



**HAL**  
open science

## Head Mounted Displays in Surgical Training and Planning: A Literature Review

Eléonore Ferrier-Barbut, Ignacio Avellino, Marie-Aude Vitrani, Geoffroy Canlorbe

► **To cite this version:**

Eléonore Ferrier-Barbut, Ignacio Avellino, Marie-Aude Vitrani, Geoffroy Canlorbe. Head Mounted Displays in Surgical Training and Planning: A Literature Review. 34e conférence internationale francophone sur l'Interaction Humain-Machine (IHM'23), AFHIM, Apr 2023, Troyes et Bar sur Aube, France. 10.1145/3583961.3583973 . hal-04019531

**HAL Id: hal-04019531**

**<https://hal.science/hal-04019531v1>**

Submitted on 8 Mar 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Head Mounted Displays in Surgical Training and Planning: A Literature Review

## Les Visiocasques dans la Formation et la Planification Chirurgicales : Une Revue de la Littérature

Eléonore Ferrier-Barbut  
eleonore.ferrier-barbut@sorbonne-universite.fr  
Sorbonne Université, ISIR  
Paris, France

Marie-Aude Vitrani  
marie-aude.vitrani@sorbonne-universite.fr  
Sorbonne Université, ISIR  
Paris, France

Ignacio Avellino  
ignacio.avellino@sorbonne-universite.fr  
Sorbonne Université, CNRS, INSERM, ISIR  
Paris, France

Geoffroy Canlorbe  
geoffroy.canlorbe@aphp.fr  
Sorbonne Université, INSERM UMRS 938, CRSA. AP-HP,  
Pitié-Salpêtrière University Hospital  
Paris, France

### ABSTRACT

Training and planning are two important preclinical activities when preparing for surgery. Although Head Mounted Displays (HMDs) are progressively being studied for training and planning in Mixed Reality (MR), clinicians need scientific evidence before making decisions when institutionalizing the technology. We investigate the up-to-date demonstrated benefits of MR-HMDs for preclinical activities through a systematic literature review using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The results indicate little evidence-based findings. Only twelve studies were eligible, nine Randomized Controlled Studies (RCT), and three non-RCT comparative studies. Regarding training, even if there is evidence demonstrating learning benefits, these do not seem to transpose to real environments. Regarding planning, we see little-to-no evidence supporting MR-HMDs compared with traditional methods. We discuss possible directions so that future studies can increase the level of evidence, as well as how the technology can evolve to better support the particular needs of surgery.

### RÉSUMÉ

L'apprentissage et la planification sont deux activités précliniques importantes lors de la préparation d'une intervention chirurgicale. Bien que les visiocasques (HMDs) soient progressivement étudiés pour la formation et la planification en Réalité Mixte (RM), les cliniciens ont besoin de preuves scientifiques avant de prendre des décisions concernant l'institutionnalisation de cette technologie dans les programmes de formation. Nous étudions les avantages actuellement démontrés des HMD-RM pour les activités précliniques par le biais d'une revue systématique de la littérature en utilisant les directives PRISMA (Preferred Reporting Items for Systematic

Reviews and Meta-Analyses). Les résultats indiquent peu de résultats basés sur la preuve. Seules douze études étaient admissibles, neuf études contrôlées randomisées (ECR) et trois études comparatives non-ECR. En ce qui concerne la formation, il existe des preuves démontrant les avantages de l'apprentissage, mais celles-ci ne semblent pas se transposer aux environnements réels. En ce qui concerne la planification, il n'existe que peu ou pas de preuves de l'utilité des HMDs-RM par rapport aux méthodes traditionnelles. Nous discutons des orientations possibles pour les études futures afin d'augmenter la qualité des études menées et de faire évoluer la technologie pour mieux répondre aux besoins particuliers de la chirurgie.

### CCS CONCEPTS

• **Human-centered computing** → **Mixed / augmented reality.**

### KEYWORDS

Head Mounted Displays, Mixed Reality, Surgical Training, Surgical Planning

### MOTS CLÉS

Visiocasques, Réalité Mixte, Formation Chirurgicale, Planification Chirurgicale

### ACM Reference Format:

Eléonore Ferrier-Barbut, Ignacio Avellino, Marie-Aude Vitrani, and Geoffroy Canlorbe. 2023. Head Mounted Displays in Surgical Training and Planning: A Literature Review. In *IHM '23: Proceedings of the 34th Conference on l'Interaction Humain-Machine (IHM '23)*, April 3–6, 2023, TROYES, France. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3583961.3583973>

## 1 INTRODUCTION

Surgical training and planning are both essential preclinical activities to anticipate and prepare for surgical interventions. *Surgical training* is a crucial part of the surgical school curricula where students learn not only by observing, but also by doing without being afraid of making mistakes. Traditional preclinical training involves either dry labs (exercises) or wet labs (animals, cadavers), where students learn the various aspects of a surgical procedure: the surgical steps, the anatomical structures involved, which instruments to use, the risks involved, the surgical gestures performed, and how to interact with other team members during the procedure. Nonetheless, most surgical training takes place in a clinical setting by means of observation. The Operating Room (OR) is the only place where students learn about specific cases and watch how experienced surgeons react to unusual situations and apply their “know-how” acquired over the years [10]. Yet, good quality training is hampered by the increasing number of students in ORs and the accelerated pace due to the high amount of daily procedures [23]. *Surgical planning* is the preoperative method of pre-visualising a surgical intervention, to define the surgical steps in advance. Pre-operative planning systems are mainly composed of visualization, modeling, analysis, and plan generation [8]. Visualization involves working with patient images (e.g. MRI), identifying structures of interest, implants and/or surgical tools in a way that is intuitive and useful for the clinician. Modeling consists of creating mathematical representations of the structures of interest, the surgical tasks and their constraints, and the physiological phenomena that are taken into account for the planning. Through analysis, clinicians explore the solution space of the planning problem by manual exploration, simulation, and/or optimization. Lastly, a plan is generated, selecting a solution that is most appropriate for the intervention based on the results of the visualization and the analysis. These tools, although essential parts of a successful surgical procedure, often require the support of an assistant to access the information they carry and may compromise sterility during surgery [34].

