2013.

42. Shuhaiber JH. Augmented reality in surgery. *Archives of Surgery.* 2004;139(2):170-174.

43. Jakimowicz JJ, Jakimowicz CM. Simulation in Surgery. Where Are We Now and Where to From Here? *Cirugia Y Cirujanos.* 2011;79(1):41-45.

44. Kneebone R. Simulatoin in surgical training: educational issues and practical implications. *Medical Education.* 2003;37(3):267-277.

45. Satava RM. Surgical Education and Surgical Simulation. *World Journal of Surgery.* 2001;25:1484-1489.

46. Bloom B, Engelhart M, Furst E, Hill W, Krathwohl D. *Taxonomy of Education Objectives: the Classification of Educational Goals.* New York: David McKay Company, Inc.; 1956.

47. Paisley A, Baldwin P, Paterson-Brown S. Validity of surgical simulation for the assessment of operative skill. *British Journal of Surgery.* 2001;88(11):1525-1532.

48. Anastakis D, Regehr G, Reznick R, et al. Assessment of technical skills transfer from the bench training model to the human model. *American Journal of Surgery.* 1999;177(2):167-170.

49. Nicholson DT, Chalk C, Funnell WRJ, Daniel SJ. Can virtual reality improve anatomy education? A randomised controlled study of a computer-generated three-dimensional anatomical ear model. *Medical Education.* 2006;40(11):1081-1087.

50. Deladisma AM, Cohen M, Stevens A, et al. Do medical students respond empathetically to a virtual patient? *American Journal of Surgery.* 2007;193(6):756-760.

51. Cook DA, Ellaway RH. Evaluating Technology-enhanced Learning: A Comprehensive Framework. *Medical Teacher.* 2015;37(10):961-970.

52. Clune WH. Methodological strength and policy usefulness of cost-effectiveness research. In: Levin HM, Mcewan PJ, eds. *Cost-effectiveness and educational policy*. Larchmont, NY: Eye On Education; 2002:55-68.

53. Hummel-Rossi B, Ashdown J. The state of cost-benefit and cost-effectiveness analyses in education. *Review of Educational Research.* 2002;72:1-30.

54. Bifulco P, Narducci F, Vertucci R, PAmbruosi P, MCesarelli M, Romano M. Telemedicine supported by Augmented Reality: an interactive guide for untrained people in performing an ECG test. *BioMedical Engineering OnLine.* 2014;13:153.

55. Abhari K, Baxter JSH, Chen ECS, et al. Training for Planning Tumour Resection: Augmented Reality and Human Factors. *Ieee Transactions on Biomedical Engineering.* 2015;62(6):1466-1477.

56. Andersen D, Popescu V, Cabrera ME, et al. Medical telementoring using an augmented reality transparent display. *Surgery.* 2016;159(6):1646-1653.

57. Huang CY, Thomas JB, Alismail A, et al. The use of augmented reality glasses in central line simulation: "see one, simulate many, do one competently, and teach everyone". *Advances in Medical Education and Practice.* 2018;9:357-363.

58. Chowriappa A, Raza SJ, Fazili A, et al. Augmented-reality-based skills training for robot-assisted urethrovesical anastomosis: a multi-institutional randomised controlled trial. *Bju International.* 2015;115(2):336-345.

59. Dickey RM, Srikishen N, Lipshultz LI, Spiess PE, Carrion RE, Hakky TS. Augmented reality assisted surgery: a urologic training tool. *Asian Journal of Andrology.* 2016;18(5):732-734.

60. Magee D, Zhu Y, Ratnalingam R, Gardner P, Kessel D. An augmented reality simulator for ultrasound guided needle placement training. *Medical & Biological Engineering & Computing.* 2007;45(10):957-967.

61. Rochlen LR, Levine R, Tait AR. First-Person Point-of-View-Augmented Reality for Central Line Insertion Training. *Simulation in Healthcare-Journal of the Society for Simulation in Healthcare.* 2017;12(1):57-62.

62. Sutherland C, Hashtrudi-Zaad K, Sellens R, Abolmaesumi P, Mousavi P. An Augmented Reality Haptic Training Simulator for Spinal Needle Procedures. *Ieee Transactions on Biomedical Engineering.* 2013;60(11):3009-3018.

63. Vera AM, Russo M, Mohsin A, Tsuda S. Augmented reality telementoring (ART) platform: a randomized controlled trial to assess the efficacy of a new surgical education technology. *Surgical Endoscopy and Other Interventional Techniques.* 2014;28(12):3467-3472.

64. Wang S, Parsons M, Stone-McLean J, et al. Augmented Reality as a Telemedicine Platform for Remote Procedural Training. *Sensors.* 2017;17(10).

65. Zahiri M, Nelson CA, Oleynikov D, Siu KC. Evaluation of Augmented Reality Feedback in Surgical Training Environment. *Surgical Innovation.* 2018;25(1):81-87.

