seminar.net
MEDIA, TECHNOLOGY & LIFELONG LEARNING
www.seminar.net

ISSN: 1504-4831 Vol 18, No 1 (2022) https://doi.org/10.7577/seminar.4692

Learning Outcomes and Learning Motivation in HMD-VR: A Literature Review

Antti Lähtevänoja
University of Helsinki
antti.lahtevanoja@helsinki.fi

Mikko Vesisenaho
University of Helsinki
mikko.vesisenaho@jyu.fi

Kati Vasalampi
University of Jyväskylä
kati.vasalampi@jyu.fi, https://orcid.org/0000-0002-1249-7566

Jani Holopainen
University of Eastern Finland
jani.holopainen@uef.fi, https://orcid.org/0000-0002-1609-8204

Päivi Häkkinen
University of Jyväskylä
paivi.m.hakkinen@jyu.fi

Abstract

While educational technology has developed to the point that extended reality (XR), including immersive virtual reality (VR), can be used in education, the overall learning outcomes of these technologies are still unknown. This literature review takes a comprehensive look at the field of immersive VR and explores the points at which the learning outcomes of head-mounted display (HMD) VR stand out and how these outcomes compare to those of other technologies and methods. Learning outcomes can be either direct learning outcomes (e.g., test results) or indirect learning outcomes which affect learning, such as learning motivation and engagement. The main result of this paper is that while HMD-based VR learning environments may not be superior to other technologies, such as desktop-based VR environments, regarding direct learning outcomes, there is a clear indication of increased learning motivation and engagement.

Keywords: virtual reality, learning outcomes, learning, learning motivation, learning engagement, head-mounted displays, immersive

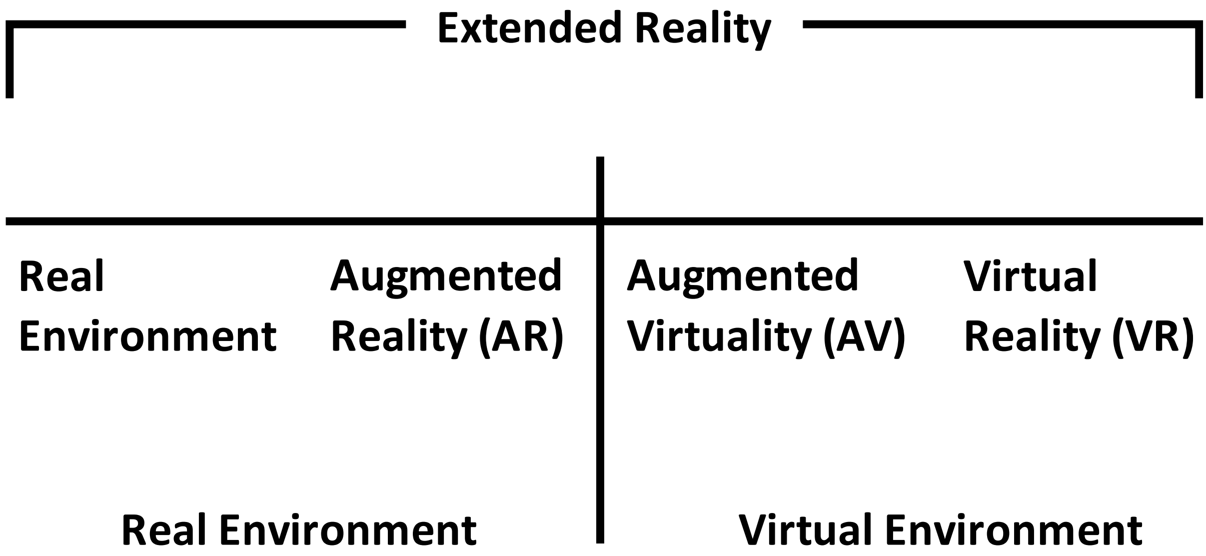
Introduction

Educational technology has experienced such remarkable development that extended reality (XR) technologies, such as augmented reality (AR), augmented virtuality (AV) and virtual reality (VR), can now be used in educational settings. Furthermore, the development of affordable head-mounted displays (HMDs) allows VR technologies to be used in educational contexts (Freina & Ott 2015; Freina et al., 2016). HMDs allow students to become immersed in virtual learning environments (VRLEs), enabling them to conduct different learning activities in these environments. This study is focused on HMD-VR i.e., a VR-based environment experienced through head-mounted-displays (HMDs) and particularly HMD-based VRLE i.e., a virtual learning environment experienced through head-mounted-displays.

The primary aim of this literature review is to investigate the effects of VRLEs in terms of learning outcomes. In other words, what are the direct or indirect learning outcomes of HMD-based VRLEs? In order to make evidence-based decisions about whether to use VR in education, there is a need for knowledge on the learning outcomes of VRLEs. It is important to remember that it is possible to undertake learning assignments with VR, which cannot be conducted in real life; however, it is not likely that VRLEs could be used as a substitute for all conventional assignments (Sathe et al., 2017). Nevertheless, this makes comparing HMD-VR to other technologies difficult as it is not always possible to conduct the same learning tasks using other technologies. Investigating the kinds of direct learning outcomes (e.g., improved knowledge, practical skills) for which HMD-VR would be most effective would be beneficial both for academia and the use of VRLEs in the field. In addition, it is important to study whether HMD-based VRLEs have any effect on indirect learning outcomes such as learning motivation and engagement. A higher learning motivation can influence direct learning outcomes in a positive manner (e.g., Triarisanti & Purnawarman 2019). For example, previous research has shown that high learning motivation and engagement can keep students coming back to learn more (Kuo, 2007), facilitate positive emotions in learning situations (Vasalampi et al., 2021), and they have a significant positive influence on deep learning (Everaert et al., 2017).

Figure 1

Reality continuum



Reality can be thought of as a continuum (Figure 1, based on Mann et al., 2018). A continuum starts from the real environment where there are no virtual objects. When computer-generated material is added to the real environment, for example, a 3D model through AR glasses or a mobile phone, the state is called AR (Mann, 2018). After AR, the environment changes from real to virtual. In the virtual environment, the first phase is AV, which, although based on a virtual environment, has real components, such as tangible objects (Neges et al., 2018). An example is a green screen capture of a human being. Finally, VR is a fully computer-generated simulation with no real components. It can be defined as a computer-generated artificial environment, which can be interactive (Guttentag, 2010). VR can be fully immersive when viewed through HMDs or semi-immersive when used through CAVE systems (VR environment projection) or via two-dimensional screens, such as computers or tablets. Immersion leads to presence, an important feature of VR that can be explained as the feeling of 'being there' (Ijsselsteijn & Riva, 2003). According to Pollard et al. (2020), immersion and the usage of immersive technology exemplify an 'enhanced training component' and can keep learners focused, help them maintain interest and motivate them to invest sufficient time on a task. HMDs can fully immerse learners into the virtual environment and allow user involvement (Vesisenaho et al., 2019). In addition, the immersion of VR-HMD can trigger emotionally significant learning situations and induce learners' engagement in learning activities (Bosse et al., 2014; Hanson & Shelton, 2008). This allows learners to be engaged in a deeper cognitive processing of the learning material (Huang et al., 2010).