Head Mounted Displays (HMDs) are being more and more studied to address the challenges in surgical training and planning. The technology enables adapting to student’s needs as well as a more step-wise learning and planning mode while ensuring patient safety [9]. Particularly for surgery, HMDs are increasingly being used because they are wearable, enabling surgeons to interact with digital information through mid-air gestures without breaking sterility, as is the case for other technologies such as tablets [1]. Moreover, HMDs have been shown to improve the learning of both technical and procedural skills in general and in surgery especially, and increase self-confidence and participant satisfaction<sup>1</sup>. We distinguish between two types of HMD: Augmented Reality (AR) or Virtual Reality (VR). AR-HMDs overlays digital information on top of a real-world environment in real time thereby “augmenting” the user’s vision while maintaining the haptics of the real world [20]. VR-HMD, on the other hand, exclude the real world and provide a fully-immersive experience. Several studies show that AR and VR are useful for learning surgical planning by defining the optimal surgical strategy [5, 6, 31, 36]. Thus, the use of AR and VR

appears promising as a supplementary operating tool [35], and has received encouraging usability score results by both Dickey et al. [7] and Rochlen et al. [27] who report feedback surveys with medical students, residents and faculty. However, we observe few studies that use AR- and VR-HMDs in the context of surgical training and planning, but, in order to institutionalize this technology, clinicians need strong evidence that demonstrates the added value. We review the literature to investigate the current level of evidence regarding the benefits of AR and VR HMDs in the field of surgical learning and surgical planning, to understand how HMDs can be used effectively in this context.

## 2 MATERIALS AND METHODS

We conducted a systematic review applying the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Figure 1). The PubMed and ClinicalTrials databases (both from U.S. National Library of Medicine) were queried with the search terms: head mounted displays, headset, surgery, augmented reality, virtual reality. We chose the PubMed and ClinicalTrials databases because the goal was to review randomized controlled studies, and this is where they are mostly published.

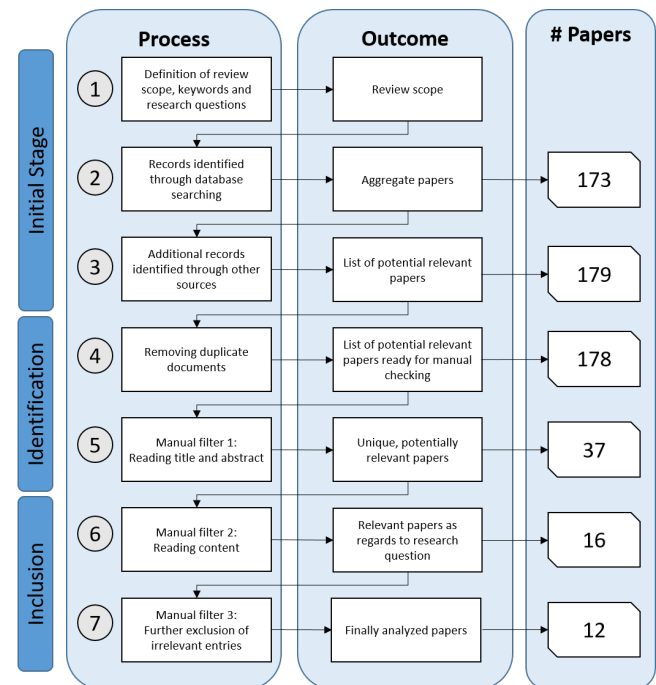


Figure 1: PRISMA flow diagram of article selection.

We followed the PRISMA guidelines performing 4 main steps.

- (1) **Identification:** We searched the PubMed and ClinicalTrials databases, as well as our personal article collections with the terms head mounted displays, headset, surgery, augmented reality, virtual reality. We only included papers published in English between January 1, 2000 to September 30, 2020. We identified 179 articles, from which two were duplicate, resulting in 178 unique articles.

<sup>1</sup>[https://www.reseau-canope.fr/fileadmin/user\\_upload/Projets/agence\\_des\\_usages/Etat\\_Art.pdf](https://www.reseau-canope.fr/fileadmin/user_upload/Projets/agence_des_usages/Etat_Art.pdf)

- (2) **Screening:** We screened the 178 articles by reading the titles and abstracts. We removed research articles that did not relate to VR- or AR-HMD visualization either as assistance to visualization or training for surgery. We excluded papers that focused on technological contributions to VR and AR, rather than their evaluation. A total of 141 articles were removed.
- (3) **Eligibility:** We assessed articles eligibility through their level of scientific evidence. We only included articles with a level A or B of scientific evidence. We decided to keep only level A or B scientific evidence, as level C is considered to be a weak level of scientific evidence by the Haute Autorite de Sante (HAS)<sup>2</sup>, the French National Authority for Health, while level A represents an *established scientific evidence* and level B a *scientific presumption*. Level A is defined as scientific evidence from high-powered randomized controlled trials (RCTs), meta-analyses of randomized controlled trials, and decision analyses based on well-conducted studies. Level B is defined as scientific evidence from low-powered RCTs, well-conducted non-randomized comparative studies, and cohort studies. A total of 37 articles were eligible for review at this step.
- (4) **Inclusion:** We fully read the articles and removed those that involved non-HMD AR or VR, or that did not specifically relate either to planning or training for surgery, and that had been missed in Screening. We also removed articles that were neither Level A nor Level B of scientific evidence, not previously detected.

We studied the benefits of using HMDs for surgical learning, including technical and non-technical surgical skills and surgeons' satisfaction, as well as for surgical planning. We describe the articles along with their level of evidence on the Kirkpatrick hierarchy [16]. The Kirkpatrick pyramid is used to rank medical education studies depending on the main outcome in four levels: 1 - student satisfaction; 2 - progress in theoretical and practical knowledge outside the clinic; 3 - benefit on clinical skills or behavior; and, 4 - impact or benefit for the patient.

