



HAL
open science

Immersive procedural training in virtual reality: A systematic literature review

Janine Jongbloed, Rawad Chaker, Elise Lavoué

► To cite this version:

Janine Jongbloed, Rawad Chaker, Elise Lavoué. Immersive procedural training in virtual reality: A systematic literature review. *Computers and Education*, 2024, 221, pp.105124. 10.1016/j.compedu.2024.105124 . hal-04674148

HAL Id: hal-04674148

<https://hal.science/hal-04674148v1>

Submitted on 21 Aug 2024

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.



Immersive procedural training in virtual reality: A systematic literature review

Janine Jongbloed^{a,*}, Rawad Chaker^a, Elise Lavoué^b

^a Laboratoire ECP, Université Lumière Lyon 2, 86 rue Pasteur, F-69365, Lyon, France

^b Université Jean Moulin Lyon 3, Iae Lyon School of Management, CNRS, INSA Lyon, UCBL, LIRIS, UMR5205, F-69008, Lyon, France

ARTICLE INFO

Keywords:

Augmented and virtual reality
Pedagogical issues
Evaluation methodologies
Simulations

ABSTRACT

Introduction: Virtual reality (VR) head-mounted displays (HMDs) offer an immersive learning experience that could replicate real-life training. Our review identifies and critically analyses studies of interventions designed to provide procedural training via these media and to test learning outcomes associated with these interventions.

Methods: We undertook a systematic search of the literature published between January 2013 and March 2023 related to immersive procedural training interventions using HMDs. Our search methods and protocol were guided by the PRISMA framework.

Results: Thirty-three studies fulfilled the broader inclusion criteria, while 23 fit the stricter criteria of measuring a procedural learning outcome. Studies were categorized into three broad types: design or usability studies, media comparisons, and value-added research. Learning environments were divided into 360° video representations, serious games, and simulations. Learning outcomes included knowledge gain, retention, and transfer. A subset of 16 study results that fit the criteria to be included in the meta-analysis show a significant positive medium effect size overall of immersive procedural training on learning outcomes as compared to less immersive environments. Knowledge transfer outcomes show the largest effect sizes.

Conclusions: Our systematic review of the literature highlights the richness and diversity of existing studies. We find strong support for immersive procedural training and suggest an important role for embodied cognition in maximizing learning outcomes. However, we also find a lack of consistency in intervention design, measured outcomes, and terminology among studies. In the future, more studies using experimental research designs on similar types of applications and outcome measures are needed to rigorously test these trends.

1. Introduction

The rapid expansion of Virtual Reality (VR) technology is revolutionizing professional training sectors by providing immersive and interactive learning experiences. VR-based training has been shown to be particularly relevant for procedural training, *i.e.*, skill acquisition, retention, and transfer, enhancing training effectiveness and efficiency (Chang & Hung, 2019; Rizzo & Bouchard, 2019). Procedural knowledge is the system containing memories of how one proceeds to do a series of tasks (Baroody, 2003; Gagne et al., 1997). Physical activities are often measured as the learning outcomes of procedural knowledge, as this knowledge is not expressed

* Corresponding author.

E-mail address: janine.jongbloed@univ-lyon2.fr (J. Jongbloed).

through verbal communication but by means of performance (Karaiskaki & Anastassiou-Hadjicharalambous, 2021). Procedural training is a cornerstone of safe practice, efficiency, and preparedness in a variety of domains, including medicine, emergency response, fire safety, laboratory work, manufacturing, and construction. Training procedural knowledge can lead to economic benefits, reduced accidents, faster response times, and other positive outcomes (Ricci et al., 2016). However, traditional classroom learning may not always be able to accurately replicate real-life situations, leading to less learner engagement and real-world skill transfer (Gao et al., 2019).

To address these limitations, VR has begun to be used in education and training to replicate dangerous or expensive workplace situations that might be impossible or unethical to let learners experience in real life. VR is well suited to learning procedural knowledge because it allows for repeated practice of complicated and challenging procedures in a safe and cost-effective environment (Hamilton et al., 2021), facilitating a risk-free, regulated setting for individuals to practice and master procedures prior to their real-world application (Xie et al., 2021). This makes VR particularly relevant for procedural tasks in fields such as the military and aviation, where pilots can replicate various flight scenarios (Buttussi et al., 2020), but also for emergency preparedness, and fire-fighting, for example (Morélot et al., 2021). In the healthcare sector, VR enables surgeons to simulate intricate surgical procedures (Lungu et al., 2021).

Furthermore, VR can provide immediate feedback and the opportunity for repeated practice until proficiency is achieved, thereby fostering a more effective learning process. The incorporation of VR into procedural training exemplifies the transformative impact of technology on conventional learning methodologies, rendering them more interactive, engaging, and efficient. As advances in VR technology continue, its utilization in procedural training is anticipated to become even more widespread, heralding a new era in professional training and development (Stefan et al., 2023).

Procedural training in VR allows learners to simulate gestures and body movements in interaction with the environment (Johnson et al., 2022; Numfu et al., 2020). The embodied cognition approach (Chemero, 2011; Corris & Chemero, 2022) considers that it is essential to integrate mental simulations, environments, situated actions, and bodily states in instructional design, as they play a pivotal role in human cognition (Barsalou, 2010). Embodied learning can be fostered through the enactment of real movement execution (Chaker et al., 2021) and hand gestures (Chaker, 2023). In this sense, immersive virtual learning using HMDs has been shown to enhance learner presence and engagement as compared to less immersive environments (Makransky et al., 2021).

However, based on the current literature, the extent to which these technological affordances lead to improved learner acquisition of procedural knowledge and skills in these environments remains unclear (Hamilton et al., 2021; Jensen & Konradsen, 2018; Johnson et al., 2022; Radianti et al., 2020). Our current literature review addresses this question, focusing on the different types of studies that have been conducted on immersive procedural training, the learning theories and measures of learning outcomes mobilized, and its efficacy as compared to traditional training environments.

1.1. Understanding learning in immersive virtual reality

Originally developed to understand less immersive multimedia learning, the Cognitive Theory of Multimedia Learning (CTML; Mayer, 2021) has been adapted and applied to understand learning in VR (Mayer et al., 2023; Parong, 2021; Parong & Mayer, 2018). Its rationale rests on the assertion that words and pictures allow people to learn more deeply than from words alone (Mayer, 2021). This theoretical approach assumes that learning occurs through active processing of selecting and organizing information as well as integration with existing information from prior learning (Mayer, 2021).

CTML has been specifically adapted into a theoretical approach for understanding learning in immersive VR (IVR) in the Cognitive Affective Model of Immersive Learning (CAMIL) (Makransky & Petersen, 2021). CAMIL illustrates how the technological affordances of immersion, control, and representational fidelity result in both a sense of presence and agency that in turn influences affective and cognitive factors that lead to learning in IVR (Petersen et al., 2022). Research using CTML and CAMIL highlight the unique immersive properties of IVR, which translate into an increased sense of presence (Meyer & Pfeiffer, 2020), and interactivity, which results in a sense of agency for the learner (Johnson-Glenberg, 2018). These two affordances have been highlighted as central to the important role that IVR could play in training (Johnson et al., 2022; Petersen et al., 2022) and have been argued to be particularly conducive to embodied learning (Lindgren & Johnson-Glenberg, 2013).

Theories of embodied cognition hold that cognitive processes are not isolated brain activities but involve the whole body and environment (Chemero, 2011; Clark, 1998) and are grounded in the sociomaterial and sensorimotor conditions of its emergence (Barsalou, 2010). Sensory and motor information from the body and environment are used to structure and control cognitive processes (Hollan et al., 2000; Versace et al., 2021), exploiting our environments to reduce cognitive workload (Wilson, 2002). Supporting this, the role of the environment in aiding recall is well-established (Godden & Baddeley, 1975; Murre, 2021). Embodied learning in IVR may occur when the learner uses “a physical gesture or movement that is well-mapped to the content to be learned” (Johnson-Glenberg, 2018, p. 1), which could provide a stronger memory trace for the content. This may offer particular advantages for non-declarative knowledge, where task performance is targeted, by engaging the physical body in the learning experience (Andrade et al., 2017).

1.2. Previous systematic reviews and meta-analyses

Early research on virtual learning and training environments showed promising effects on learning experiences, and suggested that their impact on learning outcomes needed further research (Mikropoulos & Natsis, 2011). However, most implementations before the widespread availability of current head-mounted displays (HMDs) used desktop-based VR pedagogical interventions (Merchant et al.,

2014). Since then, a growing number of literature reviews have focused on different types of XR (Kaplan et al., 2021; Papanastasiou et al., 2019), including HMD-IVR. Particularly in the last few years, several new reviews have been published on specific aspects of learning in HMD-IVR. For example, authors have focused in detail on the relationship between sense of presence and learning in IVR (Krassmann et al., 2019, 2023), student engagement in learning in IVR (J. Chen et al., 2023), and its integration into post-secondary curricula (Concannon et al., 2019).

Several literature reviews specifically relevant to IVR education and training have been published in recent years (see Table 1). Focused on procedural learning, a systematic review of serious games (SG) for evacuation training highlighted the opportunities that IVR presents for pedagogical design (Feng et al., 2018). Examining 15 studies, the authors highlighted the diversity of approaches used in terms of gaming environment, teaching methods, navigation, narrative, virtual agents, and hazards simulation. They showed generally positive results on knowledge acquisition and retention, as well as behavioural measures such as recorded in-game behaviour, but they did not provide a quantitative meta-analysis of learning outcomes.

In a more general review on the use of IVR in education and training, Jensen and Konradsen (2018) examined 21 studies incorporating HMDs from 2013 to 2018. Of these, six measured cognitive skill and five measured psychomotor skill acquisition. The results were mainly small positive or no effects of HMD as compared with traditional approaches. The authors highlight the limitations of IVR, where the immersive experience may overshadow learning and studies were often limited to user evaluations of the technology. Finally, and more positively, they underscore that while learners may be initially distracted by IVR experiences, they are also more engaged in the learning experience.

Another review focused on how IVR is used in higher education, and more specifically on the design elements and technologies, the learning theories, and the evaluation methods used to measure learning outcomes (Radianti et al., 2020). Important to our review, they classified learning contents into a number of categories, including “procedural-practical knowledge” in which learners must internalize procedures, “such as knowing how to perform a surgery or how to perform firefighting procedures” (Radianti et al., 2020, p. 13). They found that most studies did not mention any learning theories (70%) and only half specified a method for evaluating learning outcomes.

Addressing this point, another review focused specifically on quantitative learning outcomes in experimental designs where IVR was directly compared with a less immersive learning method (Hamilton et al., 2021). They showed that the greatest weakness of the studies in their review was the validity of the evaluation instruments, most of them using test scores, and particularly multiple-choice questionnaires (MCQ), to assess learning. The authors did not conduct a meta-analysis but do judge, from a qualitative perspective, that most studies in their review reported a significant advantage of using IVR over less immersive approaches. They particularly highlighted the encouraging results for procedural skills, but underscored the limitations of their own review, which only included four such studies and only two that looked at knowledge transfer to the real world.

A recent systematic review identified diverse approaches to learning in IVR and the technological and pedagogical design features used in IVR learning activities in educational contexts (Won et al., 2023). The authors outline different approaches taken by researchers when designing IVR-based learning and found five distinct patterns in the integration of design features, including one approach in which learners practice skills in realistic hybrid haptic environments (Won et al., 2023). However, they found overall limited integration and evaluation of pedagogical features, such as contextualisation and social interactions, into IVR-based learning due to technical challenges.