However, to make broad and effective use of these technologies in educational institutions or workplace training settings, their effects regarding learning outcomes need to be explored (Holopainen et al., 2020; Vesisenaho et al., 2019). For example, it is crucial to explore whether HMD-based VRLEs lead to better learning outcomes over more traditional technologies or learning methods, such as desktop-based VR applications or normal classroom teaching. HMD-VR is also faced with challenges, limiting its adoption to the educational field, in particular; while the hardware is developing, the current equipment still has an entertainment purpose (Jensen & Kondradsen, 2018). In addition, the development and content production for HMD-VR are also expensive (Carretero et al., 2021; Jensen & Konradsen, 2018). In this study, we focused on HMD-based VR systems, and therefore, only studies involving VR environments with HMD displays were included.

This paper will proceed as follows: first, some previous literature reviews on the topic are presented and discussed, followed by the research aims and questions. After that, the research methodology partly applying the categories of earlier study by Radianti et al. (2020) is presented. Lastly, we present the results of the literature review and discuss the importance of the findings.

Previous Literature Reviews

Some HMD-VR-based literature reviews (e.g., Feng et al., 2018; Hamilton et al., 2020; Jensen & Konradsen, 2018; Radianti et al., 2020) have been conducted, which have taken a more explorative approach to the previous research by listing the targeted learning outcomes. As stated before, in order to make evidence-based decisions about whether to use VR in education, there is a need for knowledge on the learning outcomes of VRLEs. Some literature reviews have focused on learning outcomes, but in a limited context, e.g., in a certain educational level such as high school.

Hamilton et al. (2020) conducted a systematic review of quantitative learning outcomes (e.g., test score, completion time, knowledge retention) and experimental design. The primary finding regarding learning outcomes was that around half of the 29 studies examined had a positive effect on learning when HMD-VR was used compared to less immersive methods. They restricted their review to high school, higher education and adult education students and included only studies with comparable HMD-VR groups which underwent other educational methods such as non-immersive technologies (e.g., laptop or desktop PC screen). Queiroz et al. (2018) conducted a review of the use of HMD-VR-based VRLEs in the context of primary school/K12 education. Noteworthy, they also included 360-degree video-based solutions, which are not regarded as VRLEs in the present article. They found that the papers under study had three learning outcomes: knowledge-based, abilities-based and skills-based. Most of these studies reported an improvement in learning outcomes. In addition, for the abilities-based studies, the VR groups were found to have more task motivation and engagement than the groups using other technologies or methods. Jensen and Konradsen (2018) conducted a review of the use of HMD-VR in education and training. They identified some key situations where HMD-VR was useful for skills acquisition: cognitive skills related to remembering and understanding spatial and visual information and knowledge, psychomotor skills related to head-movement, and affective skills related to controlling one’s emotional response. Outside these situations, HMD-VR had no advantage over other less immersive technologies.

Feng et al. (2018) conducted a systematic review of VR-based serious games for evacuation training. They included fifteen studies, five of which adopted a pedagogical approach. Using these studies, Feng et al. investigated key elements regarding the development and implementation of the games and proposed a conceptual framework for integrating and connecting these elements.

Radianti et al. (2020) conducted a systematic review of immersive VR applications for higher education. They examined the VR technologies, application domains, learning content and design elements of the 38 articles in their analysis. Learning outcomes were not their focus in the review, but they identified 18 application domains for VRLEs and pointed to some unexplored regions of the VRLE design, thereby motivating future work in the field. Bradley and Newbutt (2018) conducted a systematic literature review on the use of HMD-VR with autism. As VR technology can provide authentic, real-world conditions and can be used to train social and life skills in a safe and controlled environment (Bradley & Newbutt, 2018). This review did not focus on the learning outcomes of HMD-VR. They also found out that not much studies regarding the use of HMD-VR in autism have been conducted.

Summarising these previous reviews, it can be concluded that overviews of the learning outcomes of VRLEs have not resulted in increased clarity. It seems that in some situations (Jensen & Konradsen, 2018) and for some learning outcomes, such as those of a knowledge-based nature (Queiroz et al., 2018), HMD-based VRLEs can be beneficial. However, more data are needed to explore the learning outcomes of HMD-VR-based VRLEs. Moreover, only the review by Queiroz et al. (2018) included learning motivation and engagement as learning outcome variables. Their review indicated positive effect of HMD-VR-based VRLEs for motivation and engagement, but more evidence is needed before generalization of the results.

In addition, previous literature reviews have had quite strict inclusion criteria regarding the educational levels examined (e.g., only high school). This review aims to present a more comprehensive overview of the field as a whole and explore the points at which the learning outcomes of HMD-VR stands out and how these learning outcomes compare to those of other technologies and teaching methods. In addition, the studies under review are not limited to a specific educational level or even to educational studies in general. The results of this literature review present therefore an overall picture of the effects of VRLEs across different educational levels and contexts.

Aims and Research Questions

In this study, the primary aim is to investigate the learning outcomes of VRLEs, both from the viewpoint of direct learning outcomes (directly measurable outcomes, e.g., test results) and indirect learning outcomes (e.g., learning motivation and engagement, which will indirectly affect learning and cannot easily be measured using traditional pre- and post-tests). In this study, direct learning outcomes are categorised according to different learning goals based on primary learning outcomes (e.g., procedural-practical knowledge or learning a language).

This literature review aims to provide a comprehensive overview of the field and answer the following primary research question:

1. What research evidence is available regarding the relationship between HMD-based VRLEs and their learning outcomes?
1.1. What kinds of learning outcomes do HMD-based VRLEs offer in terms of different learning goals?
1.2. What kinds of outcomes do HMD-based VRLEs offer in terms of learning motivation and engagement?

Research Methodology

In order to answer the research questions, a systematic literature review was conducted. The process was started by following the guidelines for systematic literature reviews (Boell & Cecez-Kecmanovic, 2015; Webster & Watson, 2002), and the first stage of the present review was to conduct a keyword search. The main keyword search was focused on one database (Scopus) to ensure that the search procedure was rigorous, replicable and transparent due to the differences in the search functions and algorithms (Morschheuser et al., 2017). In addition, initial searches showed that database-based searches work better than journal searches as studies are published in many fields and journals, not just in the educational field. Thus, Scopus was chosen as the starting point of the keyword search as it indexes the relevant publication platforms on this field of study (e.g., Computers & Education, Educational Technology & Society).

The main research focus was to find studies conducted using HMDs in the context of learning. Initial searches revealed that, in some studies, the term 'training' was used as a substitute for 'learning'. Therefore, both search terms were included in the search using Boolean search operators. In order to include all variants of the search words (e.g., learn, learning, train, training), an asterisk was used. For the term 'head mounted display', both the abbreviation 'HMD' and the unabbreviated term were included. The final search term with Boolean operators was as follows: 'TITLE-ABS-KEY (virtual reality AND (train* OR learn*)) AND (hmd OR head mounted display)'). The search was conducted in October 2021 and resulted in 1,086 hits. No time frame was used in the search process as no literature review on this specific topic had been conducted before.

Article Inclusion and Exclusion Criteria

During the second stage, screening was executed based on the inclusion and exclusion criteria, which were as follows:

1) The study needs to be empirical in nature (no software reviews or literature reviews)
2) The study needs to be scientific in nature (at least 10 references)
3) The study needs to be in English

4) Studies addressing only the 'learning curves' of a specific VR hardware or software are excluded as they do not consider learning outcomes in VR
5) Studies that do not include VR with HMD (e.g., contains only AR, mixed reality or 360-degree video technologies), are excluded.
6) Studies concerning only navigation or spatial learning in VRLEs are excluded. Spatial learning is 'learning' in the sense that users learn to navigate better in virtual environments or real-life spaces with the help of VR training. However, as the focus of this literature review were mainly on exact learning outcomes, spatial learning was excluded.