### 3 RESULTS

We retained twelve studies for analysis: nine RCTs [11, 15, 18–20, 25, 32, 39], and three non-randomized controlled comparative studies [13, 14, 38]. We present our results first for surgical learning and then for planning, distinguishing between AR and VR.

#### 3.1 HMDs for Surgical Learning

We present articles that investigate the use of AR and VR through HMDs for surgical learning, summarized in Table 1 (Appendix A).

**3.1.1 Augmented Reality Head Mounted Displays.** Only one article by Logishetty Western et al. [20] compares training using an AR-HMD to traditional training (Level 2 Kirkpatrick). The authors focus on the development of technical skills for Total Hip Arthroplasty (THA) through the use of an AR-HMD with live holographic orientation feedback. Twenty-four medical students were randomized

into two groups and received four training sessions: one group with using an AR-HMD and one with one-on-one training from a hip arthroplasty surgeon. Post-training improvements and accuracy of acetabular component orientation were assessed. Authors did not find a difference between groups in the post-training test: mean difference 1.2°, 95% CI [-1.8°– 4.2°];  $p = 0.281$ . During the four training sessions, there were fewer mean errors (6 SD) in orientation of the acetabular component in the AR-guided group than in the surgeon-guided group:  $1^\circ \pm 1^\circ$  AR  $6^\circ \pm 4^\circ$ ,  $p < 0.001$ .

**3.1.2 Virtual Reality Head Mounted Displays.** Eight articles [11, 13–15, 18, 19, 25, 39] compare traditional training to training through a VR-HMD. We observe three types of simulations: Virtual Environments (VE), videos, or both combined.

- VR-HMDs presenting a VE: Two articles, both by Logishetty et al. [18, 19], focus on VE for training students to perform THA and on improvement in technical skills. One article [19] compares VR training with traditional training using the procedure-based assessment (PBA) score on a cadaver; while the other [18] compares performance in the physical world before and after VR training, as well as performance after VR training vs. experts' performance. Both articles showed a significant improvement in performance after training with VR, and the second [18] reports significantly better performance for THA when training in VR compared with traditional training. The experiments included 56 resident participants and are Level 2 on the Kirkpatrick scale.
- VR-HMDs presenting videos: Three articles compare VR training using videos to traditional training. In two of the articles [11, 39], the authors conducted an experimental protocol in training conditions (Level 2 Kirkpatrick) while in the third [15] they tested VR-HMD training as part of a veterinary students' curriculum (Level 3 Kirkpatrick). In the article by Yoganathan et al. [39], the authors compared 360° video training with a VR-HMD alone or in combination with face-to-face training with 2D videos alone or in combination with face-to-face training, for knot tying (i.e. technical skills training). Ability to tie a knot was significantly better for VR-trained participants compared to 2D video-trained participants (median knot score 5.0 vs. 4.0,  $p = 0.04$ ) even when both conditions were combined with face-to-face training (median knot score 9.5 vs. 9.0,  $p = 0.01$ ). In the article by Hunt et al. [15] the authors compared preparation time for veterinary students preparing their first live surgery after training with stereoscopic videos viewed in a VR-HMD or traditional training. Preparation time was not significantly shorter in the VR-HMD group ( $p = 0.66$ ). Finally, Harrington et al. [11] compared the engagement level as measured by Task Unrelated Image or Thoughts (TUIT) and the subjective level of engagement between training with 360° videos viewed in a VR-HMD and traditional 2D formats. The level of engagement was significantly higher in the group who viewed 360° videos ( $p < 0.0001$ ) and the number of TUITs significantly lower ( $p < 0.0001$ ).

<sup>2</sup>[https://www.has-sante.fr/upload/docs/application/pdf/2013-06/etat\\_des\\_lieux\\_niveau\\_preuve\\_gradation.pdf](https://www.has-sante.fr/upload/docs/application/pdf/2013-06/etat_des_lieux_niveau_preuve_gradation.pdf)

- VR-HMDs presenting a combination of VE and videos: Three articles describe the impact of training with a VR-HMD showing a fully VE compared with a VR-HMD showing a combination of VE and videos on performance and skills in surgery [13, 14, 25]. Huber et al. published two articles [13, 14] on satisfaction and performance while Pulijala et al. [25] focused on the interest of videos and VE viewed in a VR-HMD for perceived self-confidence of surgical students. Huber et al. [13] found that immersive reality (including a laparoscopic simulator and 360° videos of the OR) viewed in a VR-HMD enhanced a feeling of exhilaration and of “being there” without impacting performance compared with a standard VR laparoscopic simulator. Huber et al. [14] did not find a significant difference on technical performance in laparoscopic surgery ( $p > 0.5$ ) when training with a VR-HMD with 360° videos of the OR combined with a VE compared with training with a VR-HMD with a VE alone. The metrics analyzed – time, handling ergonomics, mistakes – relate to technical skills only. When analyzing perceived self-confidence of residents in oral maxillofacial surgery before and after training, Pulijala et al. [25] reported that students trained with 360° OR videos and 3D stereoscopic videos through an HMD gained in self-confidence vs. those who trained with standard 2D video. Authors also report on a non-significant improvement in knowledge acquisition.

## 3.2 HMDs for Surgical Planning

We present articles that investigate the use of AR and VR through HMDs for surgical planning, summarized in Table 2 (Appendix A).