Two recent meta-analyses exploring the impact of IVR on general learning outcomes found a small positive effect size (ES) overall (Coban et al., 2022; Wu et al., 2020). The first, which analyzed 35 randomized controlled trials in the literature between 2013 and 2019, found a generally positive effect of IVR-HMDs on learning outcomes as compared to non-immersive desktop VR or lectures, with an overall ES of $g = 0.24$ (Wu et al., 2020). They examined several potential moderating variables, including learning domain, HMD hardware, control group treatment, and type of learning outcomes. They found a particular advantage for HMDs in K-12 education and in comparison with traditional lectures. The second meta-analysis based on 48 experimental studies from 2016 to 2020 resulted in a small positive overall ES of $g = 0.38$ (Coban et al., 2022). The authors suggested that more evidence is needed on procedural task-based

Table 1
Relevant systematic reviews and meta-analyses of the effect of IVR on learning outcomes.

Reference	IVR type	Topic	Learning outcomes	Conclusions
Feng et al. (2018)	HMDs & PBDs	Evacuation serious games (SG)	Best practices, self-protection skills, and spatial knowledge	Presents advantages and limitations
Jensen and Konradsen (2018)	HMDs	Education and training	Cognitive, psychomotor, and affective skills	Limited effectiveness, low study quality
Radianti et al. (2020)	HMDs	Higher education	Eight types, including procedural-practical knowledge	Learning theories not often considered; mainly usability not learning measures
Wu et al. (2020)	HMDs	Education and training	Knowledge test and task performance	Small positive effect size
Hamilton et al. (2021)	HMDs	Education and training	Cognitive, procedural, and affective skills	Evaluation of learning outcomes inadequate
Coban et al. (2022)	HMD, CAVE, and MR	Education and training	Knowledge and task-based activities	Small positive effect size
Won et al. (2023)	HMDs	Education and training	All types of learning tasks	Limited integration and evaluation of pedagogical features

performance.

These relevant existing reviews highlight that most studies do not approach VR with an educational theoretical perspective and tend to focus on aspects of user experience such as presence, ease of use, and enjoyment. Indeed, learning outcomes were rarely studied until relatively recently, and the large majority still focus on cognitive – or declarative knowledge – rather than on procedural knowledge and learning outcomes (Hamilton et al., 2021; Matovu et al., 2022). These existing recent literature reviews and meta-analyses highlight the need for a better understanding of the evaluation of learning outcomes after IVR training, as well as the need to focus on specific types of learning, notably procedural learning outcomes.

1.3. Rationale for the review

Our systematic review, which focuses specifically on IVR procedural training using HMDs, is guided by the findings of the reviews summarized above, which show encouraging results for learning outcomes on procedural tasks and suggest that virtual skill acquisition can be successfully transferred to real world tasks (Coban et al., 2022; Hamilton et al., 2021; Kaplan et al., 2021). Indeed, procedural knowledge has been highlighted as the least evaluated learning outcome in IVR applications (Matovu et al., 2022), and the measurement of procedural learning outcomes has not been discussed in detail in the reviews to date (McGowin et al., 2021). Furthermore, it was recommended that future systematic reviews investigate the effects of IVR on learning outcomes not only in relation to learning topics or educational levels, but also in relation to types of task designs and learning experiences (Won et al., 2023).

Building on and moving beyond these existing reviews, we study the extent to which immersive procedural training improves learner procedural knowledge as compared to traditional or less immersive environments. Our purpose in the current review can thus be summarized in three research questions.

1. What types of studies have been conducted on immersive procedural training and with what types of virtual environments?
2. What learning theories, objectives, and outcome measures were employed?
3. Does immersive procedural training improve learning outcomes as compared to traditional or less immersive virtual learning environments?

2. Methods

2.1. Search strategy

The search took place between January and March 2023. In line with previous systematic reviews and our knowledge of relevant recent publications, we chose to search six digital databases: the ACM Digital Library, EBSCO (including ERIC and PsycInfo), IEEE Digital Library, Science Direct, Scopus, and Web of Science databases. We used the PICOC method to choose our search keywords. This method defines search keywords using several categories: the population, or people with which evidence is collected; the intervention, or technology under study; the comparison, or control treatment conditions; the outcomes, or dependent variable of interest; and the context, or different environments in which the research takes place (Wohlin et al., 2012). We focused our search on virtual reality interventions in training, learning or education that examine outcomes explicitly related to procedural knowledge. As described in the introduction, a variety of terms are commonly used in the multi-disciplinary research on procedural training in IVR, so we included synonyms and related terms as necessary to include a larger number of relevant studies. Our resulting base search string was:

("educat*" OR "learn*" OR "train*") AND ("head mounted display" OR "HMD" OR "head-mounted display" OR "virtual environment" OR "virtual learning environment" OR "virtual training environment" OR "virtual reality" OR "VR") AND ("behavio* skill*" OR "procedural knowledge" OR "procedural acquisition" OR "procedural training")

Our inclusion and exclusion criteria are presented in Table 2. There are at least three different systems that are often used in VR research: desktop VR, using 3D worlds or games for example; head-mounted displays (HMDs) ranging from high definition and field of vision Varjo VR to budget Google Cardboard glasses using a regular smartphone; and projection-based displays (PBD), such as CAVE systems that display 3D images on multiple screens on multiple walls of a room. We focus on HMDs because they are considered the

Table 2
Inclusion and exclusion criteria.

Inclusion criteria (IC)	Exclusion criteria (EC)
IC-1 The full text of the study is available.	EC-1 The paper is less than 6 pages in length.
IC-2 The study is peer reviewed.	EC-2 The study does not examine procedural knowledge.
IC-3 The study is written in English.	EC-3 The study is a secondary study, such as a review or theoretical paper.
IC-4 The study measures some kind of learning outcome or user-evaluated presence/immersion.	EC-4 The study presents a desktop rather than an immersive VR intervention.
IC-5 The study presents some type of evaluation of the effectiveness of the VR intervention.	EC-5 The study presents a medical simulator without an HMD.
IC-6 The study uses an HMD in at least one of the experimental conditions.	EC-6 The subjects are a special subpopulation of adults (e.g., illness, disability, etc.)
IC-7 The study was published between 2013 and 2023.	EC-7 The paper does not present an empirical study (meta-analysis only).

most immersive because they separate the user from the external world completely from a visual and auditory standpoint. Therefore, we decided to only include those studies that used an IVR environment, either computer-generated or 360° video, via an HMD. Those studies that used only a desktop VR environment, or a surgical or dental trainer without an HMD, were excluded.

Our focus on learning outcomes also led us to only include those studies who explore a measure of procedural knowledge. Procedural knowledge most often refers to the knowledge of how to perform a specific task, such as operating an instrument (Dong et al., 2022), that requires the learner to perform a sequence of actions in order (Ganier et al., 2014). During procedural training, the steps are typically demonstrated and replicated as practice, during which learners focus on assimilating both procedural knowledge and mastering motor skills (Buttussi & Chittaro, 2021). This type of knowledge, consisting of “knowing the steps required to perform a particular procedure” (Ganier et al., 2014, p. 829), is often a vitally important learning objective in a variety of contexts, from assembly lines to surgery to firefighting.

We did not limit our review to empirical studies in our systematic review, as we were also interested in how design, feasibility, and usability studies define procedural knowledge for the purposes of creating IVR applications. However, only experimental results were included in the meta-analysis. We ensured the maturity of the research reviewed by restricting our search to peer-reviewed articles and conference papers of at least 6 pages in length that were published in a scientific journal or a scientific conference proceeding.

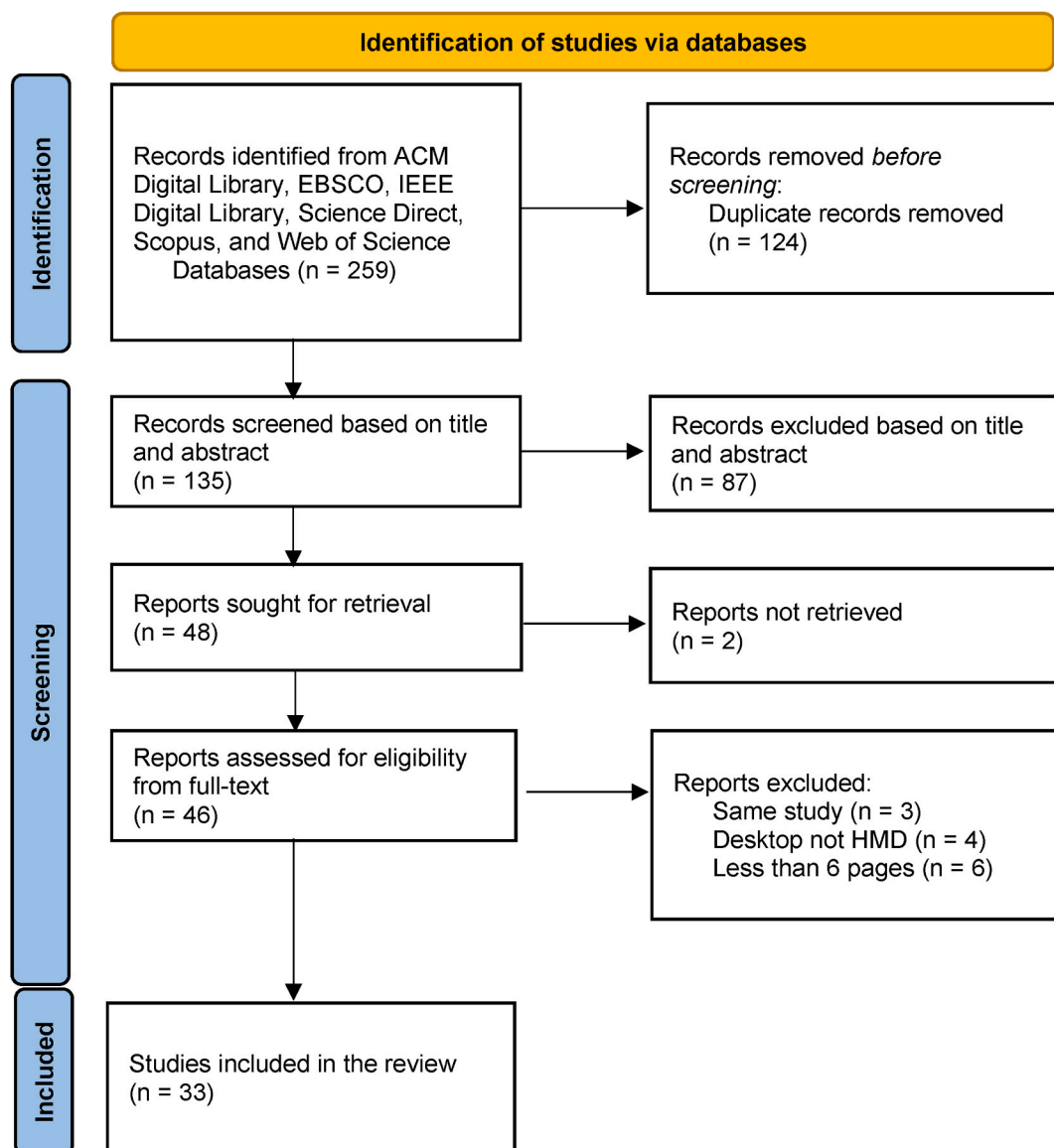


Fig. 1. PRISMA flow diagram (Page et al., 2021).