Following the eligibility screening, the inclusion and exclusion criteria were revisited to assess the coded full texts. Finally of 1,086 studies, 37 were selected for further analysis. All articles were coded according to the following information: bibliometric information, research context and participants, methods, research setting and results regarding learning outcomes.

The framework presented by Radianti et al. (2020) was used to categorise the different learning outcomes. In their literature review, they used four learning content types from previous literature: 1) analytical and problem-solving, 2) communication, collaboration and soft skills, 3) procedural–practical knowledge and 4) declarative knowledge. They also added four more during the coding process: 5) learning a language, 6) behavioural impacts, 7) others and 8) not specified. The categories 'others' and 'not specified' used by Radianti et al. (2020) were removed from this article as one of the inclusion criteria was the requirement of articulated learning outcomes. Furthermore, the category 'behavioural impacts' was removed from this analysis due to a clear focus on learning outcomes on this article, and the category 'analytical and problem-solving' was not found in the articles included on the analysis. The learning goals framework is presented in Table 1.

Table 1

Modified Framework of Learning Goals (originally presented in Radianti et al., 2020)

Categories used in this paper	Description
Declarative knowledge	VR is used to memorise factual knowledge, e.g., learning history
Procedural–practical knowledge	VR is used to assist students with internalising procedures, e.g., driving a car

Categories used in this paper	Description
Communication, collaboration, and soft skills	VR is used to improve students' ability to work in a team or improve their communication skills or soft skills
Learning a language	VR is used to improve foreign language capabilities, e.g., to speak a foreign language

Once the search process was completed, a table (Appendix 1) was constructed from the review studies. The table contained the following information from every study:

1. Bibliographic information (author(s) and year of publication)
2. Participants (including their education level and/or context and age, if specified on the article)
3. Learning goals
4. Research setting. Each study is categorised under one condition: (1) only VR, (2) VR and other technology and (3) VR and other method. In the 'only VR' group, the study only had one HMD-VR group or only HMD-VR groups. In the 'VR and other technology' group, the studies included an HMD-VR group or groups and at least one other group which used a different technology, for example, desktop-based VR. In the 'VR and other method' group, the studies included a VR group and at least one other group which used a different method for teaching and/or learning.
5. Results on learning outcomes. Learning outcomes were separated 'between time points' and 'between groups'. If the study measured overall learning outcomes between two time points (e.g., pre-intervention and post-intervention), the result was shown in the 'between time points' column. If the study did not measure or report the learning outcome change between time points, the column was marked with N/A. The 'between groups' column showed whether there was a significant difference between the groups in terms of learning outcomes.
6. Results on learning motivation/engagement.

Findings

This section presents the results of the review and focuses on the learning outcomes of VRLEs, both in terms of direct and indirect (learning motivation/engagement) learning outcomes. Furthermore, the direct learning outcomes were categorised according to different learning goals (declarative knowledge, procedural-practical knowledge, communication, collaboration and soft skills, learning a language) modified based on Radianti et al. (2020) (Table 1). In addition, studies including the learning motivation and engagement -measures are presented as one category. A summary is presented to conclude the findings of 37 studies included in the review. (Table 2). A complete list of the included studies can be found in the Appendix 1.

Table 2

Summary from the Analysis (n = 37)

Description	Number of studies (studies under a theme/total number of studies)
Studies which compared VR group and another technology/method/control group	31/37 (from total studies)
The HMD-VR group yielded better results than the technology/method/control group	7/31 (from studies with HMD-group and technology/method/control group)
The technology/method/control group yielded better results than the HMD-VR group	3/31 (from studies with HMD-group and technology/method/control group)
Studies which only compared VR group(s), a significant change between two time points or groups was found	5/8 (from studies which only compared VR group(s))
Studies categorised by different learning goals	
Studies with the goal of declarative knowledge learning	22/37 (from total studies)
Studies with the goal of procedural–practical knowledge learning	13/37 (from total studies)
Studies with the goal of language learning	2/37 (from total studies)
Studies which measured learning motivation and/or engagement. All of them found a statistically significant difference favouring the VR group	8/8 (from studies measuring learning motivation and/or engagement.).

Learning Goals - Declarative Knowledge

A total of 22 articles focused on declarative knowledge, where the primary learning goal was to memorise factual knowledge, such as biology- or history-related facts. Fifteen of them examined differences between VR methods and other technologies or methods. In five of them, the VR group was found to have statistically significantly higher learning outcomes compared to other groups. Freitas et al. (2020) studied learning about computer memory management and allocation techniques within two groups: HMD VR and traditional teaching. The VR group had significantly higher learning outcomes that the traditional teaching group. Differences regarding the learning outcomes between the time points were not reported. Liu et al. (2020) studied learning about science knowledge on three topics: leverage, animals, and plants. They compared the HMD-VR and traditional teaching groups. The former had significantly higher learning outcomes. Differences regarding the learning outcomes between the time points were not reported. Lui et al. (2020) compared learning about microbiology among three groups: HMD-VR in a standing position, HMD-VR in a sitting position and a control group (desktop VR). The sitting-position HMD-VR group had significantly higher learning outcomes than the control group and the standing-position HMD-VR group. Differences between time points were not reported. Ou et al. (2021) compared learning about a Taipei tree frog between HMD-VR and desktop-VR groups. There was a significant difference in learning outcomes between time points; in the post-test, learning outcomes were significantly higher in the HMD-VR group. Parmar et al. (2016) studied learning about electrical measurement devices between HMD-VR and desktop-VR groups. There was a significant difference in learning outcomes between the time points; in the post-test, learning outcomes were significantly higher in the VR group.

Seven studies included only a VR group(s), five of which reported a significantly improved learning outcome change between the time points: Bhargava et al. (2018) examined the study of metrology-related concepts; Chowdhury and Quarles (2021) analysed learning about multiple sclerosis; Kwon (2019) studied the learning of differences between the Moon and Earth; Rudolph et al. (2020) examined geoscience learning about Grand Canyon rocks; and Teranishi and Yamagishi (2018) studied learning about the names and positions of the parts in a PC assembly.

In Moreno and Mayer's (2004) study about how to design a plant's roots, stem and leaves to enable survival in five environments, the desktop-VR group had statistically significantly better results than the VR group. Differences regarding learning outcomes between time points were not reported.

Two studies found a significant difference between time points, but not between groups (Bertrand, 2017; Bhowmick et al., 2018; Dengel, 2020).

The remaining eight articles did not find statistically significant differences/changes between the groups and/or time points (Chen et al., 2019; Chowdhury et al., 2019; Hadjipanayi & Michael-Grigoriou, 2021; Klingenberg et al., 2020; Moesgaard et al., 2015; Moreno et al., 2002; Stepan et al., 2017; Taylor & Barnett, 2010).

Procedural-Practical Knowledge

There was a total of 12 articles in which procedural–practical knowledge (e.g., learning triage skills) was considered the main learning goal, two of which found that the VR group had significantly better learning outcomes than the comparison group (Agrawal et al., 2018; Gutierrez et al., 2007). Agrawal et al. 2018 studied learning about hazard anticipation and mitigation skills in driving between two desktop-VR groups and an HMD-VR group. Changes between the time points were not reported, but on the post-test, the VR group recorded significantly better results. Gutierrez et al. (2007) studied learning about how to conduct a physical exam and found a significant difference between the time points. Furthermore, post-test comparisons between the HMD-VR and desktop-VR groups revealed that the VR group had significantly better results.