**3.2.1 Augmented Reality Head Mounted Displays.** Wellens et al. [38] present a study investigating whether AR holograms viewed through an AR-HMD and through 3D prints improve the visibility of four anatomical structures – tumors, arteries, veins, and urinary collecting structures – compared with conventional imaging (MRI and/or CT scan), in 10 children with Wilms tumors. Seven pediatric oncology surgeons completed a questionnaire after interacting with conventional imaging, and their AR holograms obtained with 3D prints. The surgeons rated the AR holograms and the 3D prints as better because they improve anatomical understanding than conventional imaging for all the anatomical structures ( $p = 0.002$ ); arteries ( $p < 0.001$ ), veins ( $p < 0.001$ ), and urinary collecting structures ( $p < 0.001$ ) This study is of Level 1 Kirkpatrick.

**3.2.2 Virtual Reality Head Mounted Displays.** We found only two articles dealing with VR-HMDs presenting a VE [32, 33], both by the same team of researchers. They assess how 3D virtual models viewed in a VR-HMD influence operative time compared with usual preoperative planning for robot-assisted partial nephrectomy. The former [32], with 44 patients in the study group and 48 patients in the control group, found that operative time (OR, 1.00; 95% CI [0.37 – 2.70]) was not statistically different between groups in this model. In contrast, Shirk, Kwan et al. [32] found that patients with VR-assisted surgical planning had a significantly lower operative time (141 minutes vs. 201 minutes,  $p < 0.0001$ ) after controlling for case complexity and surgeon experience in 30 patients. These studies analyzed the impact of planning surgery with a VR headset on patient outcome and are Level 4 Kirkpatrick studies.

## 4 DISCUSSION

Our review reinforces the notion that HMDs are relevant for surgical training in terms of self-confidence [25], performance [18–20, 39], feeling of immersion [13, 14], as well as surgical planning in terms of reduction of operative time [32] and visibility of anatomical structures [38]. Our review is a documented trace that up to late 2020, there is scarce evidence that can be used to establish HMDs as beneficial for training and planning. There is ample room in this topic for further research, and for future reviews to show changes in the trends of such research. We now expand on the need for future studies that can demonstrate the benefits of HMDs, and future directions on the development of the technology.

### 4.1 Lack of Demonstrated Benefits and Future Studies

*Clinical Outcome Measures.* Only two [32, 33] of the 12 reviewed articles reported patient-related outcomes. It thus remains unclear if immersive training has an effect on measures such as recovery time or post-operative complications. More Level 3 and 4 Kirkpatrick studies are required to provide solid evidence about the impact of VR and AR on clinical outcomes, and to enable institutions to make decisions on how to incorporate the technology into their practice.

*The Lack of Use of Standard Protocols.* Standard protocols have existed for a number of years [12, 28]. However, the studies we reviewed use a wide variety of experimental protocols, making it difficult to draw solid conclusions. For example, while one article compares pre and post training performance [18], another compares post training performance with an HMD and without and HMD [20]. Similarly, one article evaluates TIUT (Task Unrelated Image or Thoughts) [11], while another focuses on self-confidence [33]. We see much value in future work that defines standard protocols when studying the impact of technology in preclinical activities, as these will result in studies producing similarly-structured results that can be compared, leading to meaningful generalizations. Some of the standards that such protocol could define, include participant’s level of experience, both in surgery and the technology used, the learning or planning environment (classroom, dry labs, wet labs or the OR), and the content visualized using the technology such as anatomy, patient information, feedback on performance or procedure steps to name a few. In this way, future reviews and meta analyses can make conclusions about the different technologies (AR-HMD, VR-HMD, 2D screens) by looking at studies that investigate each type of technology in isolation.

*Non-Technical Skill Measures.* Our review points towards little evidence for the impact of HMDs on non-technical skills in surgery, as we only found one study [33] that investigated self-confidence. We encourage future research to pursue investigations on non-technical skills, as they play a key role in performance when learning and planning, sometimes with a greater impact than technical skills. By incorporating these type of measures, future work can bring a more complete understanding of the factors that impact surgical learning and planning.

*Comparing Immersive and Non-Immersive AR and VR.* We observe very limited evidence on the relevance of immersive AR both for training and planning. These limited evidence include one study by Omlor et al. [22] from 2022 comparing immersive and non-immersive medical training VR videos, showing no significant difference between the two groups in the learning-success test score, although the participants in the immersive videos group rated the learning atmosphere, comprehensibility, and overall recommendation of the course significantly higher than in the non-immersive videos group. Participants also significantly perceived that they gained a better understanding of the presented procedures, and that VR can be an appropriate tool for online teaching. Another work by Carreon et al. [4] from 2021 comparing immersive and non-immersive VR on social skill acquisition for students in middle school shows no significant differences in acquisition between immersive and non-immersive conditions. The participants reported an almost identical level of presence, showing a similar perception of both devices. Still, we know from previous work studying non-immersive AR [5, 6, 17, 31, 35–37] and immersive AR [7, 27] (not included in our corpus as they are not RCTs) that AR during a surgical intervention can improve competency, surgical opinion and postoperative complication rates, although increasing operative time. Regarding immersive VR, it seems to be particularly suited for training, which is in line with previous work both in non-immersive VR in surgery [26, 30] as well as immersive VR in education [24]. Nonetheless, very few studies compare immersive vs. non-immersive AR and VR, and thus we have limited insights on what are the tradeoffs of each approach regarding its use as immersive or not. We strongly encourage that future research conducts experiments comparing AR and VR in an HMD to AR and VR on a 2D screen, providing key insights about the added value of using HMDs, regardless of the visualized content.

*Consideration of Technological Costs.* None of the reviewed articles considered the technological cost of using HMDs. Immersive training with VR and AR can add extra costs, such as the setup time, or the added cognitive load from interacting with a new system, which are important factors to consider in order to avoid unintended errors. It is important to understand these costs in order to put the demonstrated benefits in perspective.