2.2. Selection and quality assessment

After conducting our searches in the relevant databases using the article metadata, including the titles, abstracts, and keywords, we exported BibTeX files. We then used the Parsifal online tool (<https://parsif.al>) to organize the resulting BibTeX records and remove duplicate publications. From our databases searches, we identified 259 relevant records (see Fig. 1). Of these, 124 were duplicate records. We manually screened the remaining 135 records using their meta-data (title, abstract, keywords, etc.) to judge whether they met the inclusion and exclusion criteria. Eighty-seven records were removed at this stage, mainly because they examined desktop VR rather than IVR with an HMD. We then analyzed the full text of the remaining publications. Of these 46 publications, 36 studies were judged eligible for the present review. Studies were mainly excluded because they were less than six pages long (six studies), but also because they used desktop rather than HMD-IVR (four studies). Three of the articles presented the same IVR application as another study found in the review (see Fig. 1). We chose to include only the more complete, preferably experimental study, and so three studies were excluded from the systematic review of the IVR applications (Li et al., 2020, pp. 566–572; Taunk et al., 2021; Çakiroğlu & Gökoğlu, 2019b). However, the empirical results from these studies on separate samples of learners were considered for inclusion in the meta analysis, as was done for those articles presenting two separate experiments on different samples within the context of the same IVR study publication (De Lorenzis et al., 2023; Makransky et al., 2021).

2.3. Meta-analysis

For the meta-analysis, we re-examined the criteria for each study included in the systematic review. In order to reasonably compare results across studies, we limited our selection to interventions that report similar types of evaluations. All experiments that included a quantitative measure of a learning outcome for both an experimental and control group within one of the selected studies were initially chosen and examined in more detail. Due to the limited number of value-added studies (see below), only media comparison studies were included in the meta-analysis. Although several studies used multiple media comparison groups or media comparison combined with value-added conditions (e.g., Buttussi & Chittaro, 2021; Jung & Ahn, 2018), we chose to only select the control group most similar to a low-immersion, desktop condition. This decision was based on the risk of overrepresenting the same learning outcome result for the same IVR experimental group multiple times if introduced into the meta-analysis for each comparison group. However, if the same study presented multiple experiments meeting the criteria, we included these separate results. These restrictions resulted in the inclusion of eleven studies in the meta-analysis, with sixteen individual experimental results.

Effect sizes were computed from the descriptive statistics (means, standard deviations, and sample sizes) reported in the studies. We use Hedges' g , which is an adjustment to Cohen's d to correct for the fact that Cohen's d tends to overestimate the effect size in small samples (<20) because many of our studies have relatively small sample sizes. Cohen's d is defined as the difference between the mean of the experimental group and the control group divided by the pooled standard deviation. We use the "meta" command in Stata (StataCorp, 2021), using Hedges' g and analytic weighting, and present the forest plot to compare effect sizes across studies. We also conducted asymmetry tests for publication bias.

3. Results

3.1. Domains and training environments

The studies came from a variety of domains (see Table A1 in the Appendix), including medicine (twelve studies), emergency first response and rescue (five studies), industrial operations and maintenance (three studies), aviation safety, biology, construction, engineering, fire safety (two studies each), mathematics, public health, and road safety (one study each). The procedural skills to be learned were also diverse, including surgical procedures and skills, airplane evacuation, laboratory equipment training, machinery operation, first response and rescue operations, vehicle detailing and maintenance, and road safety, to mention a few. The vast majority incorporated training situations that are dangerous or expensive to replicate in real life, such as confined space rescue or cadaveric surgery.

The virtual training environments included 360° video representations, serious games, and simulations, as was also reported in previous reviews (Merchant et al., 2014; Wu et al., 2020). Importantly, only the latter two training environments incorporate interactivity, highlighted in the introduction as a key technological affordance of IVR. The majority of the IVR learning experiences (22 studies) were simulations. These simulations use learning environments that recreate real-life situations, as is often done in aviation, medical, and military training (Gormley & Murphy, 2023). Indeed, simulation training in medicine makes up the largest common group of procedural training interventions found in our systematic search of the literature. Seven used a serious game, or a game designed with an educational purpose (Buttussi & Chittaro, 2018; Çakiroğlu & Gökoğlu, 2019b; S.-Y. Chen & Chien, 2022; Longo et al., 2023; Lu et al., 2022; Luo et al., 2021; Szczerowski & Smith, 2018). Serious games typically incorporate a game plot with levels and missions that must be accomplished and corresponding points to reward performance (Lavoue et al., 2019). Finally, four studies used 360° video with corresponding narration and information points for the IVR learning experience (Grab et al., 2023; Kushiro et al., 2021; Pieterse et al., 2023; Queisner et al., 2022).

3.2. Types of study design

The reviewed papers include design and usability studies, which describe and evaluate an application through preliminary testing

with target user groups; media comparisons, which test the influence of a medium on learning or another outcome as compared to a different medium; and value-added studies, which measure the added benefit of adding an attribute or activity to a given medium as compared to the medium without this addition (see Fig. 2). Our primary focus is media comparisons, as we are interested in the effects of procedural training in IVR on learning outcomes as compared to traditional or less immersive environments. However, our educational lens makes value-added studies that focus on specific pedagogical methods also of particular interest (Meyer et al., 2019).

The experimental designs of these studies vary greatly. Some studies compared IVR with traditional paper-based instruction, e-learning, a smartphone application, simulator training, or a less immersive VR, such as a Google Cardboard low-cost HMD (see Table A2 in the Appendix). Also, some used pre- and post-tests with control group design (PPC), while others used a single-group pre- and post-tests design (SGPP) or only a post-test with control group design in an expert-novice comparison.

Design or usability studies made up the majority of the studies selected initially for review (ten studies, see Fig. 2). For example, Dong et al. (2022) describe the development of an IVR application designed to teach students how to use cryoEM laboratory instruments. They test their CryoVR application with experts to see to what extent it is realistic and instills a sense of presence. Emond et al. (2022) present an IVR application to train emergency first responders in remote areas, describing the iterative design of the application with a technical advisory group of subject matter experts.

Several medical applications involved usability studies of 360° video visits of hospitals or operating rooms to give residents an immersive experience with situated and contextualized information before their first real-life work experiences (Pieterse et al., 2023; Queisner et al., 2022; Snarby et al., 2019). Other medical applications illustrated healthcare procedures in an interactive way and got user feedback on their perceptions of enhanced readiness, mental load, and motivation through a number of questionnaires (Ganni et al., 2020; Tang et al., 2021). Two studies got user or expert feedback verbally (Luo et al., 2021; Szczurowski & Smith, 2018).

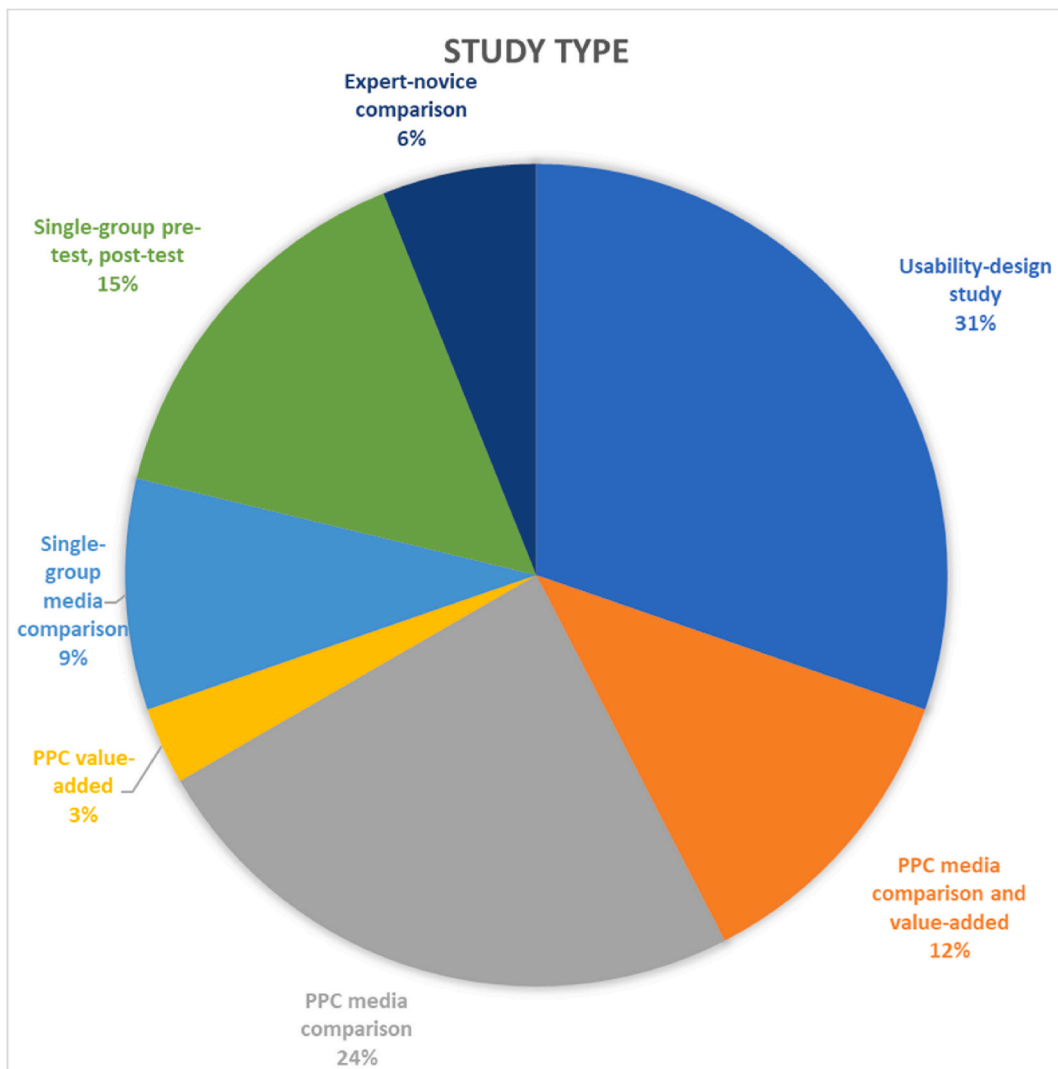


Fig. 2. Types of studies included in the review.

Several of these studies conducted an expert-novice comparison either as part of the usability study or as a single-group quasi-experimental study. For example, [Kushiro et al. \(2021\)](#) compared the hazard utterances of expert and novice field overseers immersed in a 360° video of a construction site. [Longo et al. \(2023\)](#) compared the fuzzy cognitive maps of expert and novice electric crane operators on a virtual oil rig. [Pérez-Escamirosa et al. \(2022\)](#) tested a surgical skills application on experts and novices, comparing their mental load and perceptions of learning.

Media comparison studies comparing IVR with a control group that experienced a less immersive environment were particularly pertinent to the present review because we focus specifically on the effects of immersive procedural training on learning outcomes. There were eight such studies, comprising 24% of our selected publications (see [Fig. 2](#)). For example, [Lu et al. \(2020, pp. 66–73\)](#) present an IVR application designed to provide confined space rescue training, and compare it with a traditional e-lecture training condition, measuring both knowledge gain and transfer using a quiz and a role-play in a real context with professional confined space rescuers. [Buttussi and Chittaro \(2021\)](#) compare three media in their study focused on aviation safety and emergency exit procedures: IVR-HMD, smartphone application, and printed materials. [Lohre et al. \(2020\)](#) present a surgical application, comparing IVR with video training. Similarly, [Logishetty et al. \(2019\)](#) compare an IVR application for surgical training with traditional preparatory materials, and test learning transfer with a surgical assessment of cadaveric surgery.