In two studies, the desktop-VR group had statistically significantly better results than the VR group (Lai et al., 2021; Makransky et al., 2019). Lai et al. (2021) studied learning about how to conduct different types of chemical reactions and the making of a galvanic cell, while Makransky et al. (2019), examined learning about developing an understanding of mammalian transient protein expression. Lai et al. (2021) found a significant difference between the time points, but difference between the time points was not reported in Makransky et al. (2019).

Two studies found a significant difference between time points, but not between groups (Jung & Ahn, 2018; Osti et al., 2020).

The remaining six studies found no statistically significant differences/changes between the groups and/or time points, or they were not reported (Buttussi & Chittaro, 2017; Grassini et al., 2020; Lerner et al., 2020; Nystad, 2006; Sportillo et al., 2018; Yu et al., 2021).

Communication, Collaboration and Soft Skills

One study (An et al., 2018), which was categorised under communication, collaboration and soft skills, examined the development of cross-cultural competence skills. The authors found no statistically significant between-group differences, and changes between the time points were not reported.

Learning a Language

Two studies focused on learning a language as a primary learning outcome. Ebert et al. (2016) studied the learning of Swedish language vocabulary, with the results indicating that while the traditional methods resulted in significantly better memory immediately after the training, the VR group had a statistically significant better record on word retention after a week. Nicolaidou et al. (2021) studied the learning of Greek words by comparing HMD-VR and mobile application groups. While they found a significant difference in learning outcomes between the time points, no between-group differences were found.

Learning Motivation and Engagement

In addition, we explored how HMD-VR affected students' indirect learning outcomes. Learning motivation and/or engagement were considered in eight articles (Buttusi et. al, 2017; Bhowmick et al., 2018; Klingenberg et al., 2020; Stepan et al., 2017; Lerner et al., 2020; Liu et al., 2020; Nicolaidou et al., 2020; Taylor & Barnett, 2010). Taylor and Barnett (2010) studied the learning of procedural tasks concerning military movements and found that the control group experienced significantly less engagement and interests/enjoyment than the HMD-VR and desktop-VR groups. There was no difference between the HMD-VR and desktop-VR groups in learning outcomes concerning training retention. Stepan et al. (2017) studied the learning of clinical anatomy among university students. Compared to the textbook group, learning in the HMD-VR group was significantly more engaging, enjoyable, useful and motivating. There was no difference between the groups in terms of learning outcomes regarding clinical anatomy knowledge. Engagement was measured in Buttusi et al. (2017), where knowledge about flight safety was studied among three groups: high- and low-fidelity HMD-VR and desktop-VR groups. The high-fidelity group reported significantly higher engagement than the desktop-VR group, but there was no statistically significant difference between the high-fidelity and low-fidelity HMD-VR groups. There were no significant differences between the groups in knowledge acquisition about flight safety. Bhowmick et al. (2018) studied engagement relating to the learning of the midwifery process. The HMD-VR group reported significantly more engagement than the 2D-video group, but there were no significant differences between the groups in learning outcomes. Klingenberg et al. (2020) studied the effects of teaching as a generative learning strategy in biochemistry using the desktop and immersive VR. Regarding intrinsic motivation, there was no significant difference between the groups after one intervention, but when the students experienced both conditions, there was a statistically significant difference that favoured immersive VR. Lerner et al. (2021) studied emergency simulation training in HMD-VR and found that the VR training significantly increased the intrinsic motivation of the participants. Liu et al. (2020) studied the effects of an HMD-VR-based classroom in the context of science lessons. Their results showed that students who took the science lessons using HMD-VR had better behavioural, cognitive, emotional and social engagement compared to the group which took traditional teaching lessons.

However, it is worth noting that as the primary focus of this literature review was direct learning outcomes in VR, only studies using some kind of measurement for direct learning outcomes were included. Therefore, we suggest that future research further investigate the indirect components of learning (learning motivation and engagement) outcomes.

Summary

In summary, seven of the 31 studies comparing a VR group and another technology/method/control group included significant differences/changes that favoured the HMD-VR group (Agrawal et al., 2018; Freitas et al., 2020; Gutierrez et al., 2007; Liu et al., 2020; Lui et al., 2020; Ou et al., 2021; Parmar et al., 2016). In three studies, the technology/method/control group yielded better results than the HMD-VR group (Lai et al. 2021; Makransky et al., 2019; Moreno & Mayer, 2004). The remaining studies found no statistically significant between-group differences (See also Table 2).

For the studies comprising only a VR group(s), five of them recorded a significantly improved change in learning outcomes between the two time points or between groups (Bhargava et al., 2018; Chowdhury & Quarles, 2021; Kwon, 2019; Rudolph et al., 2020; Teranishi & Yamagishi, 2018). Regarding the learning goals, the declarative knowledge goal recorded the biggest group (22 studies), followed by procedural–practical knowledge (12 studies), learning a language (2 studies) and communication, collaboration and soft skills (1 study). Between the different learning goals, there were no clear differences in the number of studies with significant differences between the groups or time points. Furthermore, all eight studies measuring learning motivation and/or engagement had a statistically significant difference that favoured the VR group.

One anomaly was removed from the summarised figures. In Ebert et al. (2016), the immediate post-test scores revealed that the traditional method was significantly better, but the VR group scored significantly higher on the retention test. This suggests a significant difference favouring both the HMD-VR and traditional methods groups.

Based on these results, it seems that HMD-VR provides generally higher engagement and motivation, though not necessarily higher learning outcomes (e.g., knowledge acquisition). However, more research is needed to have more reliable results. Here, the differences were not always statistically significant.

Discussion

The main goal of this systematic literature review was to explore the evidence in previous research regarding learning outcomes in the context of HMD-VR use in all subject areas and educational levels. The purpose was to fill the gap from previous literature reviews on the topic. A total of 37 studies were analysed. From the general trend of the results, we can conclude that there were no notable differences between the technologies (HMD-VR, desktop VR, CAVE)—only 23% of the studies comparing VR to another technology/method/control condition had better learning outcomes than the comparison group. This suggests that the exclusive use of HMD-VR does not generally yield better outcomes. However, the results regarding learning motivation and engagement suggest that HMD-VR brings more engagement and motivation to the learning process, likely influencing learning outcomes in the long run. Regarding the learning goals, there were no clear differences in the number of studies recording a significant difference between the groups or time points. In other words, no specific learning goal was found to be more suitable for HMD-VR than another method.

Furthermore, more longitudinal research is needed to determine learning outcomes in the long term. Longitudinal research could also help reduce one very important limitation, that, is the motivation caused by the VR technology itself, not the content of the HMD-VR.

In total, 26 of the studies were published in journals, while the remaining 11 studies were published in conference proceedings. In addition, as the time scope of the studies was so wide, we could not tell whether the improved learning outcomes from VR use were caused by the superiority of VR or improvements in the VR technology over the years – most studies concluding that HMD-VR yields better outcomes than comparison groups are published in recent years (Agrawal et al., 2018; Freitas et al., 2020; Gutierrez et al., 2007; Liu et al., 2020; Lui et al., 2020; Ou et al., 2021; Parmar et al., 2016).

One implication is that as the VR technology is further developed, we may see corresponding improvements in learning outcomes. Therefore, it would be beneficial to reproduce this literature review in the future to ascertain whether the trend regarding the evidence has changed. Based on the publication years of the studies, we can see that there will be a growing number of studies on this topic, including in the context of 2021.