## 4.2 Future Research in HMDs

*Considering Multiple Perceptual Senses.* None of the reviewed articles consider the study of perceptual senses other than vision, both for surgical training or planning using HMDs. They focus on visual cues for immersive experiences but overlook other perceptual senses such as haptic feedback. Nonetheless, haptic feedback is key when learning surgery [40] and when performing safe surgery [3]. When technologies, such as surgical robots, remove haptic feedback, studies show that it incurs a shift in the way surgeons work and teams collaborate [2]. Regarding HMDs, Lemole et al. [17] and Shakur et al. [29] both conducted non-RCTs (and thus not included in this review) including haptic feedback in a VR-HMD Virtual Environment for neurosurgical education, showing that HMDs were of sufficient maturity to be useful in neurosurgery where training is an especially long and laborious process requiring many years of directed hands-on training. These articles pave the way

for future work to investigate the potential of immersive learning with the addition of sensory dimensions beyond visual perception.

*Towards Techno-Pedagogical Systems.* To finish our discussion, we ask the question on the role of HMD-based systems. An area for improvement in technologies for surgical training is to not only focus on the selection of the most appropriate *technological means* used to display the training, but also on the exploration of the relevant *pedagogical means* (e.g. teaching and learning methods, student motivation, or previous skills and knowledge) to teach and convey the surgical knowledge. One future research direction is to build a surgical training on a pedagogical framework for developing the studied technology, which the review articles overlook. The way in which media, such as video and audio, is used for pedagogical purposes influences its efficacy [24]. However, current studies focus on evaluating the technology in a clinical setting and overlook pedagogical aspects when producing the media and content to be displayed in the HMD, including: the use of learning theories from human sciences, the pedagogical context, the pedagogical scenario (e.g. learning material, activities and environment) and the instructional design (e.g. the organization, methods and strategies). Additionally, studies currently focus on short-term outcomes and overlook the long-term effects of learning: while an HMD can be a great tool to visualize information, it is important to understand the learning curve and whether there is a lasting effect on the skills acquired. We believe that future work can leverage the unique characteristics of HMDs to increase the learners' acquisition of skills and knowledge. For instance, using the immersive aspect to explore different pedagogical scenarios: involving remote and co-located training, ability to embody from different points of view (e.g. surgeon, nurse or patient), giving control to access on-demand various learning materials, all while relying on well-constructed instructional design that involves real-world problems, activation of adequate knowledge and skills based on previous experience, clear explanation of learning objectives, and, integration of developed skills in the OR [21]. In this way, we would move from purely technological systems to techno-pedagogical systems.

## 5 CONCLUSION

In this literature review we studied the impact of AR and VR visualization through HMDs on surgery training and planning. Both AR- and VR-HMDs would appear to provide advantages in different ways. We believe that each modality has its uniqueness, and thus the future of HMDs in the context of learning and planning for surgery lies in the combination of different devices depending on knowledge and skills targeted in a techno-pedagogical scenario, which can be found by standardizing protocols.

## ACKNOWLEDGEMENTS

This research was supported by the Labex CAMI (project ANR-11-LABX-0004) operated by the ANR as part of the program "Investissement d'Avenir".

## REFERENCES

- [1] Daniel Andersen, Voicu Popescu, Maria Eugenia Cabrera, Aditya Shanghavi, Gerardo Gomez, Sherri Marley, Brian Mullis, and Juan P. Wachs. 2016. Medical telementoring using an augmented reality transparent display. *Surgery* 159, 6 (June 2016), 1646–1653. <https://doi.org/10.1016/j.surg.2015.12.016>