There were also a number of single-group studies, either using a pre-test post-test design (SGPP) or conducting a media comparison within one group (exposure to one condition and then the other after a wash-out period). There were five SGPP studies, including two focused on fire safety training, one with children ([Çakiroğlu & Gökoğlu, 2019a](#)) and one with college students (S.-Y. [Chen & Chien, 2022](#)). Each measured learning outcomes based on a realistic role-play situation after training. Two others examined industrial or professional applications, testing knowledge gains within the IVR environment itself ([Longo et al., 2019](#); [Tai et al., 2022](#)). A final example looked at functional communication training in medicine with continued IVR training until mastery criterion were met ([Clay et al., 2021](#)).

There were three single-group media comparison studies. One compared AR and VR with random assignment of the order of the test conditions ([Daling et al., 2020](#)). Another incorporated a two-week wash-out period between conditions ([Katz et al., 2020](#)). The third combined desktop and IVR conditions for all participants along with an expert-novice comparison of performance in the two different conditions ([Dhalmahapatra et al., 2021](#)). These applications were in the domains of industrial manufacturing assembly, advanced cardiac life support, and electric overhead crane operations, respectively. The first two focused on mental load, while the latter focused on presence and system usability.

Two of the articles presented more than one study. [De Lorenzis et al. \(2023\)](#) conducted both a media comparison and value-added research with their IVR application focused on energy audits in engineering training. In the media comparison, they compare the application with traditional lab training. In the value-added study, they integrate a learning-by-teaching approach in both media conditions to test its effect on knowledge gain and retention. [Makransky et al. \(2021\)](#) likewise include both a media comparison and a value-added study. They compare IVR with a training video focused on forensic analysis of DNA in crime scene investigation. The value-added study integrates an enactment phase outside of the IVR-HMD, where learners manipulated instruments used in the procedure. Both of these pedagogical methods were based on generative learning theory (GLS).

[Jung and Ahn \(2018\)](#) and [Schroeder et al. \(2017\)](#) combined a media comparison and value-added research in the same experiment in their studies. [Jung and Ahn \(2018\)](#) look at maritime safety and lifeboat launch procedures, testing a desktop condition and two IVR-HMD conditions: one with a computer joystick and the other with wearable sensors. [Schroeder et al. \(2017\)](#) looked at a military application training users to change an alternator in a military vehicle using a desktop condition and two IVR conditions, one gesture-based and one voice-based.

The only study that looked at a *value-added* comparison alone was the experiment by [Buttussi et al. \(2020\)](#) that compared learning of cardiopulmonary resuscitation (CPR) in IVR both using a physical mannequin and without, to test the added benefit of incorporating haptic or tactile feedback in training.

3.3. Learning theories and objectives

Learning theories were explicitly evoked in only half (16) of the 33 studies (see [Table A1](#) in the Appendix). This lack of explicit guiding principles has been highlighted by other researchers ([Chan et al., 2021](#); [Radianti et al., 2020](#)). General learning theories presented as framing for the pedagogical design of the IVR applications included a behaviorist approach ([Çakiroğlu & Gökoğlu, 2019a](#); S.-Y. [Chen & Chien, 2022](#); [Clay et al., 2021](#); [King et al., 2022](#)), an experiential learning or ‘learning by doing’ approach ([Dong et al., 2022](#); [Emond et al., 2022](#); [Longo et al., 2019, 2023](#); [Lu et al., 2022](#); [Tang et al., 2021](#)), an embodied learning approach ([Dong et al., 2022](#); [Makransky et al., 2021](#); [Queisner et al., 2022](#); [Tai et al., 2022](#)), a constructivist approach ([Buttussi et al., 2020](#)), and a situated learning approach ([Luo et al., 2021](#)). Only two referred directly to the IVR learning theories of CTML or CAMIL as guiding their pedagogical choices ([Makransky et al., 2021](#); [Tai et al., 2022](#)). These broad theoretical frameworks typically justified the use of IVR as a medium, rather than providing support for specific design choices.

Several studies directly referred to embodied cognition and embodied learning as driving the design of learning experiences and evoked the role of embodied learning in optimizing the process of learning procedural knowledge ([Dong et al., 2022](#); [Makransky et al., 2021](#); [Queisner et al., 2022](#); [Tai et al., 2022](#)). For example, [Makransky et al. \(2021\)](#) argue for the inclusion of physical manipulation on the basis that user-controlled, meaningful physical activity increases learning. They justify this choice using previous studies showing that direct bodily manipulation and motor actions congruent with learning concepts enhance learning ([Jang et al., 2017](#)). They find support for the importance of physical manipulation—or enactment—on learning outcomes, and particularly in the IVR condition ([Makransky et al., 2021](#)). [Dong et al. \(2022\)](#) also attribute much of the benefit of IVR to the facilitation of creating ‘learning by doing’

experiences where the physical body is engaged in actions that are congruent with the knowledge to be learned. They support their argument with studies of language learning (Macedonia et al., 2011; Repetto et al., 2015) showing that gestures may help learning by changing the neural pathways used to store and retrieve information. In contrast, Queisner et al. (2022) focus on embodied learning cues in the form of realistic situational cues in the virtual environment and highlight the importance of the feeling of being physically present in the real situation for deep learning.

Learning objectives and targeted competencies were often not explicitly described in the articles, although performance indicators were frequently justified as mapping onto the learning or training objectives, and thus can be inferred. More broadly, learning objectives centred on procedural training, knowledge, or skills. However, only ten studies explicitly defined this term in a conceptual way for the purposes of defining the learning objectives of the IVR application. Buttussi and Chittaro (2018) define procedural training as teaching learners to memorize complex procedures, to analyze the contextual factors that affect their execution, and to update their performance accordingly. Tang et al. (2021) describe procedural knowledge in the context of practical training, when learning standardized procedures with many steps, which demand students to repetitively revise and practice. Likewise, King et al. (2022) define it as the acquisition of steps in a mathematical questioning strategy.

Jung and Ahn (2018) divide procedural skills into two parts that both concern the execution of action sequences, including technical (sensory motor) skills, such as the manipulation of equipment, and procedural knowledge of the correct sequence of steps required for the task. Likewise, Longo et al. (2023) focus on procedural expertise on non-routine tasks, requiring “context-awareness (knowledge about the current system state), technical knowledge (capability to execute the task efficiently and safely) and decision-making skills (ability to understand which is the best task to execute to achieve the desired system state)” (p. 4). Dong et al. (2022) focus on the knowledge of how to operate an instrument, while Makransky et al. (2021) describe it as the knowledge associate with a physical skill of conducting a task.

Some studies defined more specific learning objectives. For example, in their First Responders training system, Emond et al. (2022) detail four high-level training objectives and three specific learning goals that translate into actionable steps in the IVR environment. In a training application for specialized biology laboratory equipment, Dong et al. (2022) detail three learning objectives: gaining familiarity with the sequence of steps required to use the equipment, reducing time needed training on real equipment, and showing conceptual understanding of key errors and how to avoid or respond to them. These learning objectives can then be used to define learning outcome measures.

3.4. Learning outcome measures

While most of the studies do not explicitly define procedural knowledge conceptually, they do define it implicitly in their measurement of the learning outcomes. For example, procedural knowledge is measured as verbally recounting the steps in a procedure (Johnson et al., 2022), performing or role-playing the situation (Lu et al., 2022), or describing how to apply the steps in a procedure to a new situation (Makransky et al., 2021). These measures are based on the reporting or performance of the correct actions or tasks (in the correct order) necessary to complete the targeted procedure. This may be quantified as the number of errors committed, the accuracy of manipulations, or the time taken to successfully complete the task, for example (Mao et al., 2021). Procedural knowledge was sometimes measured in a verbal or declarative manner, necessitating a verbalization of the tasks or procedures that were performed in the immersive procedural training. However, in both cases they measure learners’ understanding of the practical methods, techniques, and processes that should be applied (Matovu et al., 2022).

The most common evaluation method for these learning outcomes was oral or written knowledge tests in the form of interviews or quizzes (eleven studies). These tests typically asked the participant to report the steps in a procedure. In contrast, some studies scored task performance either in an evaluation mode in IVR (seven studies) or on real-life tasks or role-plays (eight studies). For example, participants in the fire safety studies role-played a real fire situation with professional firefighters (S.-Y. Chen & Chien, 2022). Several studies combined more than one type of learning outcome measure (four studies), using both tests and role-plays. Most studies provided multiple measures of a learning outcome of a procedure, such as the number of correct steps, incorrect steps, and out-of-order steps (Buttussi & Chittaro, 2021). The time it took to complete the task was also a common measure in evaluation modes and role-plays, as well as the quality of the task performance on an accepted evaluation scale, particularly for surgical applications (Logishetty et al., 2019).

These different measures can be defined and categorized into measures of knowledge gain, retention, and transfer (see Table A1 and A2 in the Appendix). Knowledge gain (KG) is typically measured as a change score from a pre-test to a post-test immediately after training. Nineteen studies focused at least one of their learning outcomes on KG. Knowledge retention (KR) is a post-test score after a variable time-period delay, for example, after two weeks, to see to what extent learning has successfully been incorporated into long-term memory. Five studies looked at KR. Finally, knowledge transfer (KT) is the “process by which knowledge acquired in a particular context is retrieved and applied in another” (Ganier et al., 2014, p. 828). Eight studies measured KT learning outcomes. It is typically considered a very strong indicator of the quality of learning. This is a key measure of learning outcomes in immersive procedural training, as the ultimate goal is not for learners to perform well in IVR or on a test, but rather to use these skills in real-world situations. KT was most often measured for surgical applications, where performance on a simulator or cadaveric surgery after IVR training was common (Shah et al., 2022).

3.5. Meta-analysis of learning outcomes

In order to test our research question concerning the effect of immersive procedural training on learning outcomes, we further

limited our selection of studies to those who report quantitative results in a pre-test post-test with control group (PPC) experimental design. Thirteen publications present sixteen individual studies that provide quantitative results from a between-subjects design with a control group on measures of knowledge gain, retention, or transfer outside of the IVR application. Three studies presented results for both KG and KR, and so were included for each subset. Several studies did not present the descriptive statistics necessary for a meta-analysis. In these cases, the authors of the studies were contacted via email to request the information needed. Two studies were excluded from the analysis because we could not access complete descriptive statistics for the relevant outcomes in the publication.

First, we compare results on KG, KR, or KT from media comparison studies only (see Table A2 in the Appendix for the studies included). We compare IVR treatment groups with the type of control group that most closely approximates a regular desktop learning environment, which was the most common control group, from each relevant study. In doing so, we necessarily mix types of control groups in the same meta-analysis, including video, smartphone, desktop VR, and regular lab teaching comparison groups. In the sixteen sets of results included in the meta-analysis, we find a positive overall effect size (theta) when comparing IVR with less immersive learning conditions: Hedge's $g = 0.58$, indicating a medium positive effect of IVR as compared to controls on learning outcomes (see Fig. 3).

The Q test of homogeneity indicates that not all the studies have a common effect size. The measures of heterogeneity, tau squared and H^2 , show that there is heterogeneity in the data. The I^2 statistic presents the proportion of the observed variability in effect sizes that captures the true score variance, or real differences in effect size. We see that this is high, at 89%. For this reason, asymmetry and small-study tests of publication bias are likely not appropriate for use in this meta-analysis (Ioannidis & Trikalinos, 2007). While we find that several of these tests were significant for asymmetry, we cannot interpret these results with certainty because an asymmetrical funnel plot may be due to study heterogeneity rather than publication bias (Ioannidis & Trikalinos, 2007).