Future research avenues include the wider pedagogical framework within the use of VRLEs as a part of the learning path. For example, when and how VRLEs should be used in the learning path to support learning? Furthermore, it is important to investigate the purpose of VRLEs and how they could be integrated with other learning methods and technologies (Lähtevänoja et al., 2020). It would also be fruitful to do a deeper dive into the different learning goals of HMD-VR. In this study, no differences regarding learning outcomes were found between the different learning goals.

It might also be important for future research to consider the pedagogical design of VRLEs. Based on the results of this literature review, it seems that using only the HMD-VR technology does not yield better learning outcomes. Perhaps a different pedagogical design/approach to VRLEs should be considered. Future research should, therefore, study the right pedagogical design for each technology.

Direct and Indirect Learning Outcomes

The main finding of this research is that while HMD-based VRLEs may not be superior to other technologies (such as desktop VR) regarding direct learning outcomes, a clear indication towards improved indirect learning outcomes (learning motivation, engagement) was detected. Based on the results of this literature review, it seems that HMD-VR-based VRLEs are more supportive of learning motivation and engagement than more traditional desktop-VR-based learning environments. This result is in line with a previous literature review on the topic (Queiroz et al., 2018), which also found that VR groups have more task motivation and engagement than groups deploying other technologies or methods.

Limitations

It is important to take the exclusion criteria of this literature review into consideration when making comparisons with other literature reviews. For example, the strict definition of HMD-VR used in this article implies that all studies using 360-degree videos and no computer-generated graphics were excluded. In addition, the search terms may have excluded some studies on learning motivation and engagement as they were not usually labelled as 'learning outcomes' in the articles. Therefore, future research should conduct a literature review based only on learning motivation and engagement in VRLEs.

In addition, it needs to be noted that studies that did not have a clear educational focus were still included as long as they had proper learning outcomes. These studies included papers that primarily examined the usability and feasibility of a certain VRLE; however, they still provided a subject matter to be taught to the participants and looked at learning outcome measures in order to investigate the possible learning effects of the environment. In these studies, the primary focus was not on learning.

Acknowledgements

This study has been partly conducted by the funding of University of Jyväskylä, Department of Teacher Education.

References

An, B., Matteo, F., Epstein, M., & Brown, D. (2018). Comparing the performance of an immersive virtual reality and traditional desktop cultural game. In *Proceedings of the 2nd International Conference on Computer-Human Interaction Research and Applications* (Vol. 1, pp. 54–61). https://doi.org/10.5220/0006922800540061

Agrawal, R., Knodler, M., Fisher, D. L., & Samuel, S. (2018). Virtual reality headset training: Can it be used to improve young drivers' latent hazard anticipation and mitigation skills. *Transportation Research Record, 2672*(33), 20–30. https://doi.org/10.1177/0361198118758311

Bertrand, J., Bhargava, A., Madathil, K. C., Gramopadhye, A., & Babu, S. V. (2017). The effects of presentation method and simulation fidelity on psychomotor education in a bimanual metrology training simulation. *Proceedings of the 2017 IEEE Symposium on 3D User Interfaces* (pp. 59–68). https://doi.org/10.1109/3DUI.2017.7893318

Bhargava, A., Bertrand, J. W., Gramopadhye, A. K., Madathil, K. C., & Babu, S. V. (2018). Evaluating multiple levels of an interaction fidelity continuum on performance and learning in near-field training simulations. *IEEE Transactions on Visualization and Computer Graphics, 24*(4), 1418–1427. https://doi.org/10.1109/TVCG.2018.2794639

Bhowmick, S., Darbar, R., & Sorathia, K. (2018, September). Pragati: Design and evaluation of a mobile phone-based head mounted virtual reality interface to train community health workers in rural India. *Proceedings of the NordiCHI* (pp. 299–310). https://doi.org/10.1145/3240167.3240201

Boell, S.K., Cecez-Kecmanovic, D. (2015). On being 'systematic' in literature reviews. In: Willcocks, L.P., Sauer, C., Lacity, M.C. (eds) Formulating Research Methods for Information Systems (pp. 48-78). Palgrave Macmillan, London. https://doi.org/10.1057/9781137509888_3

Bosse, T., Gerritsen, C., de Man, J., Treur, J. (2014). Towards virtual training of emotion regulation. *Brain Informatics, 1*(1–4), 27–37. https://doi.org/10.1007/s40708-014-0004-9

Buttussi, F., & Chittaro, L. (2017). 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

Bradley, R. and Newbutt, N. (2018), "Autism and virtual reality head-mounted displays: a state of the art systematic review", Journal of Enabling Technologies, 12 (3), 101-113. https://doi.org/10.1108/JET-01-2018-0004

Carretero, M. D. P., García, S., Moreno, A., Alcain, N., & Elorza, I. (2021). Methodology to create virtual reality assisted training courses within the Industry 4.0 vision. *Multimedia Tools and Applications, 80*(19), 29699–29717. https://doi.org/10.1007/s11042-021-11195-2

Chen, Y. T., Hsu, C. H., Chung, C. H., Wang, Y. S., & Babu, S. V. (2019). iVRNote: Design, creation and evaluation of an interactive note-taking interface for study and reflection in VR learning environments. In *Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 172–180). IEEE. https://doi.org/10.1109/VR.2019.8798338

Chowdhury, T. I., Ferdous, S. M. S., & Quarles, J. (2019). VR disability simulation reduces implicit bias towards persons with disabilities. IEEE Transactions on Visualization and Computer Graphics, 27(6), 3079-3090. https://doi.org/10.1109/TVCG.2019.2958332

Chowdhury, T. I., & Quarles, J. (2021). A wheelchair locomotion interface in a VR disability simulation reduces implicit bias. *IEEE Transactions on Visualization and Computer Graphics*. https://doi.org/10.1109/tvcg.2021.3099115

Deci, E. L., & Ryan, R. M. (2010). Intrinsic Motivation The Corsini Encyclopedia of Psychology: JohnWiley & Sons, Inc. https://doi.org/10.1002/9780470479216.corpsy0467

Dengel, A. (2020). How important is immersion for learning in computer science replugged games? In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (pp. 1165–1171). https://doi.org/10.1145/3328778.3366837

Ebert, D., Gupta, S., & Makedon, F. (2016). Ogma: A virtual reality language acquisition system. In Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments (pp. 1-5). https://doi.org/10.1145/2910674.2910681

Everaert, P., Opdecam, E., & Maussen, S. (2017). The relationship between motivation, learning approaches, academic performance and time spent. *Accounting Education, 26*(1), 78–107. https://doi.org/10.1080/09639284.2016.1274911

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 & Education, 127, 252–266. https://doi.org/10.1016/j.compedu.2018.09.002

Freina, L., & Ott, M. (2015). A literature review on immersive virtual reality in education: State of the art and perspectives. *The International Scientific Conference eLearning and Software for Education* 1(133), 10-1007.

Freina, L., Bottino, R. & Tavella, M. (2016). From e-learning to VR-learning: An example of learning in an immersive virtual world. *Journal of e-Learning and Knowledge Society, 12*(2).

Freitas, L. F. S., Ancioto, A. S. R., Guimarães, R. D. F. R., Martins, V. F., Dias, D. R. C., & de Paiva Guimarães, M. (2020). A virtual reality simulator to assist in memory management lectures. In *Proceedings of the International Conference on Computational Science and its Applications* (pp. 810–825). Springer. https://doi.org/10.1007/978-3-030-58820-5_58

Coulter, R., Saland, L., Caudell, T., Goldsmith, T. E., & Alverson, D. (2007). The effect of degree of immersion upon learning performance in virtual reality simulations for medical education. InMedicine Meets Virtual Reality, 15, 155.