66. da Silva ICS, Klein G, Brandao DM. Segmented and Detailed Visualization of Anatomical Structures based on Augmented Reality for Health Education and Knowledge Discovery. *Advances in Science, Technology and Engineering Systems Journal.* 2017;2(3):469-478.

67. Moro C, Stromberga Z, Raikos A, Stirling A. The Effectiveness of Virtual and Augmented Reality in Health Sciences and Medical Anatomy. *Anatomical Sciences Education.* 2017;10(6):549-559.

68. Thomas RG, John NW, Delieu JM. Augmented reality for anatomical education. *J Vis Commun Med.* 2010;33(1):6-15.

Figure 1: PRISMA Flow Diagram

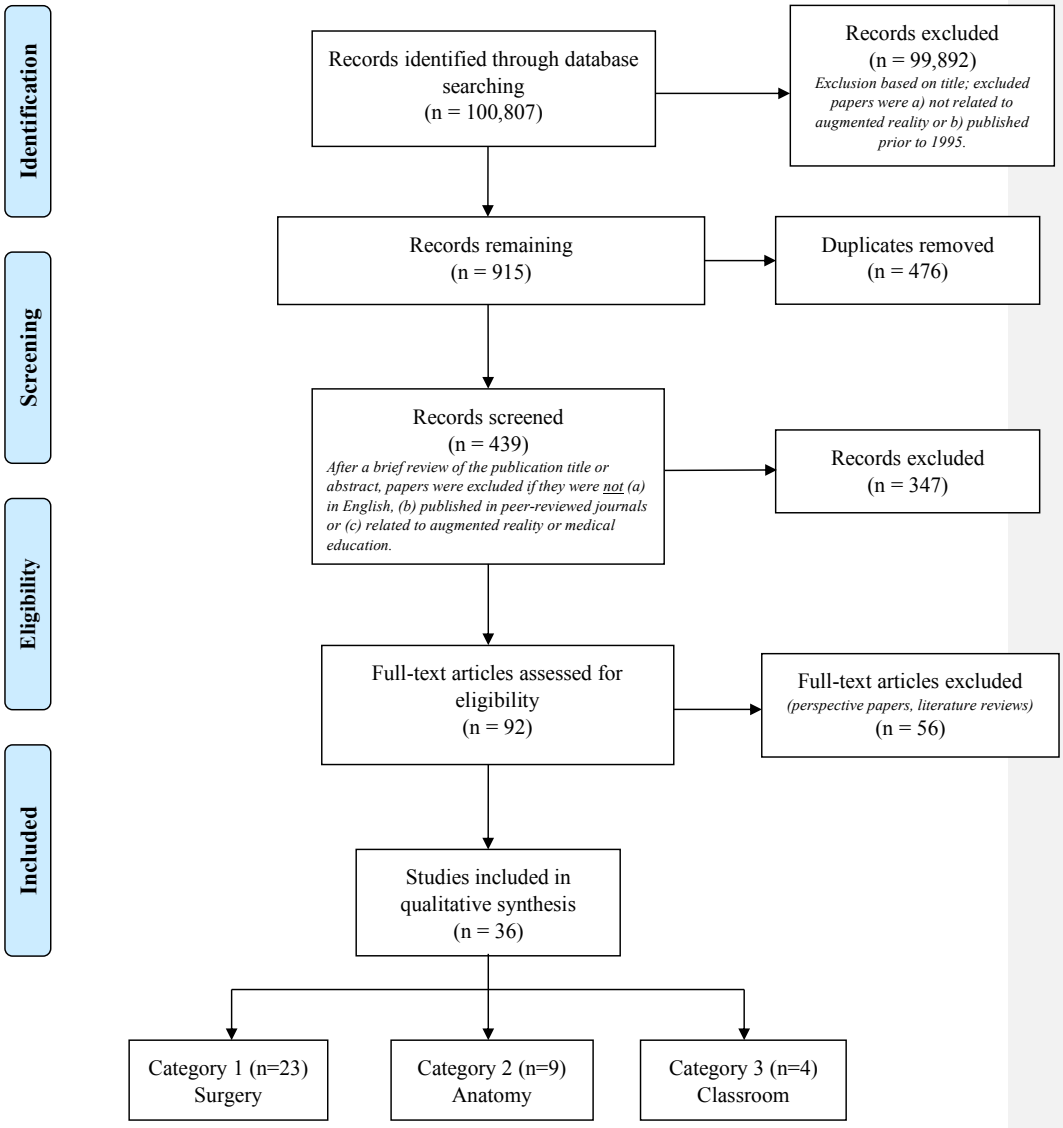
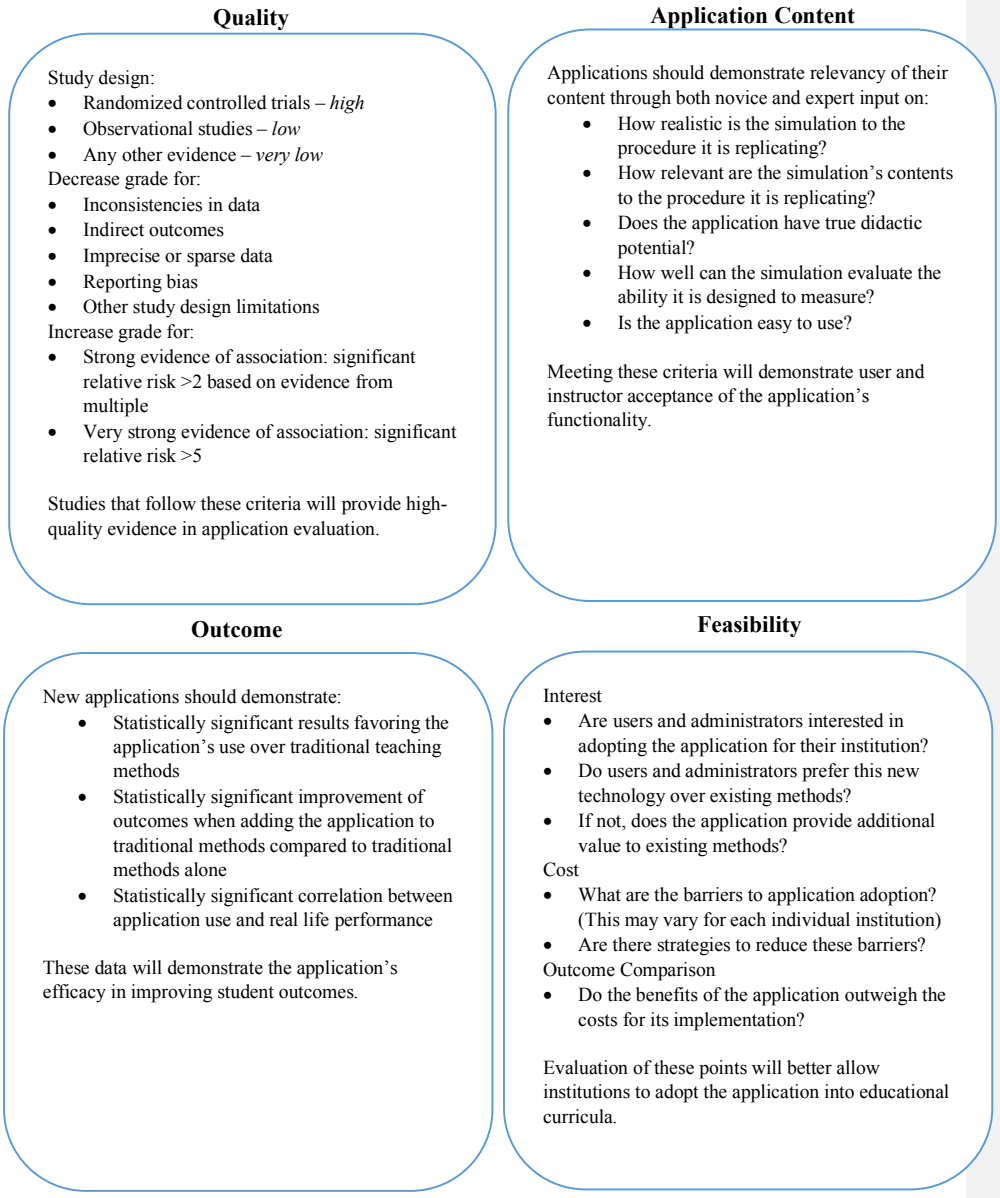


Figure 2. Augmented Reality Research Model for Curricular Integration



Appendix

Validity of Augmented Reality Applications (ARAs) in Medical Education[9,17,19]

Stages of Validity	Definition	Demonstration Criteria
1) Face Validity	The degree to which the simulation resembles the actual construct (procedure) that it seeks to replicate	Positive feedback on the realism of the ARA by both experts and learners
2) Content Validity	The degree to which the simulation's contents are relevant to the subject matter of the construct it seeks to replicate	Positive feedback on the simulation's setting and scoring system by medical experts
3) Construct Validity	The degree to which the simulation can evaluate the quality or ability it was designed to measure	Simulation outcomes are positively and significantly correlated with the user's skill level
4) Concurrent Validity	The degree to which the simulation scores correlate with the scores on an alternate "gold standard" tool or training method	Simulation outcomes are related to/like the scores on a previously established training method
5) Predictive Validity	The degree to which the simulation scores correlate with actual performance in the construct it seeks to replicate	Statistically significant correlation between simulation outcomes and actual procedural performance