To begin investigating this heterogeneity, we examined several different potential moderating variables. A subgroup analysis by type of learning measure (knowledge gain [KG], knowledge retention [KR], or knowledge transfer [KT]) shows significant differences in theta between groups, with a significant test of group differences ($\chi^2 = 8.75, p < 0.05$). Learning outcomes are most positive for knowledge transfer measures, with a large ES of Hedge's $g = 1.80$ (see Fig. 4). They are not significant for knowledge gain or knowledge retention. However, caution is warranted, as the knowledge transfer group includes only four studies. Indeed, a notable constraint in our exploration of moderating variables was the small number of studies eligible for inclusion in the meta-analysis.

4. Discussion and conclusions

In our systematic review and meta-analysis, we examined immersive procedural training interventions using HMD-IVR. We described study domains and virtual environments, learning theories, learning outcomes, and evaluation methods. We also conducted a meta-analysis on whether immersive procedural training improves learners' procedural knowledge as compared to traditional

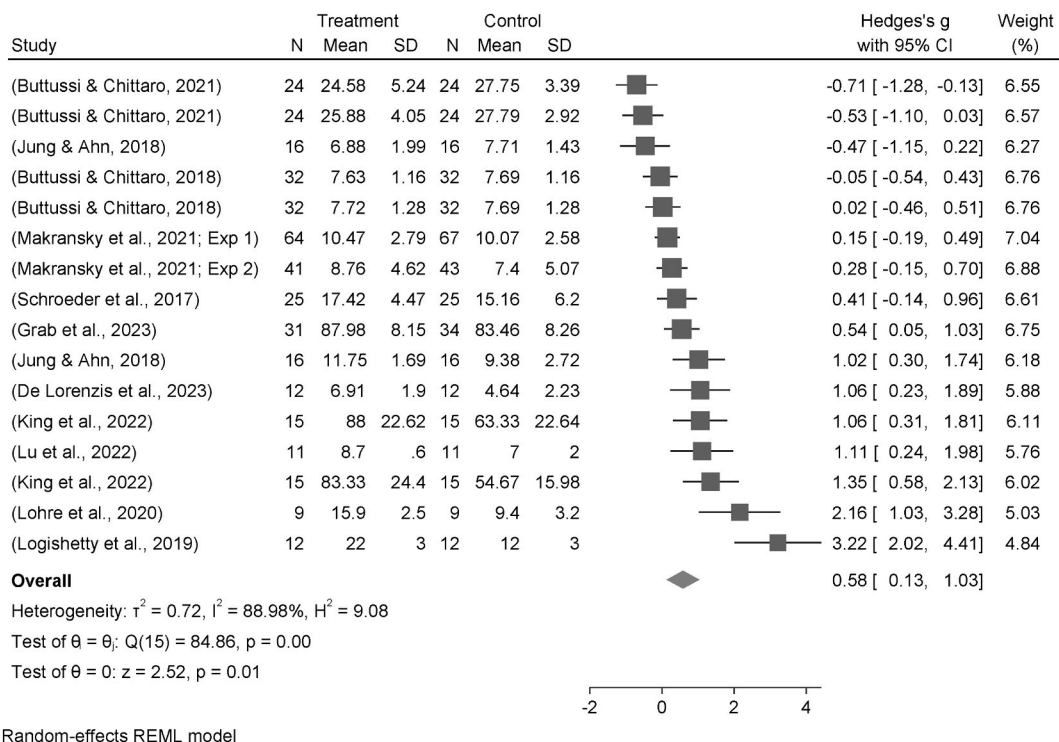


Fig. 3. Meta-analysis forest plot using Hedge's g effect size.

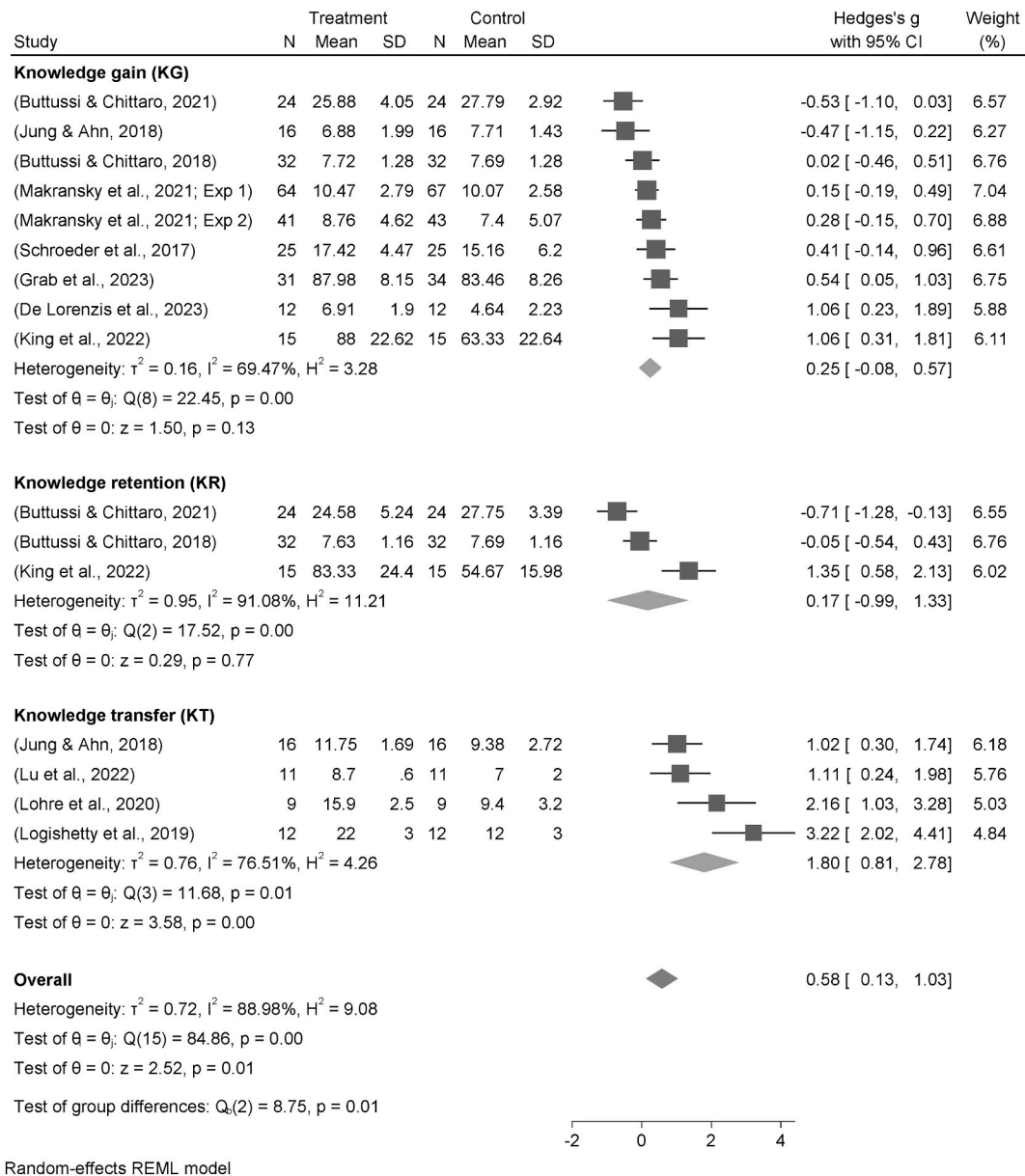


Fig. 4. Meta-analysis forest plot by type of procedural learning outcome.

environments. Most of the studies were design or usability studies rather than experimental studies measuring and comparing learning outcomes. Of those which were experimental studies, the majority were media comparisons between HMD-IVR conditions and less immersive technologies. We found that common domain was that of medicine, and most training environments were simulations of real-life situations where procedural knowledge was to be applied, such as a surgery, assembly of parts, or worksite tasks.

4.1. Pedagogical design of IVR training

In our introduction, we highlighted that IVR has been shown to be an effective medium for training, offering opportunities for learners to engage in generative processing. However, it also potentially distracts learners by necessitating extraneous processing of large amounts of information (Mayer et al., 2023). One problem underscored in the research was that studies tend to focus on technological features rather than pedagogical features (Won et al., 2023) and that user experience was studied more often than learning outcomes (Wu et al., 2020). This is problematic, as there is some evidence that IVR influences perceptions of learning more than actual learning performance (Makransky et al., 2019). Furthermore, studies examining learning transfer as opposed to evaluation modes incorporated directly within the IVR training environment have highlighted the need to avoid “getting good at the game” effects

on learning outcomes (Parong & Mayer, 2018). This occurs when learners improve their skills at interacting with the IVR training environment rather than their procedural knowledge, which may inflate learning outcomes.

Prior reviews emphasized that a variety of pedagogical approaches have been employed in IVR (Johnston et al., 2018), each relying on different theoretical approaches, but several authors found that constructivist approaches are most common (Concannon et al., 2019). Our study supports this conclusion. However, learning theories were most often not overtly discussed in the studies, as was also found in other previous reviews (Hamilton et al., 2021; Kaplan et al., 2021), and only half of the studies framed their IVR application using a cited learning theory.

Some important exceptions include the studies conducted by De Lorenzis et al. (2023) and Makransky et al. (2021), which implement value-added research based specifically on pedagogical interventions. Both use generative learning strategies (GLS), as defined by the Cognitive Theory of Multimedia Learning (Fiorella & Mayer, 2021; Mayer, 2009). According to Fiorella and Mayer (2016), “learning by enacting is [...] consistent with generative learning theory, as it helps learners use their prior knowledge to connect abstract concepts to concrete objects and actions, thereby enabling them to construct a more meaningful mental representation of the material” (p. 730). The first tested a “learning-by-teaching” approach combined with IVR, while the second used enactment with physical props. Both found particular benefits for IVR combined with these GLS, showing that the inclusion of theoretically driven design elements can maximize technological affordances. These types of approaches offer important advances in the field, which are needed to guide application development and best practices (Jensen & Konradsen, 2018; Radiani et al., 2020).

Embodied learning theory deserves special attention in this respect. One study demonstrated empirically that the inclusion of enactment is beneficial to make full use of the technological affordance of interactions in the IVR environment and to improve learning outcomes (Makransky et al., 2021). These findings are also supported by the medical simulation studies using HMDs, which incorporated interactivity as physical gestures, even though not described explicitly as embodied learning, and reported some of the largest effect sizes. As argued in the literature, IVR hand controllers allowing natural movement that is both meaningful and congruent with the content to be learned may provide important learning benefits (Johnson-Glenberg, 2018). This advantage is likely even larger for training in procedural tasks, where movements are vital components of the tasks to be learned. Thus, the effects of IVR on learning may be enhanced by incorporating theoretically based pedagogical elements that reinforce the technological affordances of the medium.

4.2. Efficacy of immersive procedural training

We analyzed sixteen experimental results in our meta-analysis and found a significant positive effect of immersive procedural training as compared to less immersive technologies. This effect was particularly large for those studies that looked at knowledge transfer. These findings support those of existing meta-analyses and underscore the potential of IVR to enhance learning outcomes as compared to less immersive media. For example, when looking at all types of learning outcomes more generally, Wu et al. (2020) found a small positive ES of $g = 0.24$ and Coban et al. (2022) found a small positive ES of $g = 0.38$. Our larger ES of $g = 0.58$ confirms the hypothesis posed in several previous reviews that procedural skills respond particularly well to this type of training (Coban et al., 2022; Hamilton et al., 2021). Our encouraging results for procedural skills specifically support these suggestions that had yet to be established empirically. However, due to divergences in study design, pedagogical interventions, and learning outcome measures, some caution is warranted when generalizing these findings to other IVR applications.