Guttentag, D. A. (2010). Virtual reality: Applications and implications for tourism. *Tourism Management, 31*(5), 637–651. https://doi.org/10.1016/j.tourman.2009.07.003

Grassini, S., Laumann, K., & Skogstad, M. R. (2020). The use of virtual reality alone does not promote training performance (but sense of presence does). *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.01743

Hadjipanayi, C., & Michael-Grigoriou, D. (2021). Arousing a wide range of emotions within educational virtual reality simulation about major depressive disorder affects knowledge retention. *Virtual Reality,* 1–17. https://doi.org/10.1007/s10055-021-00568-5

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

Hanson, K., & Shelton, B. E. (2008). Design and development of virtual reality: Analysis of challenges faced by educators. *Educational Technology & Society, 11*(1), 118–131.

Huang, H. M., Rauch, U., & Liaw, S. S. (2010). Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach. Computers & Education, 55(3), 1171-1182. https://doi.org/10.1016/j.compedu.2010.05.014

Ijsselsteijn, W. A., & Riva, G. (2003). Being There: The experience of presence in mediated environments. In G. Riva, F. Davide, & W. A. Ijsselsteijn (Eds.), *Being there: Concepts, effects and measurement of user presence in synthetic environments* (pp. 4–16). IOS Press.

Holopainen, J., Lähtevänoja, A., Mattila, O., Södervik, I., Pöyry, E., & Parvinen, P. (2020). Exploring the learning outcomes with various technologies—Proposing design principles for virtual reality learning environments. In *Proceedings of the 53rd Hawaii International Conference on System Sciences* (pp. 12–21). https://doi.org/10.24251/HICSS.2020.004

Hsieh, T. L. (2014). Motivation matters? The relationship among different types of learning motivation, engagement behaviors and learning outcomes of undergraduate students in Taiwan. *Higher Education, 68*(3), 417–433. https://doi.org/10.1007/s10734-014-9720-6

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

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), e1812. https://doi.org/10.1002/cav.1812

Klingenberg, S., Jørgensen, M. L., Dandanell, G., Skriver, K., Mottelson, A., & Makransky, G. (2020). Investigating the effect of teaching as a generative learning strategy when learning through desktop and immersive VR: A media and methods experiment. British *Journal of Educational Technology, 51*(6), 2115–2138. https://doi.org/10.1111/bjet.13029

Kuo, M. J. (2007, March). How does an online game based learning environment promote students' intrinsic motivation for learning natural science and how does it affect their learning outcomes? In *Proceedings of the 2007 First IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL'07)* (pp. 135–142). IEEE. https://doi.org/10.1109/DIGITEL.2007.28

Kwon, C. (2019). Verification of the possibility and effectiveness of experiential learning using HMD-based immersive VR technologies. *Virtual Reality, 23*(1), 101–118. https://doi.org/10.1007/s10055-018-0364-1

Lai, T. L., Lin, Y. S., Chou, C. Y., & Yueh, H. P. (2022). Evaluation of an Inquiry-Based Virtual Lab for Junior High School Science Classes. Journal of Educational Computing Research, 59(8), 1579-1600. https://doi.org/10.1177/07356331211001579

Lerner, D., Mohr, S., Schild, J., Göring, M., & Luiz, T. (2020). An immersive multi-user virtual reality for emergency simulation training: Usability study. *JMIR Serious Games, 8*(3), e18822. https://doi.org/10.2196/18822

Liu, R., Wang, L., Lei, J., Wang, Q., & Ren, Y. (2020). Effects of an immersive virtual reality-based classroom on students' learning performance in science lessons. *British Journal of Educational Technology, 51*(6), 2034–2049. https://doi.org/10.1111/bjet.13028

Lui, M., McEwen, R., & Mullally, M. (2020). Immersive virtual reality for supporting complex scientific knowledge: Augmenting our understanding with physiological monitoring. *British Journal of Educational Technology, 51*(6), 2180–2198. https://doi.org/10.1111/bjet.13022

Lähtevänoja, A., Holopainen, J., Vesisenaho, M., & Häkkinen, P. (2021). Developing design knowledge and a conceptual model for virtual reality learning environments. In

Designing, Deploying, and Evaluating Virtual and Augmented Reality in Education (pp. 100-123). IGI Global. https://doi.org/10.4018/978-1-7998-5043-4.ch005

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

Mann, S., Furness, T., Yuan, Y., Lorio, J., & Wang, Z. (2018). All reality: virtual, augmented, mixed (X), mediated (X,Y), and multimediated reality. Retrieved August 2, 2018, from Cornell University, arXiv website: https://arxiv.org/abs/1804.08386.

Moesgaard, T., Witt, M., Fiss, J., Warming, C., Klubien, J., & Schoenau-Fog, H. (2015). Implicit and explicit information mediation in a virtual reality museum installation and its effects on retention and learning outcomes. *Proceedings of the European Conference on Games-Based Learning* (pp. 387–394).

Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. Journal of educational psychology, 94(3), 598. https://doi.org/10.1037/0022-0663.94.3.598

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of educational Psychology, 96*(1), 165. https://doi.org/10.1037/0022-0663.96.1.165

Morschheuser, B., Hamari, J., Koivisto, J., & Maedche, A. (2017). Gamified crowd-sourcing: Conceptualization, literature review, and future agenda. *International Journal of Human-Computer Studies, 106*, 26-43. https://doi.org/10.1016/j.ijhcs.2017.04.005

Neges, M., Adwernat, S., & Abramovici, M. (2018). Augmented Virtuality for maintenance training simulation under various stress conditions. *Procedia Manufacturing, 19*, 171–178. https://doi.org/10.1016/j.promfg.2018.01.024

Nicolaidou, I., Pissas, P., & Boglou, D. (2021). Comparing immersive virtual reality to mobile applications in foreign language learning in higher education: A quasi-experiment. *Interactive Learning Environments,* 1–15. https://doi.org/10.1080/10494820.2020.1870504

Osti, F., de Amicis, R., Sanchez, C. A., Tilt, A. B., Prather, E., & Liverani, A. (2020). A VR training system for learning and skills development for construction workers. *Virtual Reality,* 1–16. https://doi.org/10.1007/s10055-020-00470-6

Ou, K. L., Liu, Y. H., & Tarng, W. (2021). Development of a virtual ecological environment for learning the Taipei tree frog. *Sustainability, 13*(11),

5911. https://doi.org/10.3390/su13115911

Queiroz, A. C. M., Nascimento, A. M., Tori, R., & da Silva Leme, M. I. (2018). Using HMD-based immersive virtual environments in primary/K-12 education. In *International Conference on Immersive Learning* (pp. 160–173). Springer. https://doi.org/10.1007/978-3-319-93596-6_11

Parmar, D., Bertrand, J., Babu, S. V., Madathil, K., Zelaya, M., Wang, T., ... Frady, K. (2016). A comparative evaluation of viewing metaphors on psychophysical skills education in an interactive virtual environment. *Virtual Reality, 20*(3), 141–157. https://doi.org/10.1007/s10055-016-0287-7