- [2] Ignacio Avellino, Gilles Bailly, Geoffroy Canlorbe, Jérémie Belgiht, Guillaume Morel, and Marie-Aude Vitrani. 2019. Impacts of Telemanipulation in Robotic Assisted Surgery. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 583:1–583:15. <https://doi.org/10.1145/3290605.3300813> event-place: Glasgow, Scotland, UK.
- [3] Omar S. Bholat, Randy S. Haluck, Willie B. Murray, Paul J. Gorman, and Thomas M. Krummel. 1999. Tactile feedback is present during minimally invasive surgery. *Journal of the American College of Surgeons* 189, 4 (Oct. 1999), 349–355. [https://doi.org/10.1016/S1072-7515\(99\)00184-2](https://doi.org/10.1016/S1072-7515(99)00184-2)
- [4] Adam Carreon. 2021. *A Comparison of Immersive VR and Non-Immersive VR on Social Skill Acquisition for Students in Middle School with ASD*. Ph.D. Dissertation. University of Kansas.
- [5] Matej Cimerman and Anze Kristan. 2007. Preoperative planning in pelvic and acetabular surgery: the value of advanced computerised planning modules. *Injury* 38, 4 (2007), 442–449.
- [6] Jodi L Crossingham, Jodie Jenkinson, Nick Woolridge, Steven Gallinger, Gordon A Tait, and Carol-Anne E Moulton. 2009. Interpreting three-dimensional structures from two-dimensional images: a web-based interactive 3D teaching model of surgical liver anatomy. *HPB* 11, 6 (2009), 523–528.
- [7] Ryan M Dickey, Neel Srikishen, Larry I Lipshultz, Philippe E Spiess, Rafael E Carrion, and Tariq S Hakky. 2016. Augmented reality assisted surgery: a urologic training tool. *Asian journal of andrology* 18, 5 (2016), 732.
- [8] Caroline Essert and Leo Joskowicz. 2020. Chapter 32 - Image-based surgery planning. In *Handbook of Medical Image Computing and Computer Assisted Intervention*. S. Kevin Zhou, Daniel Rueckert, and Gabor Fichtinger (Eds.). Academic Press, 795–815. <https://doi.org/10.1016/B978-0-12-816176-0.00037-5>
- [9] Randy S Haluck and Thomas M Krummel. 2000. Computers and virtual reality for surgical education in the 21st century. *Archives of surgery* 135, 7 (2000), 786–792.
- [10] JM Hamdorf and John C Hall. 2000. Acquiring surgical skills. *British Journal of Surgery* 87, 1 (2000), 28–37.
- [11] Cuan M Harrington, Dara O Kavanagh, Gemma Wright Ballester, Athena Wright Ballester, Patrick Dicker, Oscar Traynor, Arnold Hill, and Sean Tierney. 2018. 360 Operative videos: A randomised cross-over study evaluating attentiveness and information retention. *Journal of surgical education* 75, 4 (2018), 993–1000.
- [12] David J Harris, Jonathan M Bird, Philip A Smart, Mark R Wilson, and Samuel J Vine. 2020. A framework for the testing and validation of simulated environments in experimentation and training. *Frontiers in Psychology* 11 (2020), 605.
- [13] Tobias Huber, Markus Paschold, Christian Hansen, Hauke Lang, and Werner Kneist. 2018. Artificial versus video-based immersive virtual surroundings: analysis of performance and user's preference. *Surgical innovation* 25, 3 (2018), 280–285.
- [14] Tobias Huber, Markus Paschold, Christian Hansen, Tom Wunderling, Hauke Lang, and Werner Kneist. 2017. New dimensions in surgical training: immersive virtual reality laparoscopic simulation exhilarates surgical staff. *Surgical endoscopy* 31, 11 (2017), 4472–4477.
- [15] Julie A Hunt, Matthew Heydenburg, Stacy L Anderson, and R Randall Thompson. 2020. Does virtual reality training improve veterinary students' first canine surgical performance? *Veterinary Record* 186, 17 (2020), 562–562.
- [16] Donald L. Kirkpatrick. 1998. The Four Levels of Evaluation. In *Evaluating Corporate Training: Models and Issues*, Stephen M. Brown and Constance J. Seidner (Eds.). Springer Netherlands, Dordrecht, 95–112. [https://doi.org/10.1007/978-94-011-4850-4\\_5](https://doi.org/10.1007/978-94-011-4850-4_5)
- [17] G Michael Lemole Jr, P Pat Banerjee, Cristian Luciano, Sergey Neckrysh, and Fady T Charbel. 2007. Virtual reality in neurosurgical education: part-task ventriculostomy simulation with dynamic visual and haptic feedback. *Neurosurgery* 61, 1 (2007), 142–149.
- [18] Kartik Logishetty, Wade T Gofton, Branavan Rudran, Paul E Beaulé, and Justin P Cobb. 2020. Fully immersive virtual reality for total hip arthroplasty: objective measurement of skills and transfer of visuospatial performance after a competency-based simulation curriculum. *JBJS* 102, 6 (2020), e27.
- [19] Kartik Logishetty, Branavan Rudran, and Justin P Cobb. 2019. Virtual reality training improves trainee performance in total hip arthroplasty: a randomized controlled trial. *The Bone & Joint Journal* 101, 12 (2019), 1585–1592.
- [20] Kartik Logishetty, Luke Western, Ruairidh Morgan, Farhad Iranpour, Justin P Cobb, and Edouard Auvinet. 2019. Can an augmented reality headset improve accuracy of acetabular cup orientation in simulated THA? A randomized trial. *Clinical Orthopaedics and Related Research* 477, 5 (2019), 1190–1199.
- [21] M David Merrill. 2002. First principles of instruction. *Educational technology research and development* 50, 3 (2002), 43–59.
- [22] Albert J Omlor, Leonie S Schwärzel, Moritz Bewarder, Markus Casper, Ellen Damm, Guy Danziger, Felix Mahfoud, Katharina Rentz, Urban Sester, Robert Bals, et al. 2022. Comparison of immersive and non-immersive virtual reality videos as substitute for in-hospital teaching during coronavirus lockdown: A survey with graduate medical students in Germany. *Medical Education Online* 27, 1 (2022), 2101417.
- [23] Brian Andrew Parsons, Natalie S Blencowe, Andrew D Hollowood, and Janet R Grant. 2011. Surgical training: the impact of changes in curriculum and experience. *Journal of surgical education* 68, 1 (2011), 44–51.
- [24] Francesco Porpiglia, Cristian Fiori, Enrico Checcucci, Daniele Amparore, and Riccardo Bertolo. 2018. Hyperaccuracy three-dimensional reconstruction is able to maximize the efficacy of selective clamping during robot-assisted partial nephrectomy for complex renal masses. *European urology* 74, 5 (2018), 651–660.
- [25] Yeshwanth Pulijala, Minhua Ma, Matthew Pears, David Peebles, and Ashraf Ayoub. 2018. Effectiveness of immersive virtual reality in surgical training—a randomized control trial. *Journal of Oral and Maxillofacial Surgery* 76, 5 (2018), 1065–1072.
- [26] Ziazi Radianti, Tim A Majchrzak, Jennifer Fromm, and Isabell Wohlgenannt. 2020. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education* 147 (2020), 103778.
- [27] Lauryn R Rochlen, Robert Levine, and Alan R Tait. 2017. First person point of view augmented reality for central line insertion training: A usability and feasibility study. *Simulation in healthcare: journal of the Society for Simulation in Healthcare* 12, 1 (2017), 57.
- [28] Richard M Satava, Alfred Cuschieri, and Jeffrey Hamdorf. 2003. Metrics for objective assessment. *Surgical endoscopy* 17, 2 (2003), 220.
- [29] Sophia F Shakur, Cristian J Luciano, Patrick Kania, Ben Z Roitberg, P Pat Banerjee, Konstantin V Slavin, Jeffrey Sorenson, Fady T Charbel, and Ali Alaraj. 2015. Usefulness of a virtual reality percutaneous trigeminal rhizotomy simulator in neurosurgical training. *Operative Neurosurgery* 11, 3 (2015), 420–425.
- [30] Sharaf Sheik-Ali, Hilary Edgcombe, and Chris Paton. 2019. Next-generation Virtual and Augmented Reality in Surgical Education: A Narrative Review. *Surgical Technology International* 35 (Nov. 2019), 27–35.
- [31] James Shin and Quynh A Truong. 2018. Manufacturing better outcomes in cardiovascular intervention: 3D printing in clinical practice today. *Current treatment options in cardiovascular medicine* 20, 12 (2018), 95.
- [32] Joseph D Shirk, Lorna Kwan, and Christopher Saigal. 2019. The use of 3-dimensional, virtual reality models for surgical planning of robotic partial nephrectomy. *Urology* 125 (2019), 92–97.
- [33] Joseph D Shirk, David D Thiel, Eric M Wallen, Jennifer M Linehan, Wesley M White, Ketan K Badani, and James R Porter. 2019. Effect of 3-dimensional virtual reality models for surgical planning of robotic-assisted partial nephrectomy on surgical outcomes: a randomized clinical trial. *JAMA network open* 2, 9 (2019), e1911598–e1911598.
- [34] Juan A. Sánchez-Margallo, Carlos Plaza de Miguel, Roberto A. Fernández Anzules, and Francisco M. Sánchez-Margallo. 2021. Application of Mixed Reality in Medical Training and Surgical Planning Focused on Minimally Invasive Surgery. *Frontiers in Virtual Reality* 2 (2021), 144. <https://www.frontiersin.org/articles/10.3389/frvir.2021.692641>
- [35] Matthias N van Oosterom, Henk G van der Poel, Nassir Navab, Cornelis JH van de Velde, and Fijs WB van Leeuwen. 2018. Computer-assisted surgery: virtual- and augmented-reality displays for navigation during urological interventions. *Current opinion in urology* 28, 2 (2018), 205–213.
- [36] Michael W Vannier and Jeffrey L Marsh. 1996. Three-dimensional imaging, surgical planning, and image-guided therapy. *Radiologic Clinics of North America* 34, 3 (1996), 545.
- [37] Nicole Wake, Marc A Bjurlin, Pooya Rostami, Hersh Chandarana, and William C Huang. 2018. Three-dimensional printing and augmented reality: enhanced precision for robotic assisted partial nephrectomy. *Urology* 116 (2018), 227–228.
- [38] Lianne M Wellens, Jene Meulstee, Cornelis P van de Ven, CEJ Terwisscha van Scheltinga, Annemieke S Littooi, Marry M van den Heuvel-Eibrink, Marta Fiocco, Anne C Rios, Thomas Maal, and Marc HWA Wijnen. 2019. Comparison of 3-dimensional and augmented reality kidney models with conventional imaging data in the preoperative assessment of children with wilms tumors. *JAMA network open* 2, 4 (2019), e192633–e192633.
- [39] S Yoganathan, DA Finch, E Parkin, and J Pollard. 2018. 360 virtual reality video for the acquisition of knot tying skills: a randomised controlled trial. *International Journal of Surgery* 54 (2018), 24–27.
- [40] M. Zhou, S. Tse, A. Derevianko, D. B. Jones, S. D. Schweitzberg, and C. G. L. Cao. 2012. Effect of haptic feedback in laparoscopic surgery skill acquisition. *Surgical Endoscopy* 26, 4 (April 2012), 1128–1134. <https://doi.org/10.1007/s00464-011-2011-8>