4.3. Limitations and recommendations for future research

It is not easy to untangle the separate effects of media and method in our systematic review and meta-analysis. Indeed, the pedagogical methods used in the IVR applications are difficult to uncover from the often-limited descriptions in the articles, and likely differ strongly from one IVR application to the next. While several example images and a general description were often included, most did not explain the plot of the simulation or game in detail, although there were some important exceptions (for example, Buttussi & Chittaro, 2018). This heterogeneity was evident also in the meta-analytic results, where we were only able to use a subset of the studies, as several studies did not report empirical results including a control group or did not report complete descriptive statistics corresponding to their results. Furthermore, a number of different types of technological alternatives were included in the control groups, which prevented us from employing asymmetry tests for publication bias. Our results thus comprise a diversity of technological and pedagogical approaches. This highlights both the challenges and the richness of the field: the diversity of approaches and contexts show that IVR can be used successfully in a variety of domains, but the elements that distinguish a successful IVR application for procedural training remain difficult to pinpoint.

One attribute that we highlighted above is the use of enactment or gesture as an opportunity for embodied learning. The one study specifically testing this feature across non-immersive and IVR conditions, as well as other applications that compared voice versus gesture control of interactions (Jung & Ahn, 2018; Schroeder et al., 2017), suggests that this might play an important role in the learning outcomes of immersive procedural training. Thus, the interactivity offered by IVR applications, particularly in simulations that incorporate congruent, user-controlled physical gestures, appears to offer important opportunities for enhanced learning as compared to existing desktop e-learning environments.

These findings indicate that immersive procedural training may benefit from situated and embodied learning theories (for example, see Chaker et al., 2021) because pedagogical situations that mobilize the use of the physical body to perform the activity take advantage of embodied cognition to enhance learning (Johnson-Glenberg, 2018). This also helps to explain the comparatively advantageous results reported for procedural learning outcomes as opposed to theoretical or declarative learning outcomes, which may be due to an enhanced role played by embodied cognition in memory processes. Indeed, overall, our systematic review and

meta-analysis of the literature on immersive procedural training suggests that it may be more effective than less immersive technologies and that this may be due to the opportunities for embodied cognition offered by the interactivity of the medium.

Competing interests statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was supported by the RENFORCE project financed by the French National Research Agency (ANR) under Project number ANR-22-CE31-0023-03.

CRedit authorship contribution statement

Janine Jongbloed: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Rawad Chaker:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Elise Lavoué:** Writing – review & editing, Project administration, Methodology, Funding acquisition, Conceptualization.

Data availability

Data will be made available on request.

Appendix A. Studies included in the review

The 33 publications included in the systematic literature review and the 11 studies included in the meta-analysis are listed in Tables A1 and A2 below.

1.1. Studies included in the literature review

Table A1

Studies meeting the selection criteria.

Reference	Topic	Study type	Learning outcomes	Findings
Buttussi and Chittaro (2021)	Aviation safety procedures	PPC media comparison	KG: Describe steps in procedure; KR: Same test	IVR better than printed cards, no difference with smartphone
Buttussi et al. (2020)	Emergency CPR procedures	PPC value-added	KG: Describe steps in procedure & KT: Perform CPR on mannequin	No significant differences in procedures (only pressure-related errors)
Buttussi and Chittaro (2018)	Aviation safety procedures	PPC media comparison	KG: Describe steps in procedure; KR: Same test	No significant differences
Çakiroğlu and Gököglu (2019a)	Fire safety procedures	SGPP	KT: Behaviours in imitation fire situation	6 of 10 children correctly followed procedures
(S.-Y. Chen & Chien, 2022)	Fire safety procedures	SGPP	KT: Role play of fire situation	Positive change in fire safety knowledge
Clay et al. (2021)	Functional communication training	SGPP	KG: Evaluation mode in VR	Positive change in correct steps
Daling et al. (2020)	Manual assembly procedures	SG media comparison	None	NA
De Lorenzis et al. (2023)	Energy audit procedures	PPC media comparison and value-added	KG: Order steps of procedure; KG: Illustrate connections; & KR: Same test	IVR better than traditional; Learning-by-teaching better than without
Dhalmahapatra et al. (2021)	Crane operator procedures	SG media comparison	KG: # of hazards identified	IVR better than desktop
Dong et al. (2022)	Laboratory instrument procedures	Usability study	None	NA
Emond et al. (2022)	First responder procedures	Usability study	None	NA
Ganni et al. (2020)	Surgical procedures (doctor)	Usability study	None	NA
Grab et al. (2023)	Surgical procedures (patient)	PPC media comparison	KG: Knowledge of procedure	IVR better than printed and 3D models

(continued on next page)

Table A1 (continued)

Reference	Topic	Study type	Learning outcomes	Findings
Jung and Ahn (2018)	Lifeboat launching procedures	PPC media comparison and value-added	KG: Describe steps in procedure; KT: Describe how to use equipment (technical)	No difference in procedures; IVR better for technical skill
Katz et al. (2020)	Cardiac life support procedures	SG media comparison	KG: Describe steps in procedure	No difference
King et al. (2022)	Mathematical questioning procedure	PPC media comparison	KG: Role play of procedure; KR: Same test	IVR better than didactic instruction
Kushiro et al. (2021)	Field overseers procedures	Usability study	KG: # of hazards identified	Positive change in # identified
Logishetty et al. (2019)	Surgical procedures	PPC media comparison	KT: Cadaveric surgery assessment	IVR better than conventional preparation
Lohre et al. (2020)	Surgical procedures	PPC media comparison	KT: Cadaveric surgery assessment	IVR better than video instruction
Longo et al. (2023)	Crane operator procedures	Expert-novice comparison	KG: Evaluation mode in VR	Experts showed better performance
Longo et al. (2019)	Industrial emergency response procedures	SGPP	KG: Evaluation mode in VR	Positive change in performance
Lu et al. (2022)	Confined space rescue procedures	PPC media comparison	KG: Describe steps in procedure; KT: Role play of procedure	IVR better than e-lecture learning
Luo et al. (2021)	Public health emergency procedures	Usability study	None	NA
Makransky et al. (2021)	Forensic DNA procedures	PPC media comparison and value-added	KG: Describe steps in procedure; KG: Apply procedures verbally to a new situation	No difference between IVR and video; but significant interaction with enactment
Pérez-Escamirosa et al. (2022)	Surgical procedures	Expert-novice comparison	KG: Evaluation mode in VR	Experts showed better performance
Pieterse et al. (2023)	Hospital rounds procedures	Usability study	None	NA
Queisner et al. (2022)	Surgical procedure	Usability study	None	NA
Schroeder et al. (2017)	Military vehicle maintenance procedure	PPC media comparison	KG: Describe steps in procedure	No significant differences between IVR and desktop
Shah et al. (2022)	Surgical procedures	PPC media comparison	KG: Quiz on procedures; KT: medical simulator	No difference between IVR and Google Cardboard
Snarby et al. (2019)	Surgical procedures	Usability study	None	NA
Szczurowski and Smith (2018)	Road safety skills	Usability study	None	NA
Tai et al. (2022)	Car detailing procedures	SGPP	KG Evaluation mode in VR; KR: Same test	Improvement and retention
Tang et al. (2021)	Type and screen medical procedures	Usability study	None	Perceived enhanced readiness

Note: In this table, SG refers to “single group” and not “serious game.” SGPP indicates “single-group pre-test post-test” study design, while PPC indicates “pre-test post-test with control group” study design. KG denotes “knowledge gain,” KR denotes “knowledge retention,” and KT denotes “knowledge transfer” learning outcome measures.

1.2. Studies included in the meta-analysis

Table A2

PPC media comparison studies included in the meta-analysis.

Reference	Learning measure	Control group	Study n	Effect size
Buttusi and Chittaro (2021)	KG: Describe steps in procedure; KR: Same test	Smartphone	72	ns
Buttusi and Chittaro (2018)	KG: Describe steps in procedure; KR: Same test	Desktop	96	ns
De Lorenzis et al. (2023)	KG: Order steps of procedure; KG: Illustrate connections; & KR: Same test	Lab class	36	+
Grab et al. (2023)	KG: Knowledge of procedure	3-D physical model	99	+
Jung and Ahn (2018)	KG: Describe steps in procedure; KT: Describe how to use equipment (technical)	Desktop	64	ns/+
King et al. (2022)	KG: VR role play of procedure; KR: Same test	Regular instruction	30	+
Logishetty et al. (2019)	KT: Cadaveric surgery assessment	E-lecture	24	+
Lohre et al. (2020)	KT: Cadaveric surgery assessment	Video	18	+
Lu et al. (2022)	KG: Describe steps in procedure; KT: Real-life role play of procedure	E-lecture	22	+
Makransky et al. (2021)	KG: Describe steps in procedure; KG: Apply procedures verbally to a new situation	Video	131	ns
Schroeder et al. (2017)	KG: Describe steps in procedure	Desktop	75	ns

Note: In this table, KG denotes “knowledge gain,” KR denotes “knowledge retention,” and KT denotes “knowledge transfer” learning outcome measures. For the effect size, “ns” refers to “not significant” and “+” to a significant positive effect (there were no studies with significant negative effects).