Pollard, K. A., Oiknine, A. H., Files, B. T., Sinatra, A. M., Patton, D., Ericson, M., ... Khooshabeh, P. (2020). Level of immersion affects spatial learning in virtual environments: Results of a three-condition within-subjects study with long intersession intervals. *Virtual Reality,* 1–14. https://doi.org/10.1007/s10055-019-00411-y

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,* 103778. https://doi.org/10.1016/j.compedu.2019.103778

Rudolph, B., Musick, G., Wiitablake, L., Lazar, K.B., Mobley, C., Boyer, D.M., Moysey, S., Robb, A. and Babu, S.V. (2020.) Investigating the Effects of Display Fidelity of Popular Head-Mounted Displays on Spatial Updating and Learning in Virtual Reality. In International Symposium on Visual Computing (pp. 666-679). Springer, Cham. https://doi.org/10.1007/978-3-030-64556-4_52

Sathe, V., Gupta, P., Kaushik, K., Bhat, S., & Deshpande, S. (2017). Virtual reality websites (VR WEB). In *Proceedings of the 2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA)* (Vol. 1, pp. 647–652). IEEE. https://doi.org/10.1109/ICECA.2017.8203619

Sebok, A., & Nystad, E. (2006). Procedural training in virtual reality: A comparison of technology types. *Proceedings of NPIC&HMIT* (pp. 12–16).

Sportillo, D., Paljic, A., & Ojeda, L. (2018). Get ready for automated driving using virtual reality. *Accident Analysis & Prevention, 118,* 102–113. https://doi.org/10.1016/j.aap.2018.06.003

Stepan, K., Zeiger, J., Hanchuk, S., Del Signore, A., Shrivastava, R., Govindaraj, S., & Iloreta, A. (2017). Immersive virtual reality as a teaching tool for neuroanatomy. *International Forum of Allergy and Rhinology, 7*(10), 1006–1013.

https://doi.org/10.1002/alr.21986

Taylor, G. S., & Barnett, J. S. (2010, September). Training effectiveness of wearable and desktop simulator interfaces. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54*(27), 2267–2271. Sage. https://doi.org/10.1177/154193121005402710

Teranishi, S., & Yamagishi, Y. (2018). Educational effects of a virtual reality simulation system for constructing self-built PCs. *Journal of Educational Multimedia and Hypermedia, 27*(3), 411–423.

Triarisanti, R., & Purnawarman, P. (2019). The influence of interest and motivation on college students'language and art appreciation learning outcomes. International Journal of Education, 11(2), 130-135. https://doi.org/10.17509/ije.v11i2.14745

Vasalampi, K., Muotka, J., Malmberg, L. E., Aunola, K., & Lerkkanen, M. K. (2021). Intra-individual dynamics of lesson-specific engagement: Lagged and cross-lagged effects from one lesson to the next. British Journal of Educational Psychology, 91(3), 997-1014. https://doi.org/10.1111/bjep.12404

Vesisenaho, M, Juntunen, M., Häkkinen, P, Pöysä-Tarhonen, J., Miakush, I., Fagerlund, J. & Parviainen, T. (2019). Virtual reality in education: Focus on the role of emotions and physiological reactivity. *Journal of Virtual Worlds Research, 12*(1), 1–15. https://doi.org/10.4101/jvwr.v12i1.7329

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly, 26(2), xiii–xxiii. http://www.jstor.org/stable/4132319

Yu, M., Yang, M., Ku, B., & Mann, J. S. (2021). Effects of Virtual Reality Simulation Program Regarding High-risk Neonatal Infection Control on Nursing Students. Asian Nursing Research, 15(3), 189-196. https://doi.org/10.1016/j.anr.2021.03.002

Appendix

Appendix 1

Studies Included in the Analysis

Author(s) and year	Participants (education level/context, age)	Learning goals	Method	Research setting in addition to VR	Results Between time points	Between groups	Learning motivation/engagement
Agrawal et al. (2018)	36 participants: young drivers, 18–25 years	Procedural–practical knowledge. Learning young drivers' latent hazard anticipation and mitiga tion skills related to driving with VR	Post-test driving simulator assessment. In-game data	Other technology	N/A	VR group had significantly higher results	
An et al. (2018)	11 reserve officer students and 10 non-military affiliated personnel, mean age 20.58	Communication, collaboration, soft skills. Teaching cross cultural competence	Pre-and post-test, situational judgement test, Observation	Other technology	N/A	No significant difference between the groups (VR performed worse), males outperformed females	
Bertrand (2017)	41 participants: students from computing and engineering classes, 18–38 years	Declarative knowledge. Learning how to operate metrology instruments	Pre- and post-test cognition questionnaires	Other technology	Significant difference between time points	No significant differences in learning outcomes between groups	
Bhargava et al. (2018)	65 participants: students from computing and engineering classes, 18–38 years	Declarative knowledge. Learning metrology-related concepts, e.g. how to take a measurement	Pre- and post-test cognition questionnaires	Only VR	Significant difference between time points	Different levels of interaction fidelity did not have significant effect on learning outcomes	
Bhowmick et al. (2018)	57 participants, 27–51 years	Declarative knowledge. Learning midwifery process regarding childbirth	Pre- and post-test cognition questionnaires administered orally	Other technology	Significant difference between time points	No significant differences on learning outcomes between groups	HMD VR caused significantly more engagement than the 2D-video group.

Author(s) and year	**Participants (education level/context, age)**	**Learning goals**	**Method**	**Research setting in addition to VR**	**Results** **Between time points**	**Between groups**	**Learning motivation/engagement**
Buttussi & Chittaro (2017)	96 participants: volunteers, mean age 18.36	Procedural–practical knowledge. Learning flight safety procedures	Pre-, post- and retention knowledge test administered orally	Other technology	Significant difference between time points	No significant differences in learning outcomes between groups	High-fidelity HMD VR had higher engagement than the desktop VR
Chen et al. (2019)	20 participants, years 21–27	Declarative knowledge. Learning history and meteorology with note-taking feature	Post-test questionnaire	Other method	No significant difference between time points	No significant differences in learning outcomes between groups	
Chowdhury et al. (2019)	71 participants: undergraduate students, median age 20.3 years	Declarative knowledge. Learning about multiple sclerosis	Post-test knowledge multi-choice test	Other technology	N/A	No significant differences in learning outcomes between groups (VR performed better)	
Chowdhury & Quarles (2021)	40 participants: undergraduate students, median age 23.6 years	Declarative knowledge. Learning about multiple sclerosis	Post-test knowledge multi-choice test	Only VR	N/A	Significant differences in learning outcomes between groups	
Dengel (2020)	78 participants: 13–16 years	Declarative knowledge. Learning about the components of a computer, asymmetric encryption/decryption and finite state machines	Pre- and post-test questionnaire	Other technology	Significant difference between time points	No significant differences in learning outcomes between groups	

Author(s) and year	**Participants (education level/context, age)**	**Learning goals**	**Method**	**Research setting in addition to VR**	**Results** **Between time points**	**Between groups**	**Learning motivation/engagement**
Ebert et al. (2016)	19 participants: student volunteers	Learning a language. Learning 10 Swedish language words	Pre- post- and retention language test (writing, pronunciation)	Other method	N/A	On the immediate post-test scores, the traditional method was significantly better. On the later post-test (one week later) scores, there was no significant differences between the groups. On the retention test, VR scored significantly higher	
Freitas et al. (2020)	80 participants: 16–53 years	Declarative knowledge. Learning computer memory management and allocation techniques	Pre-test (baseline test only) and post-test knowledge test	Other method	N/A	VR group had significantly higher learning outcomes compared to the traditional teaching group	
Gutierrez et al. (2007)	25 participants: first-year medical students	Procedural–practical knowledge. Learning how to conduct a physical exam	Pre- and post-test knowledge test	Other technology	Significant difference between time points	VR group had significantly higher learning outcomes	
Grassini et al. (2020)	29 participants: university students	Procedural–practical knowledge. Learning how to build a small airplane model	Performance metrics: product quality, errors made during assembly, speed of assembly	Other technology	N/A	No significant difference between the groups on performance metrics	
Hadjipanayi & Michael-Grigoriou (2021)	30 participants: over 18 years	Declarative knowledge. Learning about MDD symptoms	Pre- and post-test knowledge test	Only VR	No significant difference between time points	No significant difference between groups	