## A ARTICLE TABLES FOR SURGICAL LEARNING AND PLANNING

**Table 1: Research articles on learning. RCT: Randomized Controlled Study, VR: Virtual Reality, AR: Augmented Reality, VE: Virtual Environment**

Publication	Study Type and Kirkpatrick Level	HMD and Media Type	Participants	Primary Outcome	Conclusions	Negative Effects of HMD (when mentioned)
Harrington CM. et al. 2017 [11]	RCT. Level 2	VR 360° videos in Samsung gear VR headsets	Group A (Study) 20 undergraduate students Group B (Control) 20 undergraduate students	Presence of task-unrelated images or thoughts (TUIT) and subjective level of engagement at each probe	Significantly higher engagement level with the 360° video platform ( $p < 0.0001$ ) and lower TUIT ( $p < 0.0001$ ) compared with traditional 2D formats	
Logishetty K., Rudran B. et al. 2019 [19]	RCT. Level 2	VR VE with a VR simulator (HTC Vive System, Taipei, Taiwan; THA VR Simulation v1.1, Pixelmolkerei, Chur, Switzerland)	Group A (Study) 12 surgical trainees - postgraduate years 3 to 5 Group B (Control) 12 surgical trainees -same level	Technical and non-technical surgical performance measured by a THA-specific procedure based assessment (PBA)	In the cadaveric assessment, the median PBA global summary score for VR surgeons was Level 3b (THA performed competently without guidance or intervention but lacking fluency); Control surgeons achieved a median of Level 2a (guidance or senior intervention required for most or all the procedure) $p < 0.001$ , Mann-Whitney U test)	Mean level of nausea $0.7/10 \pm 1.6$ .
Logishetty K. Western L. et al., 2019 [20]	RCT. Level 2	AR VE with a HoloLens AR headset	Group A (Study) 12 medical students Group B (Control) 12 medical students	Post training improvement of the accuracy of acetabular component positioning with AR headset compared with hands-on training by an expert surgeon	There was no difference between participants in the surgeon-trained and AR-trained group: mean difference $1.2^\circ$ , 95% CI $[-1.8^\circ - 4.2^\circ]$ ; $p = 0.281$	
Yoganathan S. et al. 2018 [39]	RCT. Level 2	VR 360° videos	Group A (Study) 20 foundation year doctors Group B (Control) 20 foundation year doctors	Ability to tie a single handed reef knot on a knot-tying jig (Ethicon, UK) using a 4 mm diameter, 45 cm length of cord was assessed against a marking criteria	Using video teaching alone, knot tying scores were significantly better in the 360° VR video arm (median knot score 5.0 vs. 4.0 $p = 0.04$ ). When video teaching was used in combination with face to face teaching, the difference persisted (median knot score 9.5 vs. 9.0 $p = 0.01$ )	