References

- Andrade, A., Danish, J. A., & Maltese, A. V. (2017). A measurement Model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, 4(3), 18–45. <https://doi.org/10.18608/jla.2017.43.3>
- Baroody, A. J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 1–33). Lawrence Erlbaum Associates Publishers.
- Barsalou, L. W. (2010). Grounded cognition: Past, present, and future. *Topics in Cognitive Science*, 2(4), 716–724. <https://doi.org/10.1111/j.1756-8765.2010.01115.x>
- Buttussi, F., & Chittaro, L. (2018). Effects of different types of virtual reality display on presence and learning in a safety training scenario. *IEEE Transactions on Visualization and Computer Graphics*, 24(2), 1063–1076. <https://doi.org/10.1109/TVCG.2017.2653117>
- Buttussi, F., & Chittaro, L. (2021). A comparison of procedural safety training in three conditions: Virtual reality headset, smartphone, and printed materials. *IEEE Transactions on Learning Technologies*, 14(1), 1–15. <https://doi.org/10.1109/TLT.2020.3033766>
- Buttussi, F., Chittaro, L., & Valent, F. (2020). A virtual reality methodology for cardiopulmonary resuscitation training with and without a physical mannequin. *Journal of Biomedical Informatics*, 111, Article 103590. <https://doi.org/10.1016/j.jbi.2020.103590>
- Çakiroğlu, Ü., & Gököglü, S. (2019a). Development of fire safety behavioral skills via virtual reality. *Computers & Education*, 133, 56–68. <https://doi.org/10.1016/j.compedu.2019.01.014>
- Çakiroğlu, Ü., & Gököglü, S. (2019b). A design Model for using virtual reality in behavioral skills training. *Journal of Educational Computing Research*, 57(7), 1723–1744. <https://doi.org/10.1177/0735633119854030>
- Chaker, R. (2023). Studying ergonomic gestures on interactive learning tabletop and tablets from a situated approach: A pilot taxonomic table and a case study. *Journal of Computers in Education*. <https://doi.org/10.1007/s40692-023-00306-7>
- Chaker, R., Gallot, M., Binay, M., & Hoyek, N. (2021). User experience of a 3D interactive human anatomy learning tool. *Educational Technology & Society*, 24(2), 136–150.
- Chan, P., Van Gerven, T., Dubois, J.-L., & Bernaerts, K. (2021). Virtual chemical laboratories: A systematic literature review of research, technologies and instructional design. *Computers and Education Open*, 2, Article 100053. <https://doi.org/10.1016/j.caeo.2021.100053>
- Chang, M. M., & Hung, H. T. (2019). Effects of technology-enhanced language learning on second language acquisition: A meta-analysis. *Educational Technology & Society*, 22(4), 1–17.
- Chemero, A. P. (2011). *Radical embodied cognitive science*. MIT Press.
- Chen, S.-Y., & Chien, W.-C. (2022). Immersive virtual reality serious games with DL-assisted learning in high-rise fire evacuation on fire safety training and research. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.786314>
- Chen, J., Fu, Z., Liu, H., & Wang, J. (2023). Effectiveness of virtual reality on learning engagement. *International Journal of Web-Based Learning and Teaching Technologies*, 19(1), 1–14. <https://doi.org/10.4018/IJWLTT.334849>
- Clark, A. (1998). Embodiment and the philosophy of mind. *Royal Institute of Philosophy Supplement*, 43, 35–51. <https://doi.org/10.1017/s135824610000429x>, 1979.
- Clay, C. J., Schmitz, B. A., Balakrishnan, B., Hopfenblatt, J. P., Evans, A., & Kahng, S. (2021). Feasibility of virtual reality behavior skills training for preservice clinicians. *Journal of Applied Behavior Analysis*, 54(2), 547–565. <https://doi.org/10.1002/jaba.809>
- Coban, M., Bolat, Y. I., & Goksu, I. (2022). The potential of immersive virtual reality to enhance learning: A meta-analysis. *Educational Research Review*, 36(March), Article 100452. <https://doi.org/10.1016/j.edurev.2022.100452>
- Concannon, B. J., Esmail, S., & Roduta Roberts, M. (2019). Head-mounted display virtual reality in post-secondary education and skill training. *Frontiers in Education*, 4. <https://doi.org/10.3389/educ.2019.00080>
- Corris, A., & Chemero, A. (2022). Embodiment and enactivism. In *Mind, cognition, and neuroscience* (pp. 414–428). Routledge. <https://doi.org/10.4324/9781003241898-30>.
- Daling, L. M., Abdelrazeq, A., & Isenhardt, I. (2020). A comparison of augmented and virtual reality features in industrial trainings. In *Lecture notes in computer science: Vol. 12191 LNCS*. Springer International Publishing. https://doi.org/10.1007/978-3-030-49698-2_4
- De Lorenzis, F., Praticò, F. G., Repetto, M., Pons, E., & Lamberti, F. (2023). Immersive Virtual Reality for procedural training: Comparing traditional and learning by teaching approaches. *Computers in Industry*, 144. <https://doi.org/10.1016/j.compind.2022.103785>
- Dhalmahapatra, K., Maiti, J., & Krishna, O. B. (2021). Assessment of virtual reality based safety training simulator for electric overhead crane operations. *Safety Science*, 139, Article 105241. <https://doi.org/10.1016/j.ssci.2021.105241>
- Dong, J., Li, D., Ozcan, K., Wan, D., Jiang, W., & Chen, Y. (2022). Development of CryoVR, a virtual reality training system for hands-on cryoEM operations. *Structural Biology*, 78(7), 903–910. <https://doi.org/10.1107/S2059798322005654>
- Emond, B., Lapointe, J. F., Kinatader, M., DeMarbre, E., Berthiaume, M., Kondratova, I., Cooper, N., Bénichou, N., & Downes, S. (2022). HCI issues, design and development for a first responders VR training system on dangerous goods transportation incidents. *Lecture Notes in Computer Science*, 13329(LNCS), 48–60. https://doi.org/10.1007/978-3-031-05675-8_5
- Feng, Z., González, V. A., Amor, R., Lovreglio, R., & Cabrera-Guerrero, G. (2018). Immersive virtual reality serious games for evacuation training and research: A systematic literature review. *Computers and Education*, 127, 252–266. <https://doi.org/10.1016/j.compedu.2018.09.002>
- Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, 28(4), 717–741. <https://doi.org/10.1007/s10648-015-9348-9>
- Fiorella, L., & Mayer, R. E. (2021). The generative activity principle in multimedia learning. In *The cambridge handbook of multimedia learning* (pp. 339–350). Cambridge University Press. <https://doi.org/10.1017/9781108894333.036>.
- Gagne, E. D., Yekovich, C. W., & Yekovich, F. R. (1997). *The cognitive psychology of school learning*. Addison Wesley Longman Inc.
- Ganier, F., Hoareau, C., & Tisseau, J. (2014). Evaluation of procedural learning transfer from a virtual environment to a real situation: A case study on tank maintenance training. *Ergonomics*, 57(6), 828–843. <https://doi.org/10.1080/00140139.2014.899628>
- Ganni, S., Li, M., Botden, S. M. B. I., Nayak, S. R., Ganni, B. R., Rutkowski, A.-F., Goossens, R. H. M., & Jakimowicz, J. (2020). Virtual operating room simulation setup (VORSS) for procedural training in minimally invasive surgery – a pilot study. *Indian Journal of Surgery*, 82(5), 810–816. <https://doi.org/10.1007/s12262-020-02131-z>
- Gao, Y., Gonzalez, V. A., & Yiu, T. W. (2019). The effectiveness of traditional tools and computer-aided technologies for health and safety training in the construction sector: A systematic review. *Computers and Education*, 138(April), 101–115. <https://doi.org/10.1016/j.compedu.2019.05.003>
- Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, 66(3), 325–331.
- Gormley, G. J., & Murphy, P. (2023). When I say ... simulation. *Medical Education*, June, 10–11. <https://doi.org/10.1111/medu.15165>
- Grab, M., Hundertmark, F., Thierfelder, N., Fairchild, M., Mela, P., Hagl, C., & Grefen, L. (2023). *New perspectives in patient education for cardiac surgery using 3D-printing and virtual reality* (Vol. 10). Frontiers in Cardiovascular Medicine. <https://doi.org/10.3389/fcvm.2023.1092007>
- Hamilton, D., McKechnie, J., Edgerton, E., & Wilson, C. (2021). Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education*, 8(1), 1–32. <https://doi.org/10.1007/s40692-020-00169-2>
- Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, 7(2), 174–196. <https://doi.org/10.1145/353485.353487>
- Ioannidis, J. P. A., & Trikalinos, T. A. (2007). The appropriateness of asymmetry tests for publication bias in meta-analyses: A large survey. *CMAJ. Canadian Medical Association Journal*, 176(8), 1091–1096. <https://doi.org/10.1503/cmaj.060410>
- Jang, S., Vitale, J. M., Jyung, R. W., & Black, J. B. (2017). Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment. *Computers and Education*, 106, 150–165. <https://doi.org/10.1016/j.compedu.2016.12.009>
- Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, 23(4), 1515–1529. <https://doi.org/10.1007/s10639-017-9676-0>

- Johnson, C. I., Bailey, S. K. T., Schroeder, B. L., & Marraffino, M. D. (2022). Procedural learning in virtual reality: The role of immersion, interactivity, and spatial ability. *Technology, Mind, and Behavior*, 3(4). <https://doi.org/10.1037/tmb0000087>
- Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI*, 5 (JUN). <https://doi.org/10.3389/frobt.2018.00081>
- Johnston, E., Olivas, G., Steele, P., Smith, C., & Bailey, L. (2018). Exploring pedagogical foundations of existing virtual reality educational applications: A content analysis study. *Journal of Educational Technology Systems*, 46(4), 414–439. <https://doi.org/10.1177/0047239517745560>
- Jung, J., & Ahn, Y. J. (2018). Effects of interface on procedural skill transfer in virtual training: Lifeboat launching operation study. *Computer Animation and Virtual Worlds*, 29(3–4), Article e1812. <https://doi.org/10.1002/cav.1812>
- Kaplan, A. D., Cruitt, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2021). The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis. *Human Factors*, 63(4), 706–726. <https://doi.org/10.1177/0018720820904229>
- Karaiskaki, A., & Anastassiou-Hadjicharalambous, X. (2021). Procedural knowledge. In T. K. Shackelford, & V. A. Weekes-Shackelford (Eds.), *Encyclopedia of evolutionary psychological science* (pp. 6298–6300). Springer International Publishing. https://doi.org/10.1007/978-3-319-19650-3_1083
- Katz, D., Levine, A., Burnett, G. W., Shah, R., Kim, E., Park, C., Shah, A., & Burnett, G. (2020). Utilization of a voice-based virtual reality advanced cardiac LifeSupport team leader refresher: Prospective observational study. *Journal of Medical Internet Research*, 22(3). <https://doi.org/10.2196/preprints.17425>
- King, S., Boyer, J., Bell, T., & Estapa, A. (2022). An automated virtual reality training system for teacher-student interaction: A randomized controlled trial. *JMIR Serious Games*, 10(4). <https://doi.org/10.2196/41097>
- Krassmann, A. L., Melo, M., Pinto, D., Peixoto, B., Bessa, M., & Bercht, M. (2019). What is the relationship between the sense of presence and learning in virtual reality? A 24-year systematic literature review. *PRESENCE: Virtual and Augmented Reality*, 28, 247–265. https://doi.org/10.1162/pres_a.00350
- Krassmann, A. L., Melo, M., Pinto, D., Peixoto, B., Bessa, M., & Bercht, M. (2023). How are the sense of presence and learning outcomes being investigated when using virtual reality? A 24 years systematic literature review. *Interactive Learning Environments*, (March), 1–24. <https://doi.org/10.1080/10494820.2023.2184388>
- Kushiro, N., Nishinaga, K., & Mega, T. (2021). Training tool on structured knowledge for risk management with VR technology. In *Lecture notes in computer science: Vol. 13095 LNCS* (pp. 262–281). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-90963-5_21
- Lavoue, E., Monterrat, B., Desmarais, M., & George, S. (2019). Adaptive gamification for learning environments. *IEEE Transactions on Learning Technologies*, 12(1), 16–28. <https://doi.org/10.1109/TLT.2018.2823710>
- Li, M., Ganni, S., Ponten, J., Albayrak, A., Rutkowski, A.-F., & Jakimowicz, J. (2020). Analysing usability and presence of a virtual reality operating room (VOR) simulator during laparoscopic surgery training. *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*, March. <https://doi.org/10.1109/VR46266.2020.1581301697128>
- Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by embodiment. *Educational Researcher*, 42(8), 445–452. <https://doi.org/10.3102/0013189X13511661>
- Logishetty, K., Rudran, B., & Cobb, J. P. (2019). Virtual reality training improves trainee performance in total hip arthroplasty: A randomized controlled trial. *The Bone & Joint Journal*, 101(B), 1585–1592.
- Lohre, R., Bois, A. J., Pollock, J. W., Lapner, P., McIlquham, K., Athwal, G. S., & Goel, D. P. (2020). Effectiveness of immersive virtual reality on orthopedic surgical skills and knowledge acquisition among senior surgical residents. *JAMA Network Open*, 3(12), Article e2031217. <https://doi.org/10.1001/jamanetworkopen.2020.31217>
- Longo, F., Nicoletti, L., & Padovano, A. (2019). Emergency preparedness in industrial plants: A forward-looking solution based on industry 4.0 enabling technologies. *Computers in Industry*, 105, 99–122. <https://doi.org/10.1016/j.compind.2018.12.003>
- Longo, F., Padovano, A., De Felice, F., Petrillo, A., & Elbasheer, M. (2023). From “prepare for the unknown” to “train for what’s coming”: A digital twin-driven and cognitive training approach for the workforce of the future in smart factories. *Journal of Industrial Information Integration*, 32, Article 100437. <https://doi.org/10.1016/j.jii.2023.100437>
- Lu, S., Wang, F., Li, X., & Shen, Q. (2022). Development and validation of a confined space rescue training prototype based on an immersive virtual reality serious game. *Advanced Engineering Informatics*, 51. <https://doi.org/10.1016/j.aei.2021.101520>
- Lu, S., Xu, W., Wang, F., Li, X., & Yang, J. (2020). Serious game: Confined space rescue based on virtual reality technology. *2020 2nd international conference on video, signal and image processing*, July. <https://doi.org/10.1145/3442705.3442716>
- Lungu, A. J., Swinkels, W., Claesen, L., Tu, P., Egger, J., & Chen, X. (2021). A review on the applications of virtual reality, augmented reality and mixed reality in surgical simulation: An extension to different kinds of surgery. *Expert Review of Medical Devices*, 18(1), 47–62. <https://doi.org/10.1080/17434440.2021.1860750>
- Luo, Y., Li, M., Tang, J., Ren, J., Zheng, Y., Yu, X., Jiang, L., Fan, D., & Chen, Y. (2021). Design of a virtual reality interactive training system for public health emergency preparedness for major emerging infectious diseases: Theory and framework. *JMIR Serious Games*, 9(4), Article e29956. <https://doi.org/10.2196/29956>
- Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, 32(6), 982–998. <https://doi.org/10.1002/hbm.21084>
- Makransky, G., Andreassen, N. K., Baceviciute, S., & Mayer, R. E. (2021). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology*, 113(4), 719–735. <https://doi.org/10.1037/edu0000473>
- Makransky, G., & Petersen, G. B. (2021). The cognitive affective Model of immersive learning (CAMIL): A theoretical research-based Model of learning in immersive virtual reality. *Educational Psychology Review*, 33(3), 937–958. <https://doi.org/10.1007/s10648-020-09586-2>. Springer.
- Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, 60, 225–236. <https://doi.org/10.1016/j.learninstruc.2017.12.007>
- Mao, R. Q., Lan, L., Kay, J., Lohre, R., Ayeni, O. R., Goel, D. P., & Sa, D. de (2021). Immersive virtual reality for surgical training: A systematic review. *Journal of Surgical Research*, 268(December), 40–58. <https://doi.org/10.1016/j.jss.2021.06.045>
- Matovu, H., Ungu, D. A. K., Won, M., Tsai, C.-C., Treagust, D. F., Mocerino, M., & Tasker, R. (2022). Immersive virtual reality for science learning: Design, implementation, and evaluation. *Studies in Science Education*, 1–40. <https://doi.org/10.1080/03057267.2022.2082680>
- Mayer, R. E. (2009). *Multimedia learning*. Cambridge University Press.
- Mayer, R. E. (2021). Cognitive theory of multimedia learning. In *The cambridge handbook of multimedia learning* (pp. 57–72). Cambridge University Press. <https://doi.org/10.1017/9781108894333.008>
- Mayer, R. E., Makransky, G., & Parong, J. (2023). The promise and pitfalls of learning in immersive virtual reality. *International Journal of Human-Computer Interaction*, 39(11), 2229–2238. <https://doi.org/10.1080/10447318.2022.2108563>
- McGowin, G., Fiore, S. M., & Oden, K. (2021). A meta-review of learning research in immersive virtual reality. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting*, 65(1), 894–898. <https://doi.org/10.1177/1071181321651326>
- Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students’ learning outcomes in K-12 and higher education: A meta-analysis. *Computers and Education*, 70, 29–40. <https://doi.org/10.1016/j.compedu.2013.07.033>
- Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education*, 140, Article 103603. <https://doi.org/10.1016/j.compedu.2019.103603>
- Meyer, L., & Pfeiffer, T. (2020). *Comparing virtual reality and screen-based training simulations in terms of learning and recalling declarative knowledge*.
- Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers and Education*, 56(3), 769–780. <https://doi.org/10.1016/j.compedu.2010.10.020>
- Morelot, S., Garrigou, A., Dedieu, J., & N’Kaoua, B. (2021). Virtual reality for fire safety training: Influence of immersion and sense of presence on conceptual and procedural acquisition. *Computers and Education*, 166. <https://doi.org/10.1016/j.compedu.2021.104145>
- Murre, J. M. J. (2021). The godden and Baddeley (1975) experiment on context-dependent memory on land and underwater: A replication. *Royal Society Open Science*, 8(11), 325–331. <https://doi.org/10.1098/rsos.200724>