Author(s) and year	**Participants (education level/context, age)**	**Learning goals**	**Method**	**Research setting in addition to VR**	**Results** **Between time points**	**Between groups**	**Learning motivation/engagement**
Jung & Ahn (2018)	64 participants: students in maritime safety training institute, 23–35 years	Procedural–practical knowledge. Learning how to launch a lifeboat from a boat	Pre- and post-test questionnaire, observation (describing the process)	Other technology and other method	Significant difference between time points	No significant difference between groups	
Klingenberg et al. (2020)	89 participants: university students, 19–36 years	Declarative knowledge. Learning about the electron transport chain	Post-test transfer and retention test	Other technology	N/A	No significant difference between groups	No difference between groups on intrinsic motivation in post-test 1, but in post-test 2, there was a significant difference favouring the VR group
Kwon (2019)	42 participants: 11 years	Declarative knowledge. Learning the differences between Moon and Earth	Survey	Only VR	N/A	HMD VR with freedom to walk and hand tracker performed significantly higher	
Lai et al. (2021)	66 participants: 9th graders (high school)	Procedural–practical knowledge. Learning to do different types of chemical reactions and making a galvanic cell	Pre- and post-test knowledge test	Other technology	Significant difference between time points	Desktop VR group had significantly higher learning outcomes	
Lerner et al. (2020)	18 participants: active emergency physicians	Procedural–practical knowledge. Learning how to conduct an emergency scenario, including assessment and procedures	Pre- and post-test knowledge test	Only VR	No significant change in knowledge between pre- and post-test	No significant change in knowledge between pre- and post-test	Intrinsic motivation increased significantly between pre- and post-test
Liu et al. (2020)	90 participants: sixth-grade students, mean age 11 years	Declarative knowledge. Learning science knowledge (leverage, animals, plants)	Pre- and post-test knowledge test	Other method	N/A	HMD VR group had significantly higher learning outcomes	HMD VR group had significantly better engagement than traditional teaching method
Lui et al. (2020)	234 participants: university students	Declarative knowledge. Learning microbiology (complex gene regulation system)	Pre and post-test knowledge test	Other technology	N/A	HMD VR had significantly higher learning outcomes	

Author(s) and year	**Participants (education level/context, age)**	**Learning goals**	**Method**	**Research setting in addition to VR**	**Results** **Between time points**	**Between groups**	**Learning motivation/engagement**
Makransky et al. (2019)	52 participants: university students, 19–42 years	Procedural–practical knowledge. Learning to develop an understanding of mammalian transient protein expression	Post-test knowledge test, mobile sensing (EEG)	Other technology	N/A	Desktop VR had significantly higher learning outcomes	
Moesgaard et al. (2015)	40 participants: 19–31 years	Declarative knowledge. Learning what happened in Mosede Fort, Denmark, during World War I	Post-test knowledge test	Only VR	N/A	No significant differences between groups	
Moreno et al. (2002)	Experiment 1: 89 participants: college students Experiment 2: 75 participants: college students	Declarative knowledge Exp1: Learning concepts about plants Exp2: Learning how to design a plant	Post-test memory test and problem-solving test	Both experiments: other technology	N/A	No significant differences between groups	
Moreno & Mayer (2004)	48 participants: college students, mean age 19.54	Declarative knowledge. Learning how to design plant roots, stem and leaves in order to survive in five different environments	Post-test memory test and problem-solving test	Other technology	N/A	Desktop VR had significantly higher results on some learning outcomes (retention, not transfer)	
Nicolaidou et al. (2021)	40 participants: undergraduate students	Learning a language. Learning Greek words	Pre- and post-test vocabulary test	Other technology	Significant difference between time points	No significant differences between groups	No significant differences between groups on engagement
Osti et al. (2020)	20 participants: university students	Procedural–practical knowledge. Learning wooden wall construction	Pre-test (baseline), real-world wall construction test	Other technology	Significant difference between time points	No statistically significant difference between groups	

Author(s) and year	Participants (education level/context, age)	Learning goals	Method	Research setting in addition to VR	Results Between time points	Between groups	Learning motivation/engagement
Ou et al. (2021)	80 university students	Declarative knowledge. Learning about the Taipei tree frog	Pre- and post-test multi-choice achievement test	Other technology	Significant difference between time points.	Learning outcomes were significantly higher for the VR group	
Parmar et al. (2016)	24 participants: college students, 19–30 years	Declarative knowledge. Learning about electrical measurement instruments	Pre- and post-test cognition questionnaire, real-world skill test	Other technology	Significant difference between time points	Learning outcomes were significantly higher for the VR group	
Rudolph et al. (2020)	30 participants: 19–51 years	Declarative knowledge. Learning geoscience—the rocks in the Grand Canyon	Pre- and post-test knowledge test	Only VR	Significant difference between time points	Significant difference in high-fidelity and medium fidelity groups but not in low-fidelity group	
Sebok & Nystad (2006)	24 participants: employees of a reactor project, 25–61 years	Procedural–practical knowledge. Learning industrial training skills related to nuclear reactors	Post-test knowledge test, observation	Other technology	No significant difference between time points	No significant differences between groups	
Sportillo et al. (2018)	60 participants: volunteers with driver's license, 22–71 years	Procedural–practical knowledge. Learning driving with automated cars in different situations: manual mode, automated mode, and take-over request	Post-test driving simulator test drive (in-game data)	Other technology	N/A	No significant differences between groups	
Stepan et al. (2017)	66 participants: university medical students	Declarative knowledge. Learning brain anatomy	Pre- and post-test cognition questionnaire	Other method	No significant difference between time points	No significant changes in learning outcomes	HMD VR group had significantly higher engagement and motivation levels
Taylor & Barnett (2010)	98 participants: university students, mean age 18.9 years	Declarative knowledge. Learning tactical movement, selecting fighting positions in an urban environment and the use of hand grenades	Post-test video test (choose the correct option)	Other technology	N/A	No significant difference	HMD VR had higher engagement and interests/enjoyment than other groups

Author(s) and year	Participants (education level/context, age)	Learning goals	Method	Research setting in addition to VR	Results Between time points	Between groups	Learning motivation/engagement
Teranishi & Yamagishi (2018)	6 participants: university students	Declarative knowledge. Learning PC assembly: names and correct positions of the parts	Pre- and post-test knowledge questionnaires	Only VR	Significant change in learning outcomes regarding the positions of the parts but not in terms of the names of the parts	Significant change in learning outcomes regarding the positions of the parts but not in terms of the names of the parts	
Yu et al. (2021)	25 participants: nursing students	Procedural-practical knowledge. Learning basic nursing situations (e.g., feeding management)	Pre-and post-test knowledge questionnaire	Other method	N/A	No significant changes between groups	