Publication	Study Type and Kirkpatrick Level	HMD and Media Type	Participants	Primary Outcome	Conclusions	Negative Effects of HMD (when mentioned)
Logishetty K. et al. 2020 [18]	RCT. Level 2	VR VE	Group A 32 orthopaedic residents (surgical postgraduate years [PGY]-1 through 4) Group B 4 expert hip surgeons	Pre-simulation and post-simulation physical world assessments with 4 primary “visuospatial” outcome measures: error (difference between planned and achieved values) of femoral neck cut height (mm) and ( $^{\circ}$ ), and acetabular cup implant anteversion and inclination ( $^{\circ}$ )	Residents improved their physical world performance from a presimulation $6 \pm 3$ mm error to a postsimulation $4 \pm 2$ mm error in femoral osteotomy height ( $p = 0.044$ ), from a $12 \pm 6$ to an $8 \pm 4$ error in femoral osteotomy angle ( $p = 0.002$ ), from a $14 \pm 4$ to a $4 \pm 2$ error in acetabular cup inclination ( $p < 0.001$ ), and from a $16 \pm 6$ to a $6 \pm 3$ error in acetabular cup anteversion ( $p < 0.001$ )	
Pulijala Y. et al. 2018 [25]	RCT. Level 2	VR VE with an Oculus Rift VR headset and a Leap Motion controller	Group A (Study) 51 residents in Oral Maxillofacial Surgery Group B (Control) 44 residents in Oral Maxillofacial Surgery	The comparative evaluation scores of the perceived self-confidence levels before and after the intervention, measured using a 5-point Likert scale	The participants in the study group showed significantly greater self confidence scores compared with those in the control group ( $p = 0.034$ )	
Hunt JA. et al. 2020 [15]	RCT. Level 3	VR Stereo Videos with a Voxkin VR headset (Kathmandu, Nepal)	Group A (Study) 29 veterinary students Group B (Control) 15 veterinary students	Preparation time for surgery	There was no difference in the total time that VR group students and control group students spent in preparation for their first live surgery ( $p = 0.66$ )	
Huber T. et al. 2017 [14]	Non-randomized Controlled Comparative Study. Level 2	VR VE + $360^{\circ}$ Videos in 2016 HTC Vive $^{\circ}$ (High Tech Computer Corp., Taiwan) with two OLED-Displays (1080 $\times$ 1200 Pixel, 90 Hz, Low-Persistence)	10 members of the surgical department	Analysis of first experiences regarding the degree of immersion, motion sickness, and performance measurements	Participants were exhilarated by the immersion of the VR surrounding during the IVR session, had a high impression of “being there” and showed no sign of nausea. Laparoscopic performance in regular and IVR sessions did not differ when analyzed according to participants’ experience ( $p > 0.05$ )	The results of the motion sickness questionnaire revealed no sign of nausea for any of the participants at any time during the IVR session.

Publication	Study Type and Kirkpatrick Level	HMD and Media Type	Participants	Primary Outcome	Conclusions	Negative Effects of HMD (when mentioned)
Huber T. et al. 2018 [13]	Non-randomized Controlled Cross Over Study. Level 2	VR VE + 360° Videos with a LapSim simulator (Surgical Science, Gothenburg, Sweden) in combination with a 2016 HTC Vive (High Tech Computer Corp, Taoyuan, Taiwan)	30 members of the surgical department	Investigation of laparoscopic performance in both setups	Participants' virtual laparoscopic performance revealed no significant differences in all 3 tasks comparing AVR and IVR, regardless of the sequence of the 2 surroundings. No differences for time, handling economics, and mistakes were detected ( $p > 0.05$ )	Three female participants (10%) suffered from cyber sickness in AVR and IVR. In 2 participants, with a predisposition, this was higher in AVR compared with IVR (Motion Sickness Scale [P1; P2]: AVR [6-10; 8-12] vs. IVR [4-7; 4-9]). The other participant without a history of motion sickness had a higher cyber sickness in IVR (8-12) than in AVR (3-7). No other participants (n = 27) had cyber sickness in either VR surrounding

**Table 2: Research articles on planning. RCT: Randomized Controlled Study, VR: Virtual Reality, AR: Augmented Reality, VE: Virtual Environment**

Publication	Study Type and Kirkpatrick Level	HMD and Media Type	Participants	Primary Outcome	Conclusions	Negative Effects of HMD (when mentioned)
Shirk JD., et al. 2019 [33]	RCT. Level 4	VR with an off-the-shelf Google Cardboard-compatible VR headset and Phone 3D Virtual Reality model with a regular 3-D format from the smartphone	Group A (Study) 44 patients Group B (Control) 48 patients	Operative time	Operative time (OR, 1.00; 95% CI [0.37 – 2.70]) was not statistically different between groups in this model	
Shirk JD., Kwan L. et al. 2019 [32]	RCT. Level 4	VR 3D Virtual Reality model with a Google cardboard headset	Group A (Study) 15 patients Group B (Control) 15 patients	Operative time	After controlling for case complexity and surgeon, patients with VR-assisted surgical planning had significantly lower OR time (141 minutes vs. 201 minutes, $p < 0.0001$ )	
Wellens LM. et al. 2019 [38]	Well-conducted non-RCT. Level 1	AR 3D print and Augmented reality holograms with a mixed reality headset (HoloLens; Microsoft Corp)	10 patients	Scoring of the visibility of the 4 anatomical structures: tumor, arteries, veins, and urinary collecting structures from 1 to 5 (1 indicates completely disagree; 2, disagree; 3, neutral; 4, agree; and 5, completely agree)	Compared with conventional imaging, the 3-D print and the AR hologram models were evaluated by the surgeons to be superior for all anatomical structures: tumor (conventional vs. 3-D print $p = 0.008$ and AR hologram $p = 0.002$ ); arteries (conventional imaging vs. 3-D print $p = 0.002$ and AR hologram; ( $p < 0.001$ ), veins (conventional imaging vs 3-D print $p < 0.001$ and AR hologram $p < 0.001$ , and urinary collecting structures (conventional imaging vs. 3-D print $p < 0.001$ and AR hologram $p < 0.001$ )	