- Numfu, M., Riel, A., & Noël, F. (2020). Virtual reality based digital chain for creating a knowledge base of hand gestures in maintenance tasks. *Procedia CIRP*, 90, 648–653. <https://doi.org/10.1016/j.procir.2020.01.122>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ*, 372. <https://doi.org/10.1136/bmj.n71>
- Papanastasiou, G., Drigas, A., Skianis, C., Lytras, M., & Papanastasiou, E. (2019). Virtual and augmented reality effects on K-12, higher and tertiary education students' twenty-first century skills. *Virtual Reality*, 23(4), 425–436. <https://doi.org/10.1007/s10055-018-0363-2>
- Parong, J. (2021). Multimedia learning in virtual and mixed reality. In *The cambridge handbook of multimedia learning* (pp. 498–509). Cambridge University Press. <https://doi.org/10.1017/9781108894333.051>
- Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*, 110(6), 785–797. <https://doi.org/10.1037/edu0000241>
- Pérez-Escamirosa, F., García-Cabra, D. A., Ortiz-Hernández, J. R., Montoya-Alvarez, S., Ruíz-Vereo, E. A., Ordorica-Flores, R. M., Minor-Martínez, A., & Tapiá-Jurado, J. (2022). Face, content, and construct validity of the virtual immersive operating room simulator for training laparoscopic procedures. *Surgical Endoscopy*, Article 0123456789. <https://doi.org/10.1007/s00464-022-09797-4>
- Petersen, G. B., Petkakis, G., & Makransky, G. (2022). A study of how immersion and interactivity drive VR learning. *Computers & Education*, 179(December 2021), Article 104429. <https://doi.org/10.1016/j.compedu.2021.104429>
- Pieterse, A. D., Hierck, B. P., de Jong, P. G. M., Ginn, T. F., Hamoen, E. C., & Reinders, M. E. J. (2023). User experiences of medical students with 360-degree virtual reality applications to prepare them for the clerkships. *Virtual Reality*. <https://doi.org/10.1007/s10055-022-00731-6>
- Queisner, M., Pogorzelskiy, M., Remde, C., Pratschke, J., & Sauer, I. M. (2022). VolumetricOR: A new approach to simulate surgical interventions in virtual reality for training and education. *Surgical Innovation*, 29(3), 406–415. <https://doi.org/10.1177/15533506211054240>
- Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147, Article 103778. <https://doi.org/10.1016/j.compedu.2019.103778>
- Repetto, C., Cipresso, P., & Riva, G. (2015). Virtual action and real action have different impacts on comprehension of concrete verbs. *Frontiers in Psychology*, 6(FEB), 1–9. <https://doi.org/10.3389/fpsyg.2015.00176>
- Ricci, F., Chiesi, A., Bisio, C., Panari, C., & Pelosi, A. (2016). Effectiveness of occupational health and safety training: A systematic review with meta-analysis. *Journal of Workplace Learning*, 28(6), 355–377. <https://doi.org/10.1108/JWL-11-2015-0087>
- Rizzo, A., & Bouchard, S. (Eds.). (2019). *Virtual reality for psychological and neurocognitive interventions*. Nature Switzerland AG: Springer. <https://doi.org/10.1007/978-1-4939-9482-3>
- Schroeder, B. L., Bailey, S. K. T., Johnson, C. I., & Gonzalez-Holland, E. (2017). Presence and usability do not directly predict procedural recall in virtual reality training. In *Communications in computer and information science* (Vol. 714, pp. 54–61). Springer Verlag. https://doi.org/10.1007/978-3-319-58753-0_9
- Shah, N. K., Taunk, N. K., Maxwell, R., Wang, X., Hubley, E., Anamalayil, S., Trotter, J. W., & Li, T. (2022). Comparison of virtual reality platforms to enhance medical education for procedures. *Frontiers in Virtual Reality*, 3. <https://doi.org/10.3389/frvir.2022.1000035>
- Snarby, H., Gåsbakk, T., Prasolova-Forland, E., Steinsbekk, A., & Lindseth, F. (2019). Procedural medical training in VR in a smart virtual university hospital. In *Smart innovation, systems and technologies* (Vol. 99, pp. 132–141). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-319-92363-5_12
- StataCorp. (2021). *Stata statistical software: Release 17*. StataCorp LLC. No. 17.
- Stefan, H., Mortimer, M., & Horan, B. (2023). Evaluating the effectiveness of virtual reality for safety-relevant training: A systematic review. In *Virtual reality*. London: Springer. <https://doi.org/10.1007/s10055-023-00843-7>. Issue 0123456789.
- Szczurowski, K., & Smith, M. (2018). 'Woodlands'-A virtual reality serious game supporting learning of practical road safety skills. *IEEE Games, Entertainment, Media Conference, GEM 2018*, 427–435. <https://doi.org/10.1109/GEM.2018.8516493>, 2018.
- Tai, K.-H., Hong, J.-C., Tsai, C.-R., Lin, C.-Z., & Hung, Y.-H. (2022). Virtual reality for car-detailing skill development: Learning outcomes of procedural accuracy and performance quality predicted by VR self-efficacy, VR using anxiety, VR learning interest and flow experience. *Computers & Education*, 182, Article 104458. <https://doi.org/10.1016/j.compedu.2022.104458>
- Tang, Y. M., Ng, G. W. Y., Chia, N. H., So, E. H. K., Wu, C. H., & Ip, W. H. (2021). Application of virtual reality (VR) technology for medical practitioners in type and screen (T&S) training. *Journal of Computer Assisted Learning*, 37(2), 359–369. <https://doi.org/10.1111/jcal.12494>
- Taunk, N. K., Shah, N. K., Hubley, E., Anamalayil, S., Trotter, J. W., & Li, T. (2021). Virtual reality-based simulation improves gynecologic brachytherapy proficiency, engagement, and trainee self-confidence. *Brachytherapy*, 20(4), 695–700. <https://doi.org/10.1016/j.brachy.2021.03.003>
- Versace, R., Bailloud, N., Magnan, A., & Ecalte, J. (2021). The impact of embodied simulation in vocabulary learning. *The Mental Lexicon*, 16(1), 2–22. <https://doi.org/10.1075/ml.20011.ver>
- Wilson, M. (2002). Six views of embodied Cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). Experimentation in software engineering. In *Experimentation in software engineering* (Vol. 9783642290). <https://doi.org/10.1007/978-3-642-29044-2>
- Won, M., Ungu, D. A. K., Matovu, H., Tregust, D. F., Tsai, C.-C., Park, J., Mocerino, M., & Tasker, R. (2023). Diverse approaches to learning with immersive Virtual Reality identified from a systematic review. *Computers & Education*, 195(May 2022), Article 104701. <https://doi.org/10.1016/j.compedu.2022.104701>
- Wu, B., Yu, X., & Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology*, 51(6), 1991–2005. <https://doi.org/10.1111/bjet.13023>
- Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F. D., Li, C., Li, W., Huang, H., Akdere, M., Mousas, C., & Yu, L.-F. (2021). A review on virtual reality skill training applications. *Frontiers in Virtual Reality*, 2(April), 1–19. <https://doi.org/10.3389/frvir.2021.